6-1991

Effects of Training Method on Inspector Accuracy in a Simulated Quality Control Task

Cathy L. Thorne
Western Michigan University

EFFECTS OF TRAINING METHOD ON INSPECTOR ACCURACY IN A SIMULATED QUALITY CONTROL TASK

by

Cathy L. Thorne

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
June 1991

Order Number 9131136

Effects of training method on inspector accuracy in a simulated quality control task

Thorne, Cathy L., Ph.D.

Western Michigan University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

TABLE OF CONTENTS

Table of Contents-Continued

APPENDICES

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Interest in and concern for quality control has increased over the past decade for a variety of reasons, including intensified international competition, consumer expectations, and legal demands (Gallway & Drury, 1986). Philip Crosby, a consultant and expert in quality control, estimated that "product companies . . . spend 20 percent of their sales dollars doing things wrong and doing them over and service companies spend 35 percent or more of their operating costs doing things wrong and doing them over" (1984, p. 5).

Comprehensive quality control programs emphasize prevention, beginning with the initial design of the product, and continuing on with frequent inspection and early detection of problems (Mainstone & Levi, 1987). Recently, increased emphasis has been placed on having individual employees serve as their own inspectors, rather than having quality control inspectors identify defects at the end of an operation (Levi & Mainstone, 1987). Therefore, the need for efficient and effective training in defect detection has never been greater.

While increasingly sophisticated equipment can perform some inspection tasks previously performed by humans,

visual inspection remains an important component in most operations (Drury & Sinclair, 1983). Unfortunately, dependence on human detection of errors may lead to problems if poor inspection methods are used by workers. Harris and Chaney (1969) pointed out that a single inspector rarely detects more than 50 to 60% of defects. Studies reviewed by Fortune (1975) and Wiener (1987) support this claim, and these studies examined a wide range of inspection tasks, including surface appearance of tin plates; surface defects of piston rings; surface defects of ball-bearings; electronic module assemblies, surface finish of small parts; a small wired unit for defects in soldering, wiring, and appearance, machine parts; metal hooks and coins; telephone equipment; reading of x-ray films; and tonsillectomy recommendations. Thus, while there is an obvious need to improve inspector performance, Fortune concluded that most research "has been concerned with the goodness or validity of inspection and the causes of failure to achieve good results have not been studied to any extent" (1975, p. 33).

Previous Research on Inspector Accuracy

Since World War II, researchers in human factors have published over 1,000 articles on the subject of error

detection (Mackie, 1987). The initial impetus for this research was the fact that enemy submarines were not always detected by individuals monitoring radar screens. Early research thus concentrated on a type of inspection task termed a "monitoring task," those in which inspectors report the occurrence of some signal which indicates an out-of-tolerance condition (e.g., a blip on a radar screen or a momentary fluctuation of a recording device). Initial research utilized various aspects of a stimulus presentation as independent variables, including signal duration, signal intensity, signal frequency, inter-signal interval, and location of the stimuli; typical dependent variables were various aspects of inspector performance, including the number of signals missed (misses), the number of non-signals reported as signals (false positives), and response latency (Drury, 1978; Howell, Johnston, & Goldstein, 1966; Mackie, 1987).

Several researchers have expressed concern about the relevance of research on monitoring tasks to typical industrial inspection tasks. One concern is that laboratory results might not generalize to actual inspection situations. For example, a vigilance decrement, a decrease in error detection and an increase in false positives, is often reported in the laboratory after a

short period of time. Smith and Lucaccini (1969) pointed out that evidence for the decrement in industrial or military applications is weak. Another concern is that the research to date has generated few methods to improve monitoring accuracy (Mackie, 1987; Smith & Lucaccini, 1969). A final concern is the fact that monitoring tasks are not representative of the complexity of many inspection tasks. Researchers (Harris & Chaney, 1969; Kibler, 1965; Mackie, 1987; Smith & Lucaccini, 1969) pointed out that many inspection tasks are not as simple as typical monitoring tasks. With scanning tasks, the inspector is expected to make judgments or compare a product with a mental image:

> Scanning tasks are included in a wide variety of inspection jobs. Included are the examination of electronic hardware for improper connections, wiring component placement, and contamination; the inspection of fruit, vegetables, peanuts, and so on for the purpose of culling out items or classifying by grades; the inspection of automobile body surfaces for blemishes in paint; and the examination, by lot of any number of machine-produced products for purposes of identifying and enumerating defective items (Harris & Chaney, 1969, p. 140).

Considerable anecdotal evidence indicates that inspectors often are unsure of the classification of an item. Fortune (1975) discussed a study involving defective piston rings where 23% of decisions were reversed upon reinspection. In a study involving the inspection of

integrated circuits, Schoonard, Gould, and Miller (1973) reported that a small number of chips accounted for most detection errors; removing 2.7% of the chips led to a 29% reduction in overall error rate. Fox and Haslegrave (1969) reported reduced detection rates "because the criterion of acceptability for the types of screw being inspected was highly subjective" (p. 718).

The wide variation in performances between and within subjects suggests that the concept of an acceptable item may not be clear. Inspectors must respond in one way (accept the item) if specified stimulus characteristics are present and respond in another manner (reject the item) if specified stimulus characteristics are not present. Yet, it is not clear that inspectors are aware of which response to make in the presence of any given stimulus configuration for the item to be inspected. Smith and Lucaccini (1969) concluded that "criteria for defining defects in many products are not well established, indicating that inspectors do not always know what to look for. Thus, the inspection task . . . necessitates that an inspector know a defect when he sees it" (p. 153).

These examples suggest that training in identification of important characteristics and in detection of their presence or absence is a critical feature of any quality

control program. Yet, little research has been done on the effects of different types of training and detection effectiveness. Just as with monitoring tasks, previous research on scanning tasks concentrated on identifying stimulus features which affect inspector performance, rather than identifying ways in which to improve inspector performance. Independent variables have included percentage of defective items (Badalamente & Ayoub, 1969; Drury & Addison, 1973; Fortune, 1975); the location and spacing of stimuli (Drury & Clement, 1978; Monk & Brown, 1975); and the number and type of flaws (Baker, Morris, & Steedman, 1960; Gallway & Drury, 1986; Geyer, Patel, & Perry, 1979; Harris, 1966).

The few studies which have reported attempts to modify inspector performance have had some degree of success. For example, more rapid feedback to glass inspectors resulted in a 50% decrease in missed defective items (Schoonard et al., 1973). In other studies, the use of specific performance aids improved inspection performance on specific defects. In the inspection of silicon wafers, an overlay improved the detection of one type of error but had no effect on another type of error (Teel, Springer, & Sadler, 1968). In the inspection of circuit boards, blink inspection (where a good sample is rapidly alternated with

the item to be inspected) reduced errors for one type of defect but not for two other types of defects (Liuzzo & Drury, 1980).

Interventions with a training component also have been successful. Chaney and Teel (1967) reported increases in defective machine parts detected when visual aids were introduced alone or with training. Training consisted of lectures and demonstrations on topics related to measurement. In a simulated inspection task, Czaja and Drury (1981) compared the effects of active training, where a response was required and feedback provided, with passive training, where subjects were merely exposed to the training materials. Significantly fewer inspection errors were reported for subjects who received active training.

These examples demonstrate that inspector performance is not invariable, but can be modified. However, Czaja and Drury (1981) pointed out that training procedures are rarely described in sufficient detail to enable others to utilize the procedures in applied settings or to replicate the procedures for additional study. Thus, while the reported results are encouraging, additional research is required.

To summarize, correct defect detection is important in many inspection tasks, and an obvious need exists for

improvement in inspection accuracy. Furthermore, relatively little research has been conducted on methods of improving inspector accuracy, and research involving training effectiveness appears to be an important area to examine.

Inspector Accuracy and Stimulus Control

Stimulus control research indicates that training may be more effective when critical differences are presented in controlled fashion, i.e., varying only one attribute at a time rather than several differences simultaneously. If several stimulus attributes are varied in one example, responding may not come to be controlled by all relevant features (Mackintosh, 1977).

In an experiment using mentally retarded children as subjects, Allen and Fuqua (1985) initially trained subjects using error stimuli (S-) which varied from the correct stimuli (S+) in at least two ways. After subjects met criterion on identification of S+ and S- stimuli, they were first tested with novel incorrect instances which also differed from the S+ in at least two ways (Multiple Differences Test) and then tested with novel S- stimuli which differed from the S+ in only one way (Minimal Differences Test).

The results indicated that stimulus control deficits which were not apparent in a Multiple Differences Test appeared on a Minimal Differences Test. Subjects could identify stimuli which contained errors, but their performance was not controlled equally well by all types of errors. Some important stimulus elements did not acquire control over responding. This led to errors in identification of correct and incorrect stimuli.

Subjects who received Minimal Differences Training, where the S- stimuli differed from the S+ stimuli in only one way, were less likely to make incorrect responses in a test situation. The authors concluded that "selecting a range of S-'s which differ along the most relevant dimensions . . . and selecting each individual S- so that it differs in only one way from the S+, will improve the sensitivity of stimulus control tests and the effectiveness of stimulus discrimination training" (1985, p. 69).

Inspectors presumably have more experience with complex stimulus configurations than the subjects used in the Allen and Fuqua (1985) research. However, this research has important implications for inspector training. In this context, the purpose of the present research was to investigate the effects of different training strategies on inspector accuracy using a simulated inspection task.

Training methods were designed based on the research of Allen and Fuqua (1985). Accuracy of error detection in a set of geometric figures was measured as a function of training method. Three training methods were compared in a group design, including Critical Differences Training (CritDiff), Multiple Differences Training (MultDiff), and Correct Samples Only (CSO).

During training, all subjects received verbal descriptions of correct and incorrect stimlus elements, and all subjects were shown correct stimuli. The only difference among groups was the type of non-examples shown. Subjects in the CSO Group were not shown non-examples. Subjects in the CritDiff Group were shown non-examples which contained only one stimulus feature error at a time. Subjects in the MultDiff group were shown non-examples which contained multiple stimulus feature errors. Subjects were then administered a test, to ensure that the learning criterion had been met. Following training and testing, subjects completed repeated inspection tasks in which both one-error stimuli and two-error stimuli were present.

METHOD

Subjects

Subjects were recruited from undergraduate courses and were paid $20.00 for their participation in the research. Thirty-eight subjects, 18 females and 20 males, were randomly assigned to one of the three training groups (*N*=12 for the CritDiff Group, *N*=13 for the MultDiff and CSO Groups).

Inspection Task

The inspection task consisted of a paper-and-pencil task selected because of its significant correlation with a number of visual inspection tasks (Harris, 1964). The inspection task was also selected because of its neutrality with respect to subject history, its specific stimulus attributes, the relative ease of training, and comparatively low costs of administration.

Four figures (Figure 1) were used in training and testing. During testing, the four figures were presented in random order, 20 times per page (for a total of 80 figures per page). Figure 2 presents an example of an inspection page. Subjects inspected twenty pages of figures per

session and placed an "X" over figures which they considered defective. The same master page was used for all inspection sheets, and errors were pasted on after random selection of error type and location.

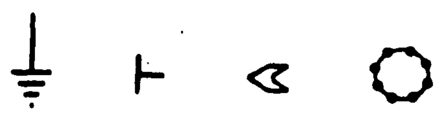


Figure 1. Correct Form of the Four Figures Used in the Inspection Task.

Figures 3-6 list correct specifications and specific defects for each figure. Each figure had three specified defects, with two variations of each defect. Thus, for each figure, it was possible to have six one-error variations and twelve two-error variations. While the placement of errors was randomized, each type of error was presented twice, to permit analysis by figure type and error type. The resulting defect rate (defective items/total number of items) was nine percent.

The fact that the task was a simulated inspection task allowed for precise control of errors. However, many of

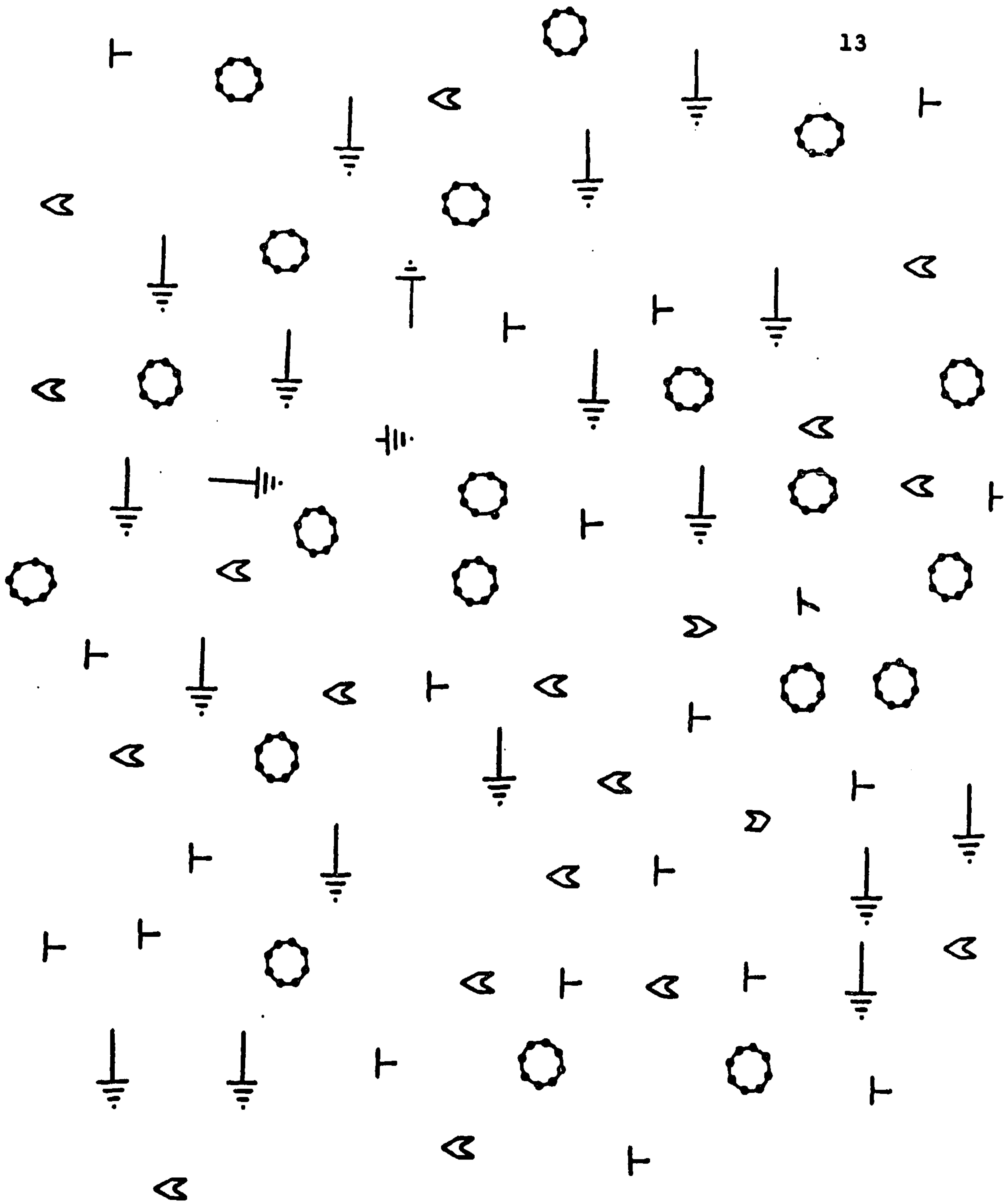


Figure 2. Example of an Inspection Page Used During Inspection Sessions.

Correct

1. Line 1 is to the left of line 2.

line 1  line 2

2. Line 2 bisects (cuts in half) line 1.

3. The two lines are at right angles to one another.

Incorrect

1a. Line 1 is above line 2.
1b. Line 1 is to the right of line 2.

2a. The two parts of line 1 are not equal in length; top segment is shorter than bottom segment.
2b. Top segment is longer than bottom segment.

3a. Line 2 slants in an upward direction.
3b. Line 2 slants in a downward direction.

One-Error Stimuli

Two-Error Stimuli

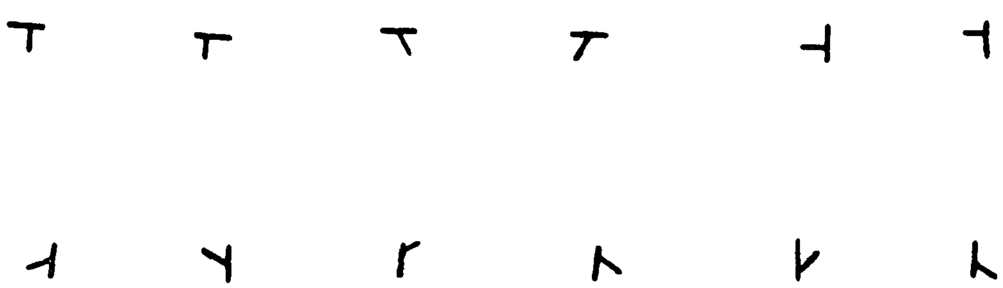


Figure 3. Correct Specifications and Specific Defects for the First Inspection Figure.

Correct	Incorrect
1. The angles point left. 	1a. The angles point down. 1b. The angles point right.
2. The figure is enclosed, a continuous line.	2a. Gaps or holes in the top half of the figure. 2b. Gaps or holes in the bottom half of the figure.
3. The sides of the angles are equal in length.	3a. Sides of angles are unequal in length; top segments are longer than bottom segments. 3b. Sides of angles are unequal in length; top segments are shorter than bottom segments.

One-Error Stimuli

Two-Error Stimuli

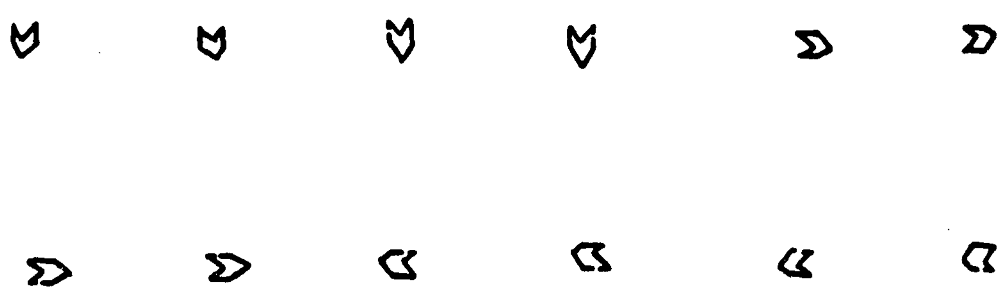


Figure 4. Correct Specifications and Specific Defects for the Second Inspection Figure.

Correct	Incorrect
1. Figure contains 8 dots. 	1a. Figure contains less than 8 dots. 1b. Figure contains more than 8 dots.
2. Dots are on the circle.	2a. Any dot inside the circle. 2b. Any dot outside the circle.
3. Complete circle figure.	3a. No gaps or holes in bottom half of circle. 3b. No gaps or holes in top half of circle.

One-Error Stimuli

Two-Error Stimuli

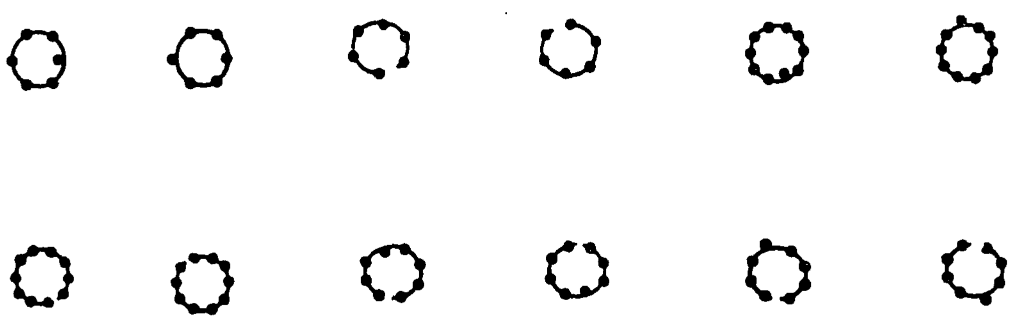
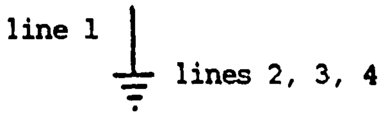


Figure 5. Correct Specifications and Specific Defects for the Third Inspection Figure.

Correct	Incorrect
1. Lines 2, 3, and 4 are below line 1.	1a. Lines 2, 3, and 4 are above line 1. 1b. Lines 2, 3, and 4 are to the right of line 1.
	
2. Three lines (and a dot) below line 1.	2a. Two lines (and a dot) below line 1. 2b. Four lines (and a dot) below line 1.
3. Line 1 is longer than line 2.	3a. Line 1 is the same length as line 2. 3b. Line 1 is shorter than line 2.

One-Error Stimuli

Two-Error Stimuli

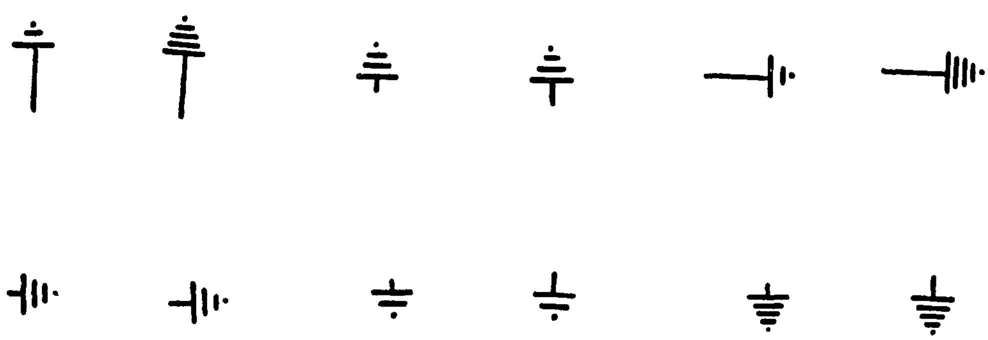


Figure 6. Correct Specifications and Specific Defects for the Fourth Inspection Figure.

the components of a complex inspection task were present, including several figures, figures closely arranged to one another, several types of errors, and varying degree of error severity.

Independent Variable

Two independent variables were applied, inspection sessions and training programs. Each subject repeated the inspection task on three separate occasions. In addition, three types of training programs were compared. All three programs included presentation of a correct sample of each figure and a verbal description of correct and incorrect attributes for each of the four stimuli. The three programs differed only in terms of non-examples presented.

The CSO Group was not shown non-examples during training; they were shown only a correct sample and a verbal description of correct and incorrect stimulus elements. To ensure that subjects made contact with the training materials, they were asked to draw the figures prior to beginning the actual inspection task.

Subjects in the CritDiff Group were shown non-examples where only one stimulus element at a time was incorrect. A separate training packet was used for each of the four stimuli, and subjects received these packets one at a time.

In each training packet, each error variation was described and shown on a separate page. A self-test was provided for each error variation; subjects designated defective stimuli and then compared their answers with the correct answers on the following page. When subjects completed a training packet, they turned in the materials and completed a quiz on that particular stimulus. Subjects were required to score at least 90% before proceeding to the next training packet. If they scored less than 90%, they received another training packet on the same stimulus and took another quiz.

Subjects in the MultDiff Group were shown non-examples containing two defects. Procedures were similar to those utilized for the CritDiff Group. Four separate training packets were used, each followed by a quiz. Because only two-error stimuli were shown, some differences occurred in the manner in which the figures, descriptions, and quizzes were presented. Each page contained verbal descriptions of all stimuli, i.e., the descriptions of all correct and incorrect stimuli. Two incorrect figures were shown on each page. Each packet contained a self-quiz and answer sheet at the end, following all the presentations of incorrect samples. Appendices A, B, and C show representative pages from each of the three training programs.

Dependent Variable

The dependent variables were the various types of inspection errors that could be made: (a) false positives, correct figures which were marked as incorrect; and (b) misses, errors which were missed. The total number of misses were further analyzed as one-error and two-error misses, stimuli containing one or two errors respectively.

Inter-Observer Agreement

Inter-observer agreement was assessed by having a second independent observer score 33% of the inspection task sheets (one inspection task per subject, with the session randomized). Percentage agreement was computed as follows:

$$\frac{\#\ \text{Agreements}}{\#\ \text{Agreements} + \#\ \text{Disagreements}} \times 100$$

An agreement was recorded when both observers scored an item as either a miss or a false positive; a disagreement was recorded when any item was scored differently by the two observers.

Procedures

Subjects in all three conditions participated in two phases: (1) training and (2) inspection. All training and testing sessions were conducted in two offices adjacent to one another. The experimenter utilized a third office where materials were distributed; subjects could be observed from this third office. Each office contained sufficient working space for two subjects, and desks were placed in such a way so that subjects could not observe each others' materials.

Subjects were allowed to select session times, with the requirement that sessions had to be separated by a minimum of 48 hours. Subjects generally scheduled all inspection sessions within two to four days of each other.

Training

During a single training session, subjects received training packets according to their group assignment; the order of the figure presentation was the same for all groups. A checklist was attached to each packet, with specific instructions regarding required activities. The instructional materials have been described previously; particular procedures are described below.

1. The CSO Group received one training packet which

contained information on each of the four stimuli, each figure having its own page. Written instructions told subjects to study the figures and descriptions as long as they wanted, to draw each figure at the bottom of the page, and to turn in all materials when they were finished. At that point, the experimenter verified that the four figures had been drawn, and presented the first inspection task to the subject.

2. The CritDiff Group and MultDiff Groups received four training packets. The differences in the contents of the packets have been described previously; the administration of the training was similar. Subjects received one packet at a time, and written instructions directed subjects to study the materials and complete the self-test(s). After subjects turned in the materials and the experimenter verified that the self-test(s) had been completed, subjects received the quiz for that figure. If the quiz score was at least 90%, subjects were given the next training packet. If the quiz score was below 90%, subjects were asked to complete the same training packet, followed by a different quiz. When all four training packets had been completed, subjects were given their first inspection task.

If subjects had questions or comments, they were told to perform as well as they could according to the instructions in training materials they received.

Inspection Task

Subjects completed three inspection tasks; the first task occurred immediately following training. During each session, subjects inspected 20 sheets with 80 figures per sheet. All subjects received the same inspection task, and the inspection task remained the same for all three inspection sessions.

Written directions for the inspection were provided at the beginning of each session. Subjects were told to place an "X" over defective items, which could contain one or more errors. When they had completed a page, they were instructed to place their initials at the bottom of the page, to designate that they had inspected that page. They were also told not to return to previous pages. No time limit to complete the task was imposed, but subjects were directed to complete the task as quickly and accurately as possible. After their final inspection session, subjects completed a questionnaire (Appendix D) regarding the training and inspection activities. One section of the questionnaire asked subjects to rate activities according

to level of difficulty, interest, etc. The final question was open-ended and asked subjects to explain any strategy that they used during the inspection task.

After the initial training session, subjects had no additional contact with the instructional materials. In other words, there was no review of materials prior to subsequent inspection sessions.

RESULTS

All subjects completed all sessions. With the exception of one subject, all subjects met criterion on their first try for quizzes. One subject had to repeat the training on only one figure. No difference in scores was observed based on the sex of the subject ($\underline{F}(1,36)=0.27$, $\underline{p}>.05$).

Interobserver Agreement

Overall interobserver agreement was 96%: 95% for the first session, 97% for the second session, and 99% for the third session.

Misses

A split-plot analysis of variance (Huitema, 1980) was used to determine the effects of Training Program, Sessions, and Training Program x Sessions. The split-plot analysis of variance was used because the Sessions factor was repeated three times for each subject.

While the effects of Training alone failed to reach statistical significance, significant effects for Sessions for all types of misses and a significant Training x Sessions effect for one-error misses were observed. The results indicated that the CritDiff Group maintained low error rates across all three sessions. Error rates for the

MultDiff and CSO Groups were higher than for the CritDiff Group during initial sessions and decreased across sessions, but the improvement was more pronounced with two-error stimuli than with one-error stimuli.

Total Misses

Total misses include the total number of defective items which subjects failed to detect. The Sessions effect was statistically significant ($F(2,4)=8.03$, $p<.05$). Post hoc analysis using the Least Significant Differences (LSD) test indicated that mean performance in session one differed significantly from that of session two ($t(29)=3.09$, $p<.01$) and that mean performance in session one differed significantly from that of session three ($t(29)=3.77$, $p<.001$). No difference was observed between mean performance in session two and mean performance in session three. Figure 7 shows the average number of total misses by group for each session (Appendix E provides summary data). Misses remained virtually unchanged for the CritDiff Group across sessions. Misses decreased for the MultDiff and CSO Groups across sessions, but were still higher for these two groups than for the CritDiff Group, even after three sessions.

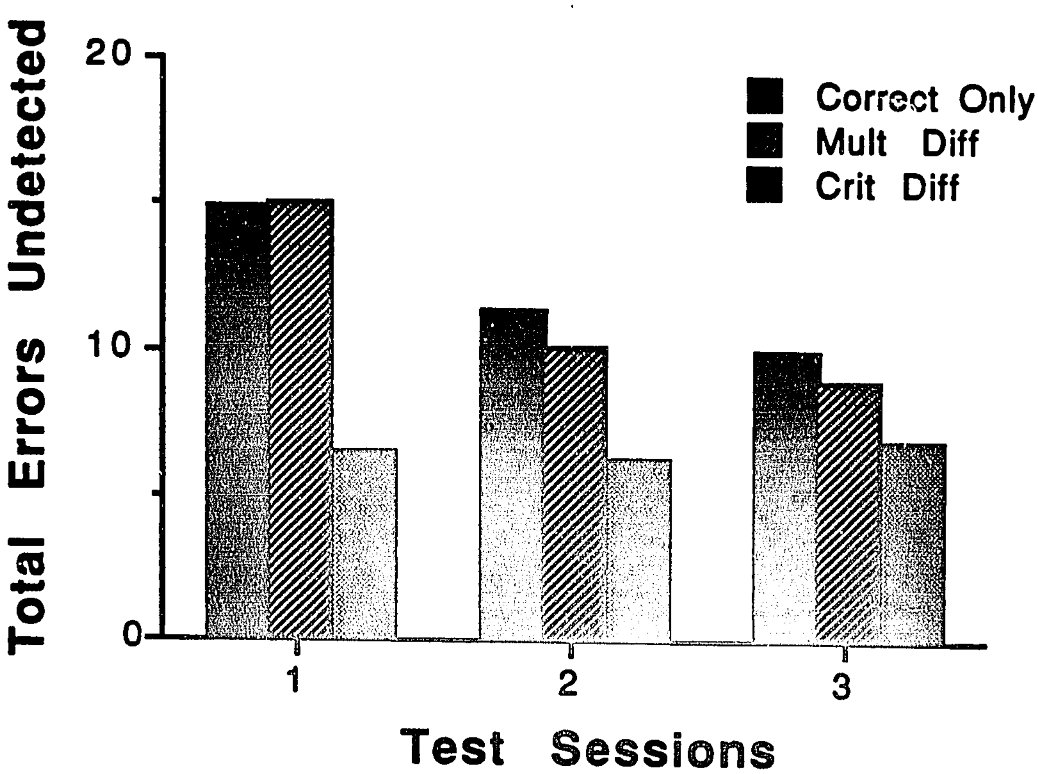


Figure 7. Mean Number of Total Errors Undetected for Each Inspection Session for Each Training Group.

Two-Error Misses

Two-error misses are failures to detect stimuli containing two errors. The Sessions effect was statistically significant (*F*(2,4)=5.42, *p*<.05). Post hoc analysis using the LSD test indicated that mean performance in session one differed significantly from that of session two (*t*(29)=2.13, *p*<.05) and that mean performance in session one differed significantly from that of session three (*t*(29)=3.32, *p*<.01). No difference was observed between mean performance in session two and mean performance in session three. Figure 8 shows the average number of two-error misses by group for each session (Appendix E provides summary data). Two-error misses remained virtually unchanged for the CritDiff Group across sessions. Two-error misses decreased for the MultDiff and CSO Groups across sessions; by the third session, the two-error misses for these groups were only slightly higher than for the CritDiff Group.

One-Error Misses

One-error misses are failures to detect stimuli containing only one error. The Sessions effect was statistically significant (*F*(2,4)=6.91, *p*<.05). Post hoc

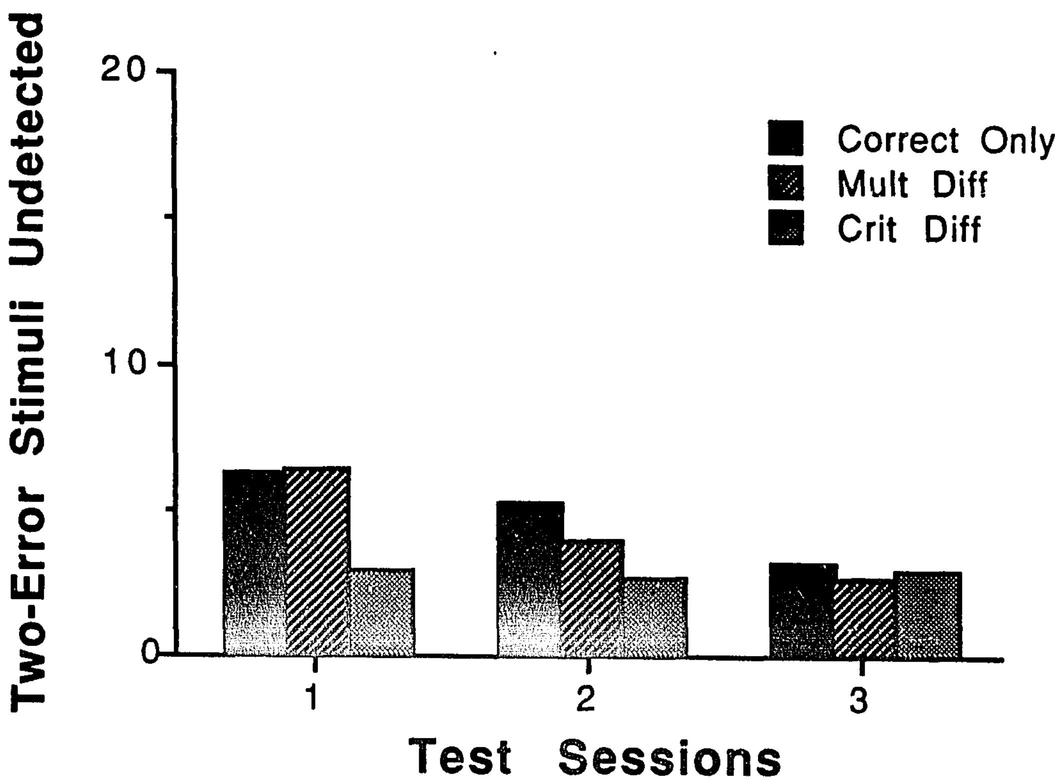


Figure 8. Mean Number of Two-Error Stimuli Undetected for Each Inspection Session for Each Training Group.

analysis using the LSD test indicated that mean performance in session one differed significantly from that of session two ($t(29)=3.32$, $p<.01$) and that mean performance in session one differed significantly from that of session three ($t(29)=3.25$, $p<.01$). No difference in mean performance was observed between session two and session three.

The Training x Sessions effect was also statistically significant ($F(4,70)=2.54$, $p<.05$), indicating that the difference in performance among groups was statistically significant in some sessions, but not in others. Subsequent analyses were done by comparing the mean performance levels of the three groups for each of the performance sessions individually using a Between Groups Analysis of Variance. Group performances did not differ significantly in session two or three. However, a significant difference was observed in session one ($F(2, 35)=3.86$, $p<.05$). Post hoc analysis using the LSD test to assess mean differences within session one produced no evidence of significant effects. Figure 9 shows the average number of one-error misses by group for each session (Appendix E provides summary data). One-error misses remained virtually unchanged for the CritDiff Group. While one-error misses declined somewhat for the MultDiff

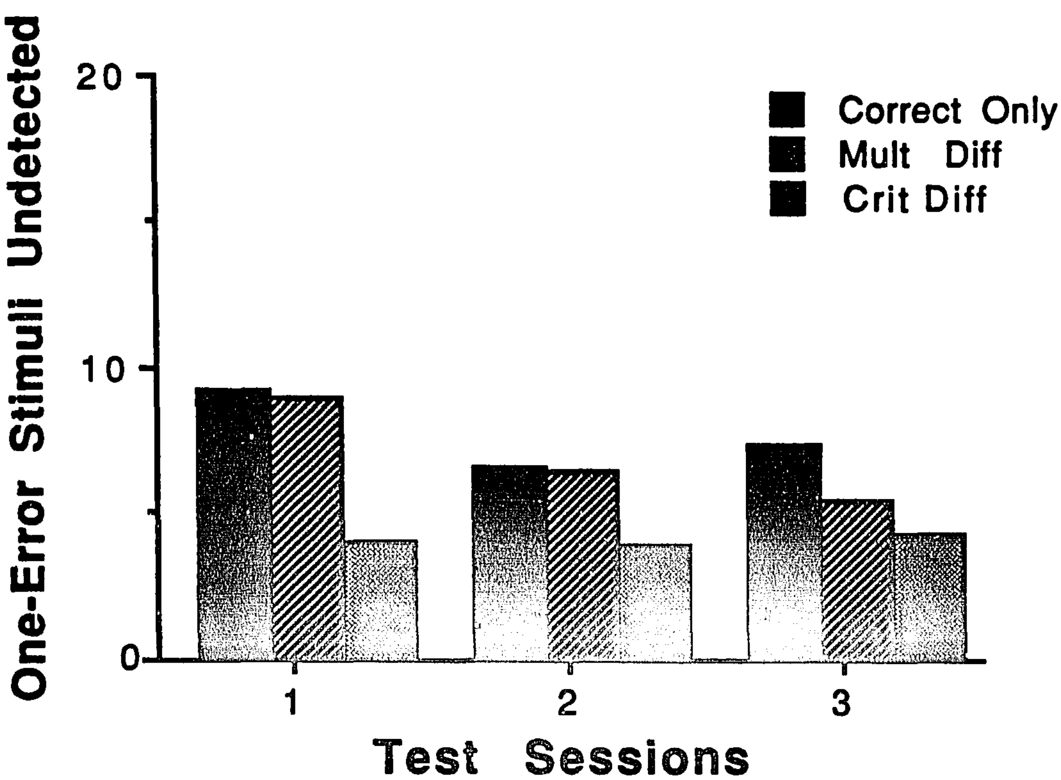


Figure 9. Mean Number of One-Error Stimuli Undetected for Each Inspection Session for Each Training Group.

and CSO Groups, the pattern was different from the pattern exhibited with two-error misses for these two groups. With one-error misses, there was a larger initial difference between the CritDiff Group and the MultDiff and CSO Groups. In addition, the decrease in one-error misses was not as great as for two-error misses for the MultDiff and CSO Groups.

False Positives

False positives occurred when a figure was reported as defective when the figure did not contain an error. No significant differences between groups were observed. Figure 10 shows the average number of false positives by group for each session (Appendix E provides summary data). While the CSO Group average was considerably higher than the other two groups, the difference was almost entirely attributable to two subjects with very high rates of false positives, 290 false positives for one subject and 635 false positives for the other subject.

Subject Questionnaires

Questionnaires (Appendix D) were obtained from all but one subject. Most subjects rated the activities very highly, and no differences between groups were observed

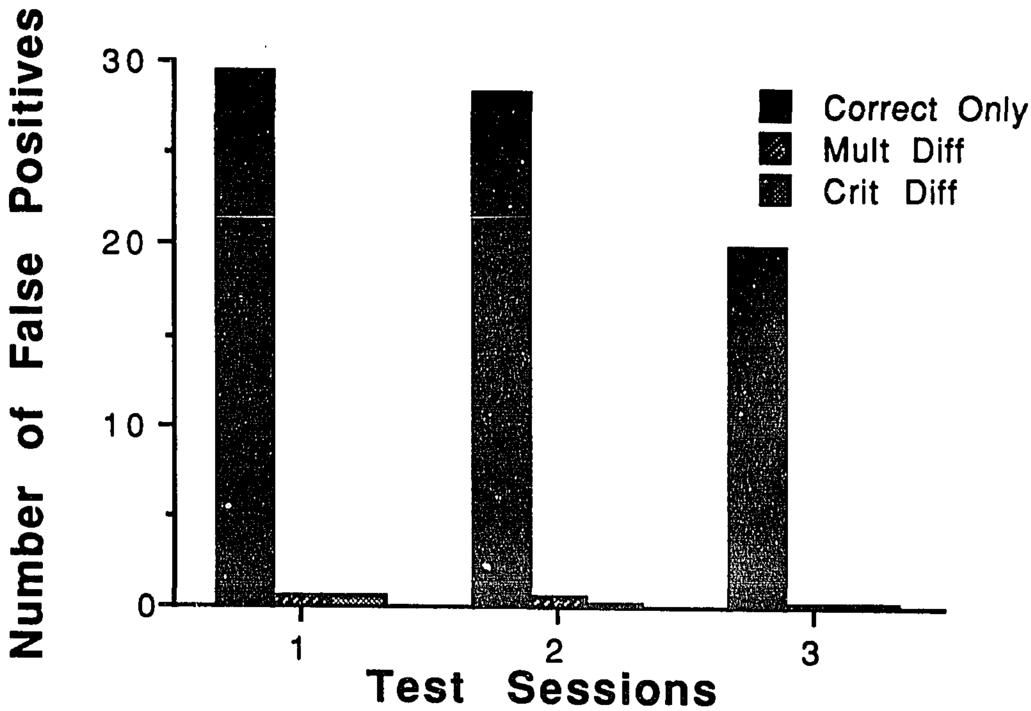


Figure 10. Mean Number of False Positive Errors for Each Inspection Session for Each Training Group.

On the final question, which asked subjects to explain any strategy that they used during the inspection task, responses fell into four distinct categories: (1) inspection by figure type, where subjects examined one type of figure before proceeding to the next type; (2) scanning inspection, where subjects scanned pages from left to right, row by row, etc.; (3) no inspection strategy, including such descriptions as remembering training or looking carefully; and (4) another strategy, where subjects indicated a systematic inspection method unique to them. Table 1 summarizes responses to the open-ended question.

Differences in strategies adopted by subjects were noted as a function of the type of training received. At least half of the subjects in the CritDiff and MultDiff Groups reported inspecting by figure type. Only two subjects in the CSO group reported inspecting by figure type; the other subjects in this group either reported scanning (five subjects) or reported no strategy (six subjects).

Table 1

Number of Subjects Reporting Specific Inspection Strategies According to Training Group Assignment

Inspection Strategy	Training Group		
	CSO	MultDiff	CritDiff
Figure Type	2	6	7
Scanning	5	2	2
Another strategy	6	1	2
No strategy	0	3	1

DISCUSSION

The results indicated that selection of training stimuli may be an important component in accurate inspection performance. Performance trends indicated that the CritDiff Group maintained higher rates of inspection accuracy across the three inspection sessions than the CSO and MultDiff Groups. Furthermore, the differences were most evident with one-error stimuli, where the performance of the CritDiff Group was significantly better than that of the other two groups during the first test session following training. Although significant group differences dissipated with practice across test sessions (i.e., group performance differences were not statistically significant during test sessions two or three), the relative accuracy of the CritDiff Group was greater than the other groups throughout the study.

Several controls utilized during the study support the conclusion that differences in inspection accuracy can be attributed to selection of training stimuli. Because a training criterion was adopted for each group, it cannot be argued that subjects in different groups performed differently because of different levels of mastery during the training component. Nor can it be argued that subjects simply performed better on only those stimuli to which they

had been exposed during training. If this were the case, not only would subjects in the MultDiff Group have trouble with one-error stimuli, but subjects in the CritDiff Group would have trouble with two-error stimuli. This clearly was not the case for the CritDiff Group.

While the reliability of self-reports can be questioned (Perone, 1988), it is interesting to compare the self-reports of inspection strategies reported by the three groups. Over half of the subjects in both the CritDiff and MultDiff Groups reported using an inspection by figure-type approach, which has been demonstrated to be more effective than other inspection methods (Harris & Chaney, 1969). The fact that subjects in the CritDiff Group performed better than those in the MultDiff Group suggests that inspection strategy alone cannot overcome the effects of training stimuli selection.

The fact that the inspection performance of subjects in the CSO and MultDiff Groups improved across sessions is interesting, particularly given the fact that subjects were not told the results of their performance, or allowed to review training materials. One area for additional research would be whether verbal descriptions assisted subjects in the CSO and MultDiff groups in improving their performance. Because the same verbal descriptions were

provided to all subjects, verbal descriptions alone are obviously not sufficient to provide initial optimum inspection performance. However, the fact that performance in the CSO and MultDiff Groups improved suggests that there may have been increased reliance on the rules provided as the sessions progressed. It would be interesting to compare inspection performance with and without verbal descriptions, to determine more precisely the contribution of fairly specific rules to inspection accuracy.

High false positive rates for two subjects in the CSO Group were not unexpected. Because they had not been exposed to non-examples, even miniscule changes in the appearance of figures caused some subjects to mark a figure as defective. Most of the false positives occurred with the circle figure (Figure 5). Depending on its placement on the page, some circles had a slightly more "flattened" look than others. This feature was not identified in the training materials as an error, nor did subjects in the other two groups mark figures as defective based on this feature.

The fact that subjects in the CSO Group were not exposed to non-examples during training increased the probability that control by irrelevant stimuli would develop (Mackintosh, 1977). Had non-examples been

presented, the likelihood that the irrelevant stimuli would have been present in some non-examples and absent in other non-examples would have decreased the correlation between any irrelevant stimuli and responding based on their presence or absence. Once again, the stimuli presented during training appear to make a difference in the types of errors made.

Some of the quality control literature explains errors by referring to types of criteria which subjects develop (Fox & Haslegrave, 1969; Schoonard et al., 1973). Subjects with high rates of false positives are said to develop a strict criterion; those with high rates of misses are said to develop a lax criterion. The present experiment suggests how the particular stimuli used during training may affect the development of different criteria. A lax criterion may indicate that subjects are not under the control of all relevant stimulus features. A strict criterion may indicate that subjects are controlled by stimulus features other than those, or in addition to those, specified by the experimenter.

While the discussion thus far has focused on the differences between the three groups in the present experiment, it is interesting to note that all subjects in this experiment performed considerably better than subjects

in another study using the same figures. Harris (1968) provided subjects with sample defects; subjects could also "ask for clarification on any item about which they were undecided" (p. 378). For a nine percent defect rate (used in the present experiment), Harris' graph shows a miss rate of approximately 25% and a false positive rate of approximately 15%. In the present study, with the exception of the two subjects in the CSO Group with high rates of false positives, no subject made anywhere near the percentages of errors reported by Harris, even in the first inspection session. Because Harris did not report specific training procedures, one can only speculate as to the causes for the performance differences between the two experiments. However, the procedures used by Harris did not require that subjects meet a criterion for initial learning. Thus, while the present study compared three different methods of training, the Harris study may demonstrate performance on this task when subjects are merely exposed to the figures, without being required to meet a performance criterion prior to the inspection task.

The present study also is consistent with the results reported by Allen and Fuqua (1985) using mentally retarded subjects. In both experiments, training with one-error stimuli was more effective in producing accurate responding

than training with two-error stimuli. The fact that subjects in the present experiment could read allowed training to proceed very rapidly, but descriptions alone were not sufficient to produce accurate responding in many instances.

The results of the present experiment also suggest a specific method to improve inspector training: specify relevant attributes and vary them one at a time. This approach emphasizes the specific stimulus characteristics of selected samples and not time spent in training or the total number of training samples inspected.

In a time of increased concern over quality, this training approach suggests one cost-effective solution for improving individual inspector accuracy.

Appendix A

Sample Page Used in Training Subjects in the Correct Sample Only (CSO) Group

Correct	Incorrect
1. The angles point left. 	1a. The angles point down. 1b. The angles point right.
2. The figure is completely enclosed, a continuous line. 	2a. Gaps or holes in the top half of the figure. 2b. Gaps or holes in the bottom half of the figure.
3. The sides of the angles are equal in length. 	3a. Sides of angles are unequal in length; the top segments are longer than the bottom segments. 3b. Sides of angles are unequal in length; the top segments are shorter than the bottom segments.

In the space provided, please draw the correct figure. If you draw the figure more than once, please circle the one which you consider to be correct.

Appendix B

Sample Page Used in Training Subjects in the
Multiple Differences (MultDiff) Group

<u>Correct</u>

1. The angles point left.

2. The figure is completely enclosed, a continuous line.

3. The sides of the angles are equal in length.

<u>Incorrect</u>

1a. The angles point down.

1b. The angles point right.

2a. Gaps or holes in the top half of the figure.

2b. Gaps or holes in the bottom half of the figure.

3a. Sides of angles are unequal in length; the top segments are longer than the bottom segments.

3b. Sides of angles are unequal in length; the top segments are shorter than the bottom segments.

Here are some examples of incorrect figures:

Appendix C

Sample Page Used in Training Subjects in the Critical Differences (CritDiff) Group

<u>Correct</u>

1. The angles point left.

<u>Incorrect</u>

1a. The angles point down.

The next page contains a self-test. When you are ready, turn the page and complete the self-test.

Appendix D

Questionnaire Administered to Subjects Upon Completion of the Experiment

Inspector Training Questionnaire

1. The TRAINING (the materials you received before the inspection task):

1	2	3	4	5
complicated				simple

1	2	3	4	5
boring				interesting

1	2	3	4	5
confusing				clear

1	2	3	4	5
useless				helpful

2. The INSPECTION TASK (the grey pages you examined for errors):

1	2	3	4	5
difficult				easy

3. Overall, participating in the study was:

1	2	3	4	5
boring				interesting

4. I thought the training was:

5. I thought the inspection task was:

6. I thought the purpose of the study was:

7. The incentive ($20) I received for participating in this research : a) is about the right amount b) is too much money c) is too little money.

8. Please describe briefly how you accomplished the inspection task - any strategy or system you used which helped you.

Appendix E

Mean (Mn) Number of Errors and Standard Deviations (SD) for the Three Training Groups

Mean (Mn) Number of Errors and Standard Deviations (SD) for the Three Training Groups

Group Error Type	Test Sessions 1		2		3	
	Mn	SD	Mn	SD	Mn	SD
Correct Only						
Total Errors	14.85	10.48	11.23	7.97	9.85	6.97
One-Stimulus	8.77	5.80	6.15	4.08	6.85	3.93
Two-Stimuli	6.08	5.02	5.08	4.72	3.00	3.34
False Positives	29.38	61.98	28.31	68.59	19.92	56.78
Multiple Differences						
Total Errors	14.92	10.05	9.92	6.97	8.85	6.57
One-Stimulus	8.69	5.76	6.08	4.03	5.08	3.52
Two-Stimuli	6.23	5.31	3.85	3.69	3.77	3.47
False Positives	0.62	0.77	0.62	1.45	0.15	0.38
Critical Differences						
Total Errors	6.42	5.66	6.17	7.51	6.75	7.14
One-Stimulus	3.67	3.73	3.58	4.21	4.00	4.39
Two-Stimuli	2.75	2.38	2.58	3.73	2.75	3.08
False Positives	0.58	0.90	0.25	0.62	0.25	0.45

Appendix F

Approval Letter From the Human Subjects Institutional Review Board

Human Subjects Institutional Review Board

Kalamazoo, Michigan 49008-3899

WESTERN MICHIGAN UNIVERSITY

Date: March 21, 1990

To: Cathy L. Thorne

From: Mary Anne Bunda, Chair Mary Anne Bunda

This letter will serve as confirmation that your research protocol, "The Effects of Training Methods on Inspector Accuracy", has been approved as expedited by the HSIRB. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the approval application.

You must seek reapproval for any change in this design. You must also seek reapproval if the project extends beyond the termination date.

The Board wishes you success in the pursuit of your research goals.

xc: W. Redmon, Psychology

HSIRB Project Number 90-03-01

Approval Termination March 21, 1991

BIBLIOGRAPHY

Allen, K.D., & Fuqua, R.W. (1985). Eliminating selective stimulus control: A comparison of two procedures for teaching mentally retarded children to respond to compound stimuli. Journal of Experimental Child Psychology, 39, 55-71.

Badalamente, R.V., & Ayoub, M.M. (1969). A behavioral analysis of an assembly line inspection task. Human Factors, 11, 339-352.

Baker, C.A., Morris, D.F., & Steedman, W.C. (1960). Target recognition on complex displays. Human Factors, 2, 51-61.

Chaney, F.B., & Teel, K.S. (1967). Improving inspector performance through training and visual aids. Journal of Applied Psychology, 51, 311-315.

Crosby, P.B. (1984). Quality without tears. New York: McGraw-Hill.

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

Drury, C.G. (1978). Integrating human factors models into statistical quality control. Human Factors, 20, 561-572.

Drury, C.G., & Addison, J.L. (1973). An industrial study of the effects of feedback and fault density on inspection accuracy. Ergonomics, 16, 159-169.

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.

Drury, C.G., & Sinclair, M.A. (1983). Human and machine performance in an inspection task. Human Factors, 25, 391-399.

Fortune, B.D. (1975). An experimental investigation of the effect of selected vigilance variables on inspection accuracy in a visually aided scanning task (Doctoral dissertation, University of Arkansas, 1975). Dissertation Abstracts International, 36, 3846A.

Fox, J.G., & Haslegrave, C.M. (1969). Industrial inspection efficiency and the probablity of a defect occurring. Ergonomics, 12, 713-721.

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

Geyer, L.H., Patel, S., & Perry, R.R. (1979). Detectability of multiple flaws. Human Factors, 21, 7-12.

Harris, D.H. (1964). Development and validation of an aptitude test for inspectors of electronic equipment. Journal of Industrial Psychology, 2, 29-35.

Harris, D. (1966). Effect of equipment complexity on inspection performance. Journal of Applied Psychology, 50, 236-237.

Harris, D.H. (1968). Effect of defect rate on inspection accuracy. Journal of Applied Psychology, 52, 377-379.

Harris, D.H., & Chaney, F.B. (1969). Human factors in quality assurance. New York: John Wiley.

Howell, W.C., Johnston, W.A., & Goldstein, I.L. (1966). Complex monitoring and its relation to the classical problem of vigilance. Organizational Behavior and Human Performance, 1, 129-150.

Huitema, B.E. (1980). The analysis of covariance and alternatives. New York: John Wiley.

Kibler, A.W. (1965). The relevance of vigilance research to aerospace monitoring tasks. Human Factors, 7, 91-99.

Levi, A.A., & Mainstone, L.E. (1987). Fundamentals of statistical process control. Journal of Organizational Behavior Management, 9, 5-21.

Liuzzo, J.G., & Drury, C.G. (1980). An evaluation of blink inspection. Human Factors, 22, 201-210.

Mackie, R.R. (1987). Vigilance research-Are we ready for countermeasures? Human Factors, 29, 707-723.

Mackintosh, N.J. (1977). Stimulus control, attentional factors. In W.K. Honig & J.E.R. Staddon (Eds.), Handbook of Operant Behavior (pp. 481-513). Englewood Cliffs, NJ: Prentice-Hall.

Mainstone, L.E., & Levi, A.S. (1987). Obstacles to understanding and using statistical process control as a productivity improvement approach. Journal of Organizational Behavior Management, 9, 23-32.

Monk, T.H., & Brown, B. (1975). The effect of target surround density on visual search performance. Human Factors, 17, 356-360.

Perone, M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Use and abuse of subjects' verbal reports. The Behavior Analyst, 11, 71-75.

Schoonard, J.W., Gould, J.D., & Miller, L.A. (1973). Studies of visual inspection. Ergonomics, 16, 365-379.

Smith, R.L., & Lucaccini, L.F. (1969). Vigilance research: Its application to industrial problems. Human Factors, 11, 149-156.

Teel, K.S., Springer, R.M., & Sadler, E.E. (1968). Assembly and inspection of microelectronic systems. Human Factors, 10, 217-224.

Wiener, E.L. (1987). Application of vigilance research: Rare, medium, or well done? Human Factors, 29, 725-736.

EFFECTS OF TRAINING METHOD ON INSPECTOR ACCURACY IN A SIMULATED QUALITY CONTROL TASK

Cathy L. Thorne, Ph.D.

Western Michigan University, 1991

This study compared the effects of three training programs on inspector accuracy. The inspection task required 38 subjects to detect errors in geometric figures. All three training programs provided the same written descriptions of correct and incorrect figures, and all three training programs provided examples of correct figures. What differed among the programs was the type of incorrect figures. Individuals assigned to the Correct Sample Only Group were not shown any incorrect figures. Individuals assigned to the Critical Differences Group were shown incorrect figures which contained only one error at a time. Individuals assigned to the Multiple Differences Group were shown incorrect figures which contained more than one error at a time.

Following training, all participants inspected the same sheets of figures, on three separate occasions. Individuals in the Critical Differences Group had very high rates of inspection accuracy across all three inspection sessions. Inspection accuracy for individuals in the other two groups improved across the three inspection sessions, but was still not as accurate as for individuals in the

Critical Differences Group. This was particularly true for figures containing only one error, which were less obvious and harder to detect.

Earlier research demonstrated that selection of incorrect figures shown during training plays an important role in accuracy during testing, but the finding had not previously been applied in the context of inspector accuracy. The present research suggests a way in which inspector training programs can be improved, by specifying the types of errors which training figures contain.