AN EXPERIMENTAL INVESTIGATION OF INDUSTRIAL INSPECTOR ACCURACY UNDER VARYING LEVELS OF PRODUCT DEFECTIVENESS

By

PAUL MARK WALLACK

Bachelor of Science
University of Tulsa
Tulsa, Oklahoma
1950

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1956

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 1967

OKLAHOMA
STATE UNIVERSITY
LIBRARY

JAN 18 1968

AN EXPERIMENTAL INVESTIGATION OF INDUSTRIAL INSPECTOR ACCURACY UNDER VARYING LEVELS OF PRODUCT DEFECTIVENESS

Thesis Approved:

S. Keith Adams
Thesis Adviser

W. J. Bentley

James E. Shamblin

Wa Dreinbach

M. R. Lohmann

N. N. Durham
Dean of the Graduate College

660103

PREFACE

This research is concerned with the problem of accuracy of industrial electronics inspectors on a visual isnpection task. In conducting the study, it was necessary to develop a decision model for inspector accuracy in order to assure that satisfactory and complete measures of performance were used in reporting the experimental results. Reporting of inspector accuracy is usually treated in terms of the percentage of correct detections of non-conforming product without equal consideration for the number of times good product is correctly identified. This results in an incomplete picture of the total situation.

As the result of this study, I feel that the signal detection model is most suitable for analyzing problems of inspector accuracy. The model formalizes and unifies the elements of the inspector accuracy problem and, thus, provides a convenient framework within which the experimenter can work. It is hoped that the identification of the magnitude of error in correctly identifying conforming product and the treatment of visual inspection as a possible vigilance task will alert the professional quality control person to the significance of these elements and so give them more weight in inspection performance improvement efforts.

To acknowledge my indebtedness in full measure to each of the persons who has in some way given me support, encouragement, and counsel when these were needed during the long course of this study would require a volume as long as this dissertation. However, I would like to devote a few words of thanks and appreciation to each of the following: To my family for their faith, patience, and understanding; to Larry Snodgrass for continued encouragement; to A. Scott Crossfield and Luther Lawrence for their interest and support; to Carl Winfield, John Harris and their inspection team for participating in the experiment; and to my colleagues, Dave Graves, Don Kussee, H. R. LaPorte and Bob Haney who listened to my ideas. In particular, Dave Graves' assistance with the data collection in the pilot study was invaluable.

In addition to these, I would like to acknowledge sincere thanks to my advisory committee for their patience with my rate of progress over the term of my study program. In particular I appreciate the counsel and encouragement offered by Professors Wilson J. Bentley and Paul E. Torgersen during that time. Also, I appreciate very much the assumption of the role of thesis adviser by Dr. S. K. Adams and the valuable suggestions he has provided.

Thanks also are due Miss Velda Davis who provided typing and other assistance.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 1

The Inspection Task Model 2
Statement of the Problem 5
Importance of the Problem 6
Summary 9

II. INSPECTOR ACCURACY 11

Review of the Literature 13
Summary 19

III. VIGILANCE TASK CRITERIA 22

Vigilance Task Description 24
Vigilance Decrement Pattern 30
Summary 31

IV. SIGNAL DETECTION THEORY 32

Fundamental Detection Problem 33
Signal Detection Criteria 35
The Decision Matrix 41
Summary 43

V. PERFORMANCE CRITERIA 44

McCornack's Criteria 44
Bayesian Criteria 46
Bayesian Criteria Versus McCornack's Criteria 48
Receiver Operating Characteristic Curve 49
Apparent Fraction Defective 49
Summary 51

VI. EXPERIMENTAL DESIGN AND TECHNIQUE 53

Data Collection Model 54
Test Materials - Kind 56
Test Materials - Quantity 57
Inspectors Tested 60

Chapter Page

VI. (CONTINUED)

Test Schedule 61
Pretraining Schedule 62
Test Procedure 64
Data Analysis 66
Summary 71

VII. RESULTS AND ANALYSIS 72

Effect of p_o: McCornack's Criteria 72
Effect of p_o: Bayesian Criteria 76
Vigilance Performance: McCornack's Criteria 80
Vigilance Performance: Bayesian Criteria 83
Index of Detectability: d′ 85
Apparent Fraction Defective 88
Summary 91

VIII. TWO POPULATION ANALYSIS 93

Classical Method 93
Signal Detection Method 95
Summary 98

IX. SUMMARY AND CONCLUSIONS 99

Summary 99
Problems in Experimentation 102
Recommendations for Additional Study 105

BIBLIOGRAPHY 107

APPENDICES 110

A. TEST INSTRUCTIONS 111

B. INSPECTION DATA SHEET 125

C. SUMMARY OF PERFORMANCE MEASURES FOR SEVEN INSPECTORS 126

D. DATA SUMMARY FOR SEVEN INSPECTORS 127

E. CONTINGENCY TABLES 131

F. METHOD FOR CALCULATING SIGNIFICANCE OF REGRESSION 134

APPENDICES Page

G. NORMALIZED DEVIATION OF p_3 FROM p_0 FOR EACH TIME BLOCK 135

H. INSPECTOR ERROR EFFECTS ON SAMPLING DECISIONS 136

LIST OF TABLES

Table Page

I. Testing Schedule 63

II. Test Results: McCornack's Criteria 75

III. Tests of Hypotheses for Bayesian Criteria 78

IV. Operating Characteristic Curve Computations 136

V. Calculations of Probability of Acceptance for p_3 for Two Inspector Groups 139

LIST OF FIGURES

Figure Page

1. Inspection Task Model 3

2. Typical Vigilance Performance Decrements 23

3. Probability Density Functions of Observations 34

4. Critical Regions for Observations 36

5. Receiver Operating Characteristic Curve 40

6. Inspection Decision Matrix 42

7. Test Materials 58

8. Wire Preparation Inspection Training 65

9. McCornack's Criteria Versus p_o 73

10. Bayesian Criteria Versus p_o 77

11. McCornack's Criteria Versus Time Blocks 81

12. Bayesian Criteria Versus Time Blocks 84

13. d' and x_c for Each p_o 86

14. Normalized Deviations from p_o for Time Blocks 89

15. d' and x_c for Two Inspector Groups 96

16. Effect of Inspection Error on an O.C. Curve 137

17. O.C. Curves for Two Inspector Groups 140

CHAPTER I

INTRODUCTION

The purpose of this research was to study the performance of industrial inspectors both in detecting defects and in identifying acceptable product. As will be shown later, most research related to inspection performance considers defect detection to the complete exclusion of performance quality in product acceptance. The result is that the only inspection performance measure available to industrial management is the number of defects missed. This criterion does not fully establish the quality of inspector performance.

The experimental design used in this study provided for the collection of data pertinent to the ability of the industrial inspector to accept good product as well as to detect defects. As a result, the secondary task of the research became that of studying performance criteria which accommodate both of these elements of the inspection task and evaluating their utility to industrial management. A useful inspector performance criterion is one that not only enables the inspection supervisor to meet his responsibility for utilizing his personnel to the best possible advantage, but also is one which company executives can readily use in communicating with the customer concerning the quality of performance of the inspection function.

Before suitable performance criteria can be defined, the magnitude of inspection error, both in detecting defects and in identifying acceptable product, must be quantified for a range of inspection tasks. This research, which is concerned with a visual inspection task performed by electronics inspectors, is a start toward that objective.

The Inspection Task Model

A visual inspection situation may be modeled in the following manner. Visual inspection, as a processing task, is defined as the aided or unaided observation of details without measurement to determine the conformance and completeness of the part of finished product (18). The inspector is required to separate the fraction of the input that is nonconforming with respect to applicable specifications, drawings, and workmanship standards, and forward to the user only the conforming items (Figure 1). Thus, the process input can be represented as follows:

$$I = p_o I + q_o I \tag{1}$$

where:

I = total input

p_o = fraction of nonconforming items

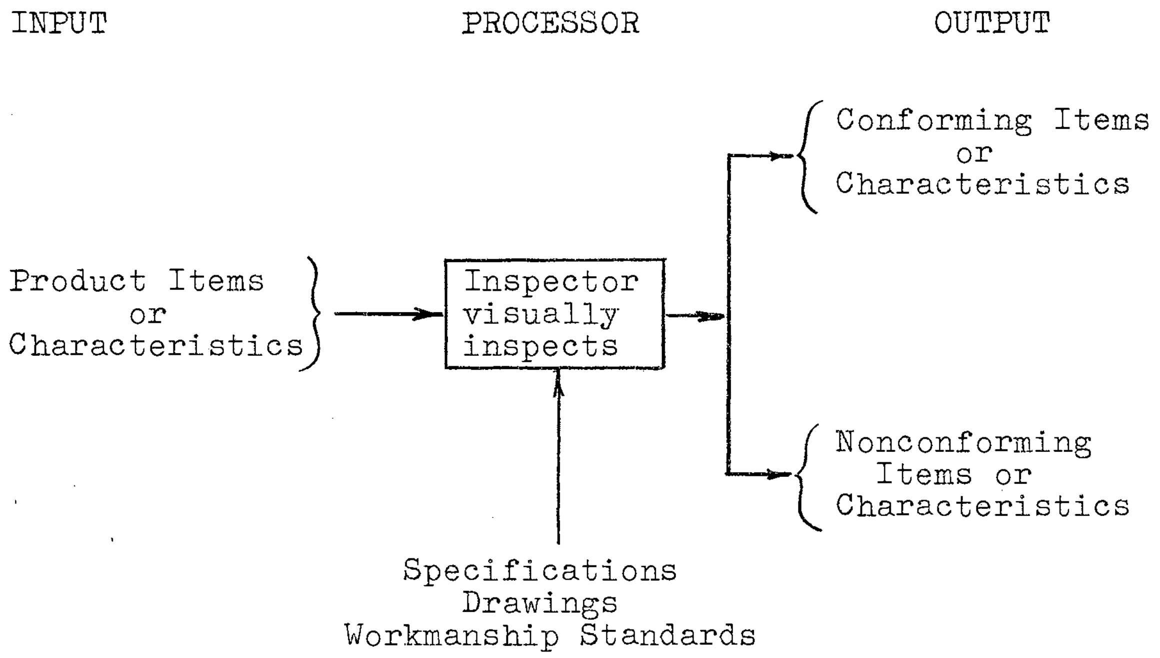


Figure 1. Inspection Task Model

q_0 = fraction of conforming items = $1 - p_0$.

In the inspection task it is desirable to identify correctly all the nonconforming items and segregate them from the conforming items. The inspector may err in this identification process in either or both of two ways: he may classify conforming items as being nonconforming, a Type I error; or he may classify nonconforming items as being conforming, a Type II error.

If p_1 is designated as the probability of a Type I error and p_2 is designated as the probability of a Type II error, then a probability matrix can be constructed for the inspection process as follows:

		Decision Based on Inspection		
		Accept	Reject	Total
Product Mix	Conforming	$(1-p_1)q_0$	$q_0(p_1)$	q_0
	Nonconforming	$p_2 p_0$	$(1-p_2)p_0$	p_0
	Total	$(1-p_1)q_0 + p_2 p_0$	$q_0 p_1 + p_0(1-p_2)$	1

where:

q_0 = a priori probability of conforming product in an inspection lot

p_0 = a priori probability of nonconforming product in an inspection lot = $(1 - q_0)$

$1 - p_1$ = probability of a correct acceptance decision for conforming product

$1 - p_2$ = probability of a correct reject decision for nonconforming product.

The products in each cell of the matrix are the joint probability elements of the conditional probability statements associated with each stimulus-response combination of the matrix. These statements are of the form

$$P(X|Y) = \frac{P(X \cdot Y)}{P(Y)}. \qquad (2)$$

The following conditional probability statements can be written for the probability matrix for inspector decisions:

$$P\ (\text{Accept}|\text{Conforming}) = \frac{(1 - p_1)q_0}{q_0} \qquad (3)$$

$$P\ (\text{Reject}|\text{Nonconforming}) = \frac{(1 - p_2)p_0}{p_0} \qquad (4)$$

$$P\ (\text{Reject}|\text{Conforming}) = \frac{q_0(p_1)}{q_0} \qquad (5)$$

$$P\ (\text{Accept}|\text{Nonconforming}) = \frac{p_0 p_2}{p_0} \qquad (6)$$

Statement of the Problem

Development of the probability matrix leads to identification of the basic questions whose answers are sought by this research.

The questions are as follows:

1. Does the a priori probability of a nonconforming item affect inspector performance? Stated another way, is inspector performance independent of the fraction defective of the product mix?
2. For a given range of a priori probabilities

of non-conformance, is there an inspector performance decrement over time that does not exist over another range of a priori probabilities of nonconformance?

3. What performance measures are most useful in describing the results obtained in answering questions one and two?

The balance of this chapter deals with the importance of these basic questions, while Chapter II reviews the literature on inspection accuracy. As question number two relates to the problem of vigilance performance decrement, Chapter III presents those aspects of the subject that relate to the inspection task. As the research problem relates to the theory of signal detection by human observers, Chapter IV discusses the inspection task in the framework of this theory. In Chapter V are presented a number of criteria found useful in describing the quality of performance of industrial inspectors. The experimental design for collecting and analyzing the data required for answering the basic questions is described in Chapter VI. Results of the experiment and their analysis are presented in Chapters VII and VIII. Chapter IX summarizes the research and suggests additional experiments which may aid in answering questions arising from this work.

Importance of the Problem

The area of concern here is that of industrial inspection wherein the inspector with high error rates can cause extensive damage by failure to identify defective material. If such material, known as "escapes" in industrial terminology, is used in subsequent assembly operations, it may cause serious damage to material or injury to personnel.

Although the technology of the industrial inspection has made rapid advances in the past few years, there still exists a large amount of waste resulting from inspector error. Inspection processes that seem to be associated with high error rates are those in which the inspector plays a passive role. A by-product of this situation is loss of attention or vigilance on the part of the inspector. Because of his loss of attention, he may fail to recognize nonconforming quality characteristics (Type II error) and may even fail to recognize conforming characteristics (Type I error).

Research in the area of industrial inspection has been minimal. What has been accomplished is, according to one author (22), characterized by incomplete reporting and inadequate experimental design. There has also been raised the question as to what constitutes a proper measure of accuracy for industrial inspectors.

The inspection problem is essentially the same as the signal detection problem in military operations and it appears that signal detection theory as applied to human observers has considerable merit in the examination of inspection accuracy. This subject is discussed in Chapter IV.

The developing theory of vigilance has a number of implications for the inspection problem cited. Psychologists studying problems of monitoring radar screens, control panels, etc., have found that, at best, man is a poor monitor under a variety of conditions (3). Vigilance in this sense denotes man's degree of attention to his monitoring task. Therefore, a vigilance task is one in which the observer of a sequence of signals attempts to detect the occurrence of an infrequent, "wanted" signal which for some reason must be identified. For instance, the "wanted" signal may be that one blip in a population of blips on a radar screen that designates an unfriendly aircraft.

A parallel may be drawn between the vigilance task and industrial inspection processes. In the case of critical quality characteristics, nonconforming items must be detected. If these characteristics are designated prior to production, then in-process controls may be established in an effort to minimize the production of nonconforming items. The result should be a low level of defectiveness (perhaps $p \leq 1.0\%$) in an inspection lot presented for examination. The inspector who has just inspected a long series (k) of these items may not observe that the ($k + 1$) item is nonconforming. The lack of a defect in the previous sequence of items whose characteristics conform to specification has resulted in a vigilance decrement such that he is not able to detect the existence of the nonconforming characteristic on item (k + 1). Vigilance research has revealed that the detection of a wanted signal serves as reinforcement and so stimulates the subject to maintain his initial performance level. It has also shown that the introduction of artificial wanted signals is significant in the maintenance of performance levels (8). This suggests that in the industrial inspection task there should be introduced into the product sequence a number of items with known nonconforming characteristics in an effort to maintain the initial performance level of the inspector so as to minimize the likelihood of a vigilance performance decrement.

Introduction of known defectives into the product inspection sequence is not new to industrial inspection. This effort was made in an attempt to gage an inspector's accuracy by comparing known defects found to the number introduced. However, there does not appear to have been any effort to introduce known defectives into an inspection lot for the purpose of manipulating the vigilance performance of the inspector.

Summary

In this chapter there has been presented a probability matrix for the inspection tasks, statements of the basic questions to be answered in this research, and an indication of the importance of the problem in the area of industrial inspection.

Chapter II contains a review of the literature relative to inspection accuracy, while Chapter III is devoted to the developing theory of vigilance performance.

CHAPTER II

INSPECTOR ACCURACY

As stated in Chapter I, this research concerns the accuracy of inspectors in an industrial environment. Preliminary to the review of related research, it is necessary to describe the industrial inspection process so that the problem of accuracy can be put in its proper perspective.

The basis for providing an inspection function in a manufacturing organization results from the fact that humans err and machines do not perform at a constant level. The industrial process requires a product design which is communicated to the manufacturer in the form of drawings and specifications. The design is translated to materials, processes, and skills and the product is made.

Before passing the completed item over to the ultimate user, it is necessary to assure that it conforms to design intent. Items not conforming are rejected and may be either reworked until they do conform or scrapped.

The inspection process consists of measuring, examining, testing, or otherwise comparing a unit of product with the process requirements. The comparison may be made either with or without the use of instruments. Instruments consist of one or more of a variety of gages, standard measuring devices, or special test equipment. A comparison without the use of instruments, generally referred to as visual inspection, requires testing by sight, touch, or manipulation. It is this latter class of inspection which is the concern of this research.

Thus, the inspection function exists because industrial management recognizes that nonconformances occur in manufacturing. Because machine processes are nonconstant and humans err, it is necessary to find and identify those errors in order to provide a product of satisfactory quality. In past years, management, although recognizing the potential error of the production worker, has failed to give adequate recognition to the error potential of the inspector. The advent of sampling plans and control charts did much to overcome this complete dependence on the inspector's accuracy. The philosophy of modern quality control is that 100 percent inspection is usually ineffective and expensive. Yet many organizations revert to 100 percent inspection at the start of a new design program. In the event of small or short product runs, screening all units of product may be the only way to assure that quality requirements are met in production. Regardless of how small the size of the run or how minimum the complexity of the product, inspection error is possible and, therefore, must be studied in greater depth in order to minimize potential costs. This research is designed to be a contribution to that study.

Review of the Literature

One of the earliest studies dealing with inspector-accuracy was that of Tifflin and Rogers in 1941 (28) wherein they were concerned with the accuracy of tinplate inspectors. In this study, 150 inspectors were each given 150 plates to examine, of which 61 were defective. Correct identifications averaged 78.5 percent with a range of from 55 to 96 percent.

In 1945 Lawshe and Tifflin (28) reported a study concerned with inspector-accuracy when reading various micrometers and calipers to establish tolerances. The percentage of inspectors reading within the required tolerances ranged from nine percent to 64 percent for 11 inspection tasks. Also, the experimenters found that the accuracy of reading a micrometer did not correlate with age, experience, or time on the present job.

Hayes (17) reported on the inspection of piston rings for surface defects (sand and gas holes from the foundry) with the result of 67 percent inspector-accuracy; i.e., 67 percent of the nonconforming items were classified as such. It is interesting to note that Hayes then sent the set of rings back through inspection, and the inspectors were led to believe that the lot had been reworked prior to being resubmitted. Accuracy then fell to 33 percent which indicates some considerable amount of loss of attentiveness on the part of the inspectors.

A study by Jacobsen (19) on 17 quality control inspectors thought to be 95 to 98 percent effective reported that these inspectors were only 80.5 percent accurate in identifying solder defects and ranged from 32 to 65 percent accurate in identifying wiring and appearance defects on a wired unit with built-in defects. In a second experiment (20) wherein 39 inspectors were tested on a similar wired unit with built-in defects, the result was 82.8 percent accuracy.

A number of other studies of inspector-accuracy are reported which indicate similar low levels of inspector-accuracy. For instance, Carter (4) reported on the accuracy of inspectors of acoustical tiles. There was a visual inspection for defects in fabrication, coating, cutting, drilling, or bevelling. The results showed the inspectors correctly identified 95 percent of the good tiles, but rejected only 76 percent of the defective tiles. Carter also pointed out that whereas the tiles were 86 percent good when received, they were only 96 percent good after being inspected, thus realizing only 71 percent of the maximum possible improvement. Kennedy (21) reported on four groups of inspectors performing both visual and gaging operations with accuracies for three groups classed as "regulars" found to be 68 percent, 56.25 percent, and 57.14 percent, and accuracy for one group classed as "expert" found to be 66.67 percent.

The cited studies were all concerned with industrial-type inspection of various kinds of product. The problem is not restricted to this environment, however. Adams (1) reported on the inspection of hard red winter wheat where wheat grain is compared to grain standards by experienced wheat inspectors. The results are reported as "... 40 percent of the initial estimates placed the sample improperly."

A critique of literature on inspector accuracy by McCornack in 1961 is of particular interest here (22). First, McCornack suggests that there are in effect four measures of performance that should be considered in evaluating inspectors. These he defined as:

A_1 = Percent of correct inspections.

A_2 = Percent of satisfactory product accepted.

A_3 = Percent of defective product rejected.

E_f = Efficiency in improving product.

The use of A_1 permits maximizing the total number of correct inspections and, thus, minimizing errors of misclassification. It assumes that all errors are equally important, and that both correct decisions are important.

The use of A_2 permits maximizing the probability of accepting satisfactory product, and so minimizes the probability of rejecting satisfactory products. In this case, it is assumed that cost of product is high and there exists little concern about accepting defective product; therefore, little emphasis is placed on Type II errors.

The measure A_3 permits maximizing the probability of rejecting defective product and, thus, minimizes the probability of accepting defective product. This measure concentrates on defects and ignores Type I errors. This is the measure of accuracy commonly used because the accurate identification of defects is the inspection problem as seen by the inspectors themselves.

The proposed accuracy measure E_f is defined by McCornack as the ratio of actual improvement achieved to the maximum possible improvement in percentage of satisfactory product. Use of this measure permits maximization of the probability that the accepted product is defective. Not only is this index sensitive to both Type I and Type II errors, but also it takes into account defects created in the product by the inspector himself.

All of the studies reviewed earlier in this chapter used the measure A_3 in reporting results, and only one of them considered the percent improvement in the product by the inspector.

The second item of importance to this study in McCornack's critique is his suggestion that some inspection tasks may well be treated as vigilance tasks in an effort to broaden the examination of the problem of "escapes." McCornack suggests that as man is a poor monitor, A_3 may drop to about 0.5 in a short time and remain at this level for extended periods.

One of McCornack's comments dealt with the lack of theory as to how an inspector does his job. The studies of Colquhoun (6, 7, 8), Thomas (27), and Seaborne (24), which seem to be an effort to overcome this deficiency, are summarized here because of their relation to the research problem.

Colquhoun defined inspection efficiency as the proportion of faults to which a response was made in a simple machine paced inspection task. He studied the effect of a short rest pause on efficiency and found that when work was uninterrupted for a period of one hour, efficiency, although high, declined after about 30 minutes. When a five minute rest was given at the end of 30 minutes, the efficiency level was maintained throughout the hour. The study was for a one hour period only and so did not simulate the workday situation. Colquhoun justifies this limitation as follows:

> It is not uncommon for checking to continue for periods of two or three hours at a stretch without an 'official' rest pause. However, although the flow of work may appear continuous, in the majority of factories, the time spent over the examination of individual items, or of particular parts of the material, can vary within quite wide limits, and this enables the inspector to take occasional brief rests during what might otherwise be an unbroken session of checking (6).

Colquhoun was also able to determine that differences in efficiency were unrelated to intelligence. In a subsequent experiment (7), he demonstrated that time of day and temperament had an effect on accuracy, but could not at the time of writing offer reasons for the observed effects.

Thomas (27) treats the perceptual organization of industrial inspectors and reports the following:

1. An inspector may require a number of months to build up to the accepted level of skill.

2. A difference exists between the formal description of the inspection task and the description built up by observation and discussion.

3. A thoroughly experienced inspector seems to acquire a picture of "good" product such that defects stand out; and in addition, he gets to know what to inspect in certain situations -- that is, he uses the cues available to him such as familiarity with product, process, and end use of product.

Seaborne (24) reported on the social effects of an inspection task with regard to accuracy. His results indicated that the proportions of objects rejected by a subject on the basis of measured characteristics would be affected by the perceived proportions of rejections made by other persons working near at hand. Seaborne demonstrated that group rejection norms developed in the laboratory as well as in the industrial situation and that the participants were not aware of making changes in their judging behavior.

The ongoing research program of Harris (10, 11, 12, 13, 14, 15, 16) has resulted in a number of findings which relate to this research. Of major interest in Harris' work is his consideration of the percentage of false detections made by inspectors. This percentage relates to McCornack's factor A_2 which has not been considered to any extent in other research. Harris' findings are summarized below:

1. Defect detection performance does not depend on: experience of the inspector, the number of false detections, or the time taken to perform the inspection.
2. Some inspectors have bias for certain types of defects.
3. Defect detection performance becomes worse as complexity of the product being inspected increases.
4. The average percentage of defects detected in ten studies ranged from 14 percent to 74 percent, and the average percentage of false detections ranged from 15 percent to 70 percent.

Summary

In this chapter there have been presented summaries of a number of studies relative to the accuracy of industrial inspectors along with a discussion of the measures of accuracy used. The following points were brought out in these studies:

1. Inspector-accuracy ranged from nine percent to 96 percent with an unweighted mean of 64 percent. The measure of accuracy used corresponds to McCornack's "A_3."
2. The percentage of false detections ranged from an average of 15 percent to 70 percent. The measure used relates to McCornack's "A_2."
3. There is no correlation between accuracy and

either age, experience, or time on present job.

4. So called "expert" inspectors did not fare better than "regular" inspectors on a selected task.
5. The belief that product was "good" when in fact it was not reduced the level of accuracy.
6. A rest pause after the first 30 minutes of an inspection task seems to enhance the inspector's efficiency for the next 30 minutes.
7. An inspector continually adjusts his discrimination as regards the inspection process and uses whatever cues are available in performing his task.
8. In a group of inspectors, one individual's rate of accuracy may be affected by his perceived proportion of rejections made by his associates.
9. Defect detection performance, while not dependent upon the number of false detections or upon the time taken for the inspection task, does appear to become worse as complexity of the items being inspected increases.
10. Some inspectors have a habitual bias for

certain types of defects.

The next chapter will contain a brief discussion of vigilance task criteria along with some aspects of the vigilance concept that have bearing on this research.

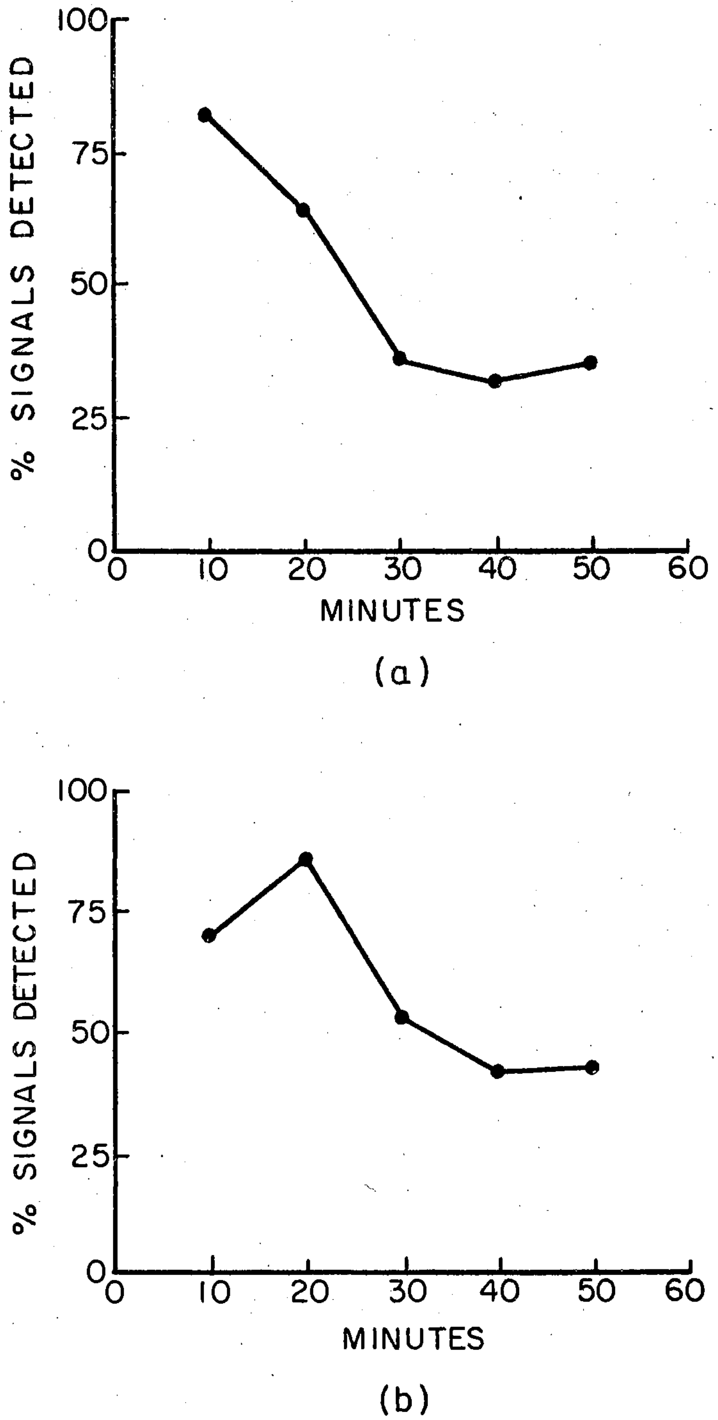
CHAPTER III

VIGILANCE TASK CRITERIA

Vigilance is defined as a state of alertness in responding to a stimulus (9). Hence, a vigilance task is one which requires an individual to be alert for a particular infrequently appearing stimulus.

The industrial inspector who must distinguish between units just within and/or just beyond the tolerance limit of acceptability is performing a vigilance task in that he may be making a visual determination of an infrequently appearing characteristic. Vigilance research (3) has indicated that alertness of the individual performing a task tends to drop off during the first 30 minutes of a vigilance task and then to stabilize at some lower level (Figure 2) (3). At this lower level, the probability of detecting these infrequently appearing signals is lower, thus causing the operator to miss some number of what may be vital signals. In industrial inspection, these "escapes" may seriously hinder assembly or performance of one or more components.

The problem of improving the accuracy of industrial inspectors is difficult and is one which needs more attention than that given in the past. The identification of



Adapted from Reference (3): <u>Vigilance</u>: <u>A Symposium</u>. Buckner, D. N. and J. J. McGrath

Figure 2. Typical Vigilance Performance Decrements

industrial inspection tasks as vigilance tasks relates them to the research accomplished in the field of vigilance performance. If the results of this relationship appear useful in planning inspection work, immediate improvements are suggested. For this reason, the effort in this research is directed toward constructing an experiment which simulates an inspection situation without regard to the kinds of tasks used in current vigilance performance studies.

Vigilance Task Description

The concern in this research is only with the generally accepted criteria for defining a vigilance task since a suitable operating definition enables one to specify task performance requirements without recourse to any of the contemporary theories which have been used in explaining various kinds of behavior observed in vigilance experimentation. Thus, interest is limited to the variables that affect the vigilance decrement.

The nine major variables which may affect the vigilance task and their effects are noted below (3).

1. Rate of signal presentation - An increase in this rate results in a greater percentage of wanted signals detected.
2. Intersignal magnitude - The more regular the interval, the more signals are detected.
3. Signal magnitude - Signals of large magnitude

(size, intensity, duration) escape detection less frequently than those of smaller magnitude.

4. Knowledge of results - Knowledge as to the fact of a missed wanted signal tends to minimize or postpone a vigilance decrement on the task.
5. Environmental factors - Distractions which compete for the attention of the inspector will lower the apparent signal frequency and, consequently, result in deteriorated performance.
6. Knowledge of signal location - When visual search is involved, performance is generally poorer than when location is known.
7. Rest - After a short period of rest, the level of performance returns to that of the beginning of the task.
8. Motivation level - Motivation determines the initial level of performance and may either expedite or postpone the onset of a decrement in performance.
9. Extraneous stimuli - Such stimuli operate as a form of knowledge of results when applied at the time a signal is not detected.

From knowledge of these variables and his research in the field, McGrath (3) has proposed a vigilance task description which was used as the basis for designing the task used in this experiment. This description is reproduced below.

1. The task will require an operator to perceive and report a change in his operating environment.

2. The observer's response behavior should have no effect upon the probability of signal occurrence.

3. The signal to be detected may be either a discrete stimulus added to or taken away from the environment or a change in a continuously or intermittently presented stimulus.

4. Any type of stimulus may be used as a signal provided it is perceivable by the observer when he is alerted or directed to it.

5. The signal to be detected must be specified for the observer.

6. When the signal is a stimulus not requiring an orientation response, the intensity of the signal stimulus should be close to the observer's threshold; otherwise, the less relevance the task has for the study of vigilance performance.

7. Signals should occur infrequently because as the rate of signal occurrence approaches the maximum responding rate of the subject, the less relevance the task has for the study of vigilance performance.

8. The ratio of nonsignal stimuli to signal stimuli should be high. The closer this ratio is to 1:1, the less relevance the task has for the study of vigilance performance.

9. The temporal order of signal occurrence should be irregular. As the temporal order of signal occurrence approaches absolute regularity, the task becomes less relevant to the study of vigilance performance.

10. The task should be prolonged and continuous. The briefer the work period, the less relevance the task has for the study of vigilance performance.

The major performance criterion for a vigilance task meeting the requirements of McGrath's description is both the detection of signals and the probability of that detection within some specified time limit following the occurrence of the signal.

Some discussion of McGrath's vigilance task description is appropriate from the view of the inspection process in order to establish the relevance of each of the items in this operating definition to the inspection process.

With reference to item one, it is the responsibility of the inspector to identify and report defects in a series of items or characteristics being inspected, as noted in the previous definition of visual inspection. These defects are the changes in the operating environment which must be perceived and reported. All inspection tasks conform to this requirement.

Likewise, all inspection tasks conform to the requirement of item two since the probability of a defect is a function of the manufacturing process which has concluded prior to inspection. Defect or signal occurrence probabilities thus are a priori probabilities with reference to the inspection task. These cannot be changed by the inspector.

Item three relates to the nature of the defect itself. For instance, the task used in this study employed a discrete stimulus added to the environment, in this case, a scratch or nick in the strands of a 22 gauge wire. The absence of a stimulus might refer to a missing part. Thus, a defect may exist in either sense - a stimulus added to or taken away from the environment.

To be consistent with item four, the inspector must know what the defect is and be able to see and identify it when it is pointed out to him. If he does not know or understand the defect specification, he then becomes involved in some form of problem solving rather than simply perceiving and reporting a change in his operating environment.

Item five states simply that the inspector must know what kind of defect he is looking for. The foreman does not give him a binful of parts and say simply, "Inspect these." Rather, he says "Inspect these for x, y, and z defects."

Intensity of the signal relative to the observer's threshold is difficult to translate to terms of the inspection tasks. In general, the defect must be detectable when the observer is alerted to it but it must not be so intense as to be completely disruptive to the observer. An unusual or extreme condition of a defect such as a cut wire when only nicks were expected, would be considered to be significantly above the threshold.

Infrequent occurrence of a signal is characteristic of inspection lots with low a priori probabilities of defectiveness. From item seven, it might be inferred that inspection tasks with low levels of defectiveness are more likely to be vigilance tasks than those with higher levels of defectiveness. This is because the inspector's ability is improved in an environment of frequently occurring signals.

Item eight may be interpreted in the same manner as item seven. Inspection lots containing a low level of defectiveness have more relevance for vigilance performance than ones in which the level of defectiveness is high, perhaps as much as 50%.

Item nine relates to the random occurrence of signals over time. If the between signal interval is regular so that the observer can perceive it, then the task is not a vigilance task. In the inspection situation, the inspector may draw items from a bin using some randomizing procedure, or it may be that the defects are manufactured in some random order of occurrence. It is assumed that the occurrence of a defect in an inspection situation is more irregular than not.

Some inspection tasks are prolonged and continuous; others are not. In the factory situation, there is usually a mid-morning and a mid-afternoon break which coupled with the lunch break means that the maximum uninterrupted work period is two hours. Thus, an inspection task, though continuous, would not last longer than two hours without interruption. In the experimenter's experience, the inspector embarked on a task of this duration would very likely stop after 30-40 minutes for a smoke, a drink of water, or for some other purpose, in order to relieve the monotony or fatigue associated with the task. This unscheduled relief enables him to resume work at his normal level of performance and, thus, avoid the decrement usually associated with vigilance tasks.

Of the ten items listed in McGrath's vigilance task description, the first five relate directly to industrial inspection tasks. It is conformance to this description that enables industrial inspection tasks to be classed as vigilance tasks or not. Probably the most difficult requirement to adhere to is number ten, since unscheduled relief periods tend to stimulate performance and so make the task less relevant for the study of vigilance performance.

Vigilance Decrement Pattern

Review of the research results reported by Buckner (3) indicates that there are three characteristics typical of patterns which represent a vigilance performance decrement. The first of these is the pattern configuration. In general, the patterns are similar to those in Figure 2, with a minimum decrement of ten percentage points. Secondly, the decrement occurs in the second or third period assuming ten to fifteen minute intervals. The third characteristic is that the curve should stabilize at the lower level of maximum decrement.

Summary

In this chapter, there has been presented a brief exposition of the major variables pertaining to a vigilance task, a vigilance task description, a discussion of the inspection process relative to this operating definition of vigilance performance, and an operating description of the vigilance pattern. The chapter which follows presents a discussion of signal detection theory in terms of the inspection process.

CHAPTER IV

SIGNAL DETECTION THEORY

Signal detection theory relative to human observers not only provides a theoretic framework for the inspection task, but also identifies performance parameters which may be more suitable for task evaluation than those of McCornack.

Originally, Wald's theory of statistical decision (29) was translated into a theory of signal detection with initial applications in the field of radar design. The detection theory that resulted was a general theory. Its generality suggested its relevance to the detection of signals by human observers. Research on this application of the theory was undertaken by Swets, Tanner and Birdsall (26), whose paper "Decision Processes in Perception" is included in *Signal Detection and Recognition by Human Observers: Contemporary Readings*. Their work served as the base for additional experiments in the field, many of which are reported in the same text.

This chapter describes the inspection task in terms of signal detection theory. The discussion is based on the original paper by Swets, Tanner and Birdsall.

Fundamental Detection Problem

In the fundamental detection problem, an observation is made of events occurring in a fixed interval of time, and a decision is then made, based on this observation whether the time interval contained only background interference (noise) or background interference plus a signal (signal plus noise). Only these alternatives exist in the fundamental problem. Noise is always present, but the signal may or may not be present during a given time interval. When the observer has completed each of his observations he reports, "Yes, a signal was present." or "No, a signal was not present."

Visual inspection tasks as described earlier are a variation of the signal detection problem. Units of conforming product represent the element of noise while a defect on a unit of product represents the signal. The report made by the inspector takes the form of writing the defective item number on a defect report and/or removing it from the inspection lot. In the terminology used in this report, nonconforming units are equivalent to signal plus noise and conforming units are equivalent to noise alone.

This situation may be represented graphically as in Figure 3 if the term observation is taken to refer to the sensory datum on which the decision is based and it is assumed to vary continuously along a single dimension. It

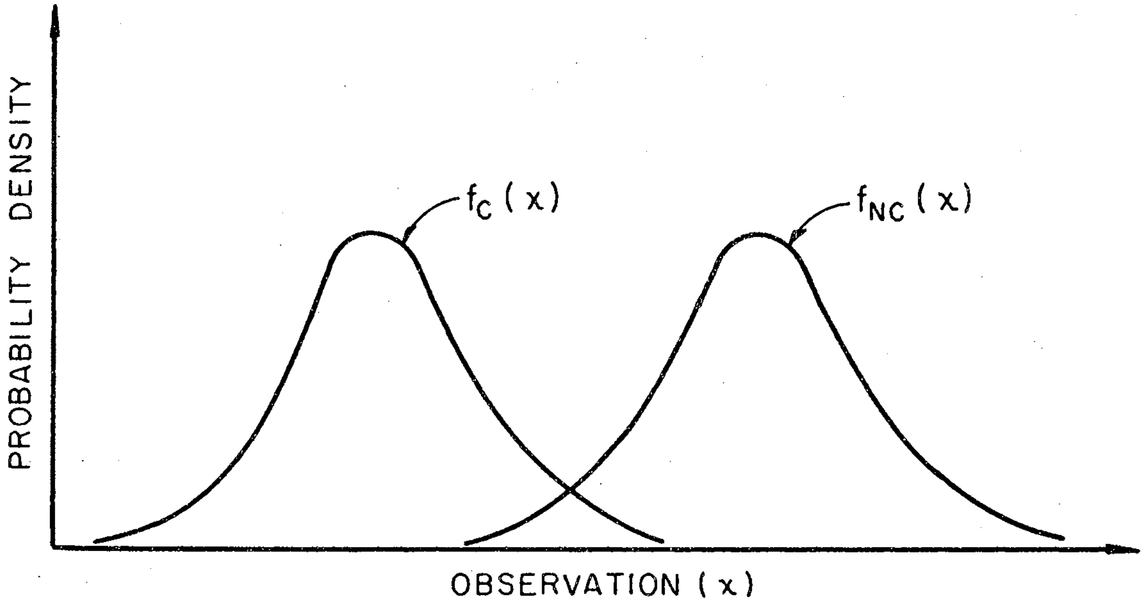


Figure 3. Probability Density Functions of Observations

is assumed further that any observation may arise either from a nonconforming item or a conforming item.

The observation is labeled x and plotted on the abscissa. The distribution $f_C(x)$ represents the probability density that x will occur given conforming product, while the distribution $f_{NC}(x)$ represents the probability density that x will occur given nonconforming product.

It is assumed that all observations may be described in terms of the continuum of a single axis and that both density functions may be assumed to be normal.

Signal Detection Criteria

Referring again to Figure 3, the inspector, given an observation, must decide from which hypothesis the observation resulted; that is, he must state that it came from one distribution or the other. In effect, the inspector must establish somewhere on the continuum a criterion x_c such that if any observation x_i is greater than x_c, he says "Yes, a defect occurred.", while if x_i is less than x_c, he says "No, a defect did not occur."

In Figure 4 there are identified critical regions A (Accept) and R (Reject) as defined by x_c.

Now, it may be seen that the decision process has four outcomes. The inspector may say yes (R) or no (A), and he may be either correct or incorrect. The decision outcome relative to the nonconforming item, thus, may be a hit or correct rejection, (NC · R, the joint occurrence of

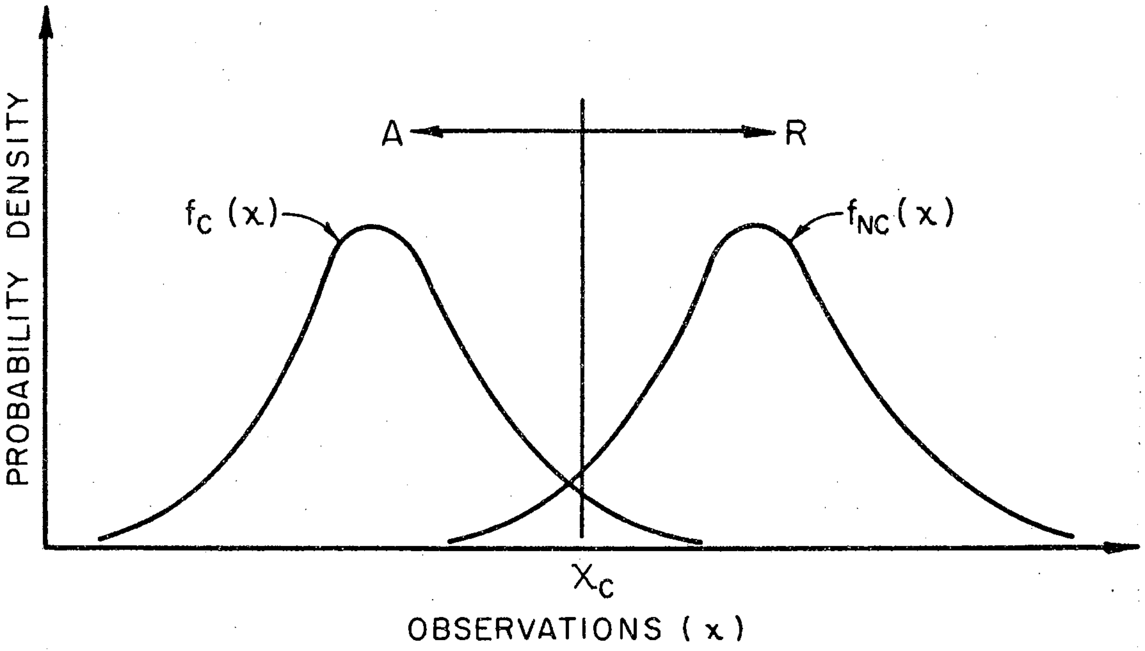


Figure 4. Critical Regions for Observations

the hypothesis NC in and observation in Region R), a miss (NC · A), a correct acceptance (C · A) or a false alarm (C · R). The interdependence of the four probabilities is obvious. Adjustment of the criterion x to increase the probability of a hit, p(NC · R), is achieved only by accepting an increase in the probability of a false alaram, p(C · R) and a decrease in the other probabilities. Whatever the inspector's objectives, the optimal criterion can be specified in terms of the likelihood ratio. That is, if the observation is defined in terms of the likelihood ratio, $\lambda(x) = f_{NC}(x)/f_C(x)$, then the optimal criterion is specified by some value β of $\lambda(x)$.

To illustrate this manner of specifying the optimal criterion, it is necessary to rely on the concept of conditional probabilities as opposed to the statement of joint occurrence noted before. These conditional probabilities are the fundamental quantities in evaluating an observer's performance.

The two probabilities of principal interest are the probability of a reject decision conditional upon the occurrence of a nonconforming item, or $P(R|NC)$, and the probability of an acceptance decision given the occurrence of a conforming item, $P(A|C)$. These two are sufficient for the others are merely their complements.
$P(A|NC) = 1 - P(R|NC)$ and $P(R|C) = 1 - P(A|C)$.

The conditional and joint probabilities are related as follows:

$$P(R|NC) = \frac{P(R \circ NC)}{P(NC)}$$

$$P(A|C) = \frac{P(R \cdot C)}{P(C)} \qquad (7)$$

where

P(NC) = a priori probability of defect occurrence.

P(C) = 1 - P(NC) = a priori probability of occurrence of conforming units.

With reference to Figure 4, P(R|NC) may be defined as the integral of $f_{NC}(x)$ over the critical region R, and P(A|C) as the integral of $f_C(x)$ over A. In other words, the two probabilities represent areas under the curves of Figure 4 on either side of some value x_c.

The expected value of the decision as the sum, over the potential outcomes of the decision, of the products of probability of outcome and desirability of outcome defines an optimal criterion by a critical value of likelihood ratio β. If V equals positive value and K equals cost, then

$$E(V) = V_{NC \cdot R}P(NC \cdot R) + V_{C \cdot A}P(C \cdot A) - K_{NC \cdot A}P(NC \cdot A) - K_{C \cdot R}P(C \cdot R) \qquad (8)$$

Substituting a priori and conditional probabilities from Equation (2) into Equation (3), and collecting terms yields the result that maximizing E(V), the expected value is equivalent to maximizing

$$P(R|NC) - \beta P(R|C) \qquad (9)$$

where

$$\beta = \frac{p(C)}{p(NC)} \cdot \left(\frac{V_{C \cdot A} + K_{C \cdot R}}{V_{NC \cdot R} + K_{NC \cdot A}}\right). \qquad (10)$$

This value of β is equal to the value of the likelihood ratio $\lambda(x)$. In Equation (9) β simply weighs the hits and false alarms, while in Equation (10) β is determined by the a priori probability of the occurrence of nonconforming items and of conforming items and the values for each of the decision outcomes.

Thus, the optimal cutoff x_c is that point on the axis where the ratio of the ordinate value of $f_{NC}(x)$ to the ordinate value of $f_C(x)$ is β. The location of x_c along the x-axis depends not only on the a priori probabilities and values, but also upon the overlap of the two density functions; i.e., the signal strength.

If one plots $P(R|NC)$ as an ordinate versus $P(R|C)$ as the abscissa as in Figure 5, there is obtained a curve known as the receiver operating characteristic, or ROC curve. This curve with d′ as a parameter is based on the assumption that the probability density functions, $f_C(x)$ and $f_{NC}(x)$, are normal and of equal variance. The parameter of the family of ROC curves is called d′ where

$$d' = \frac{\mu_{f_{NC}(x)} - \mu_{f_C(x)}}{\sigma_{f_C(x)}} \qquad (11)$$

Figure 5 is plotted for three values of d′. Assume that

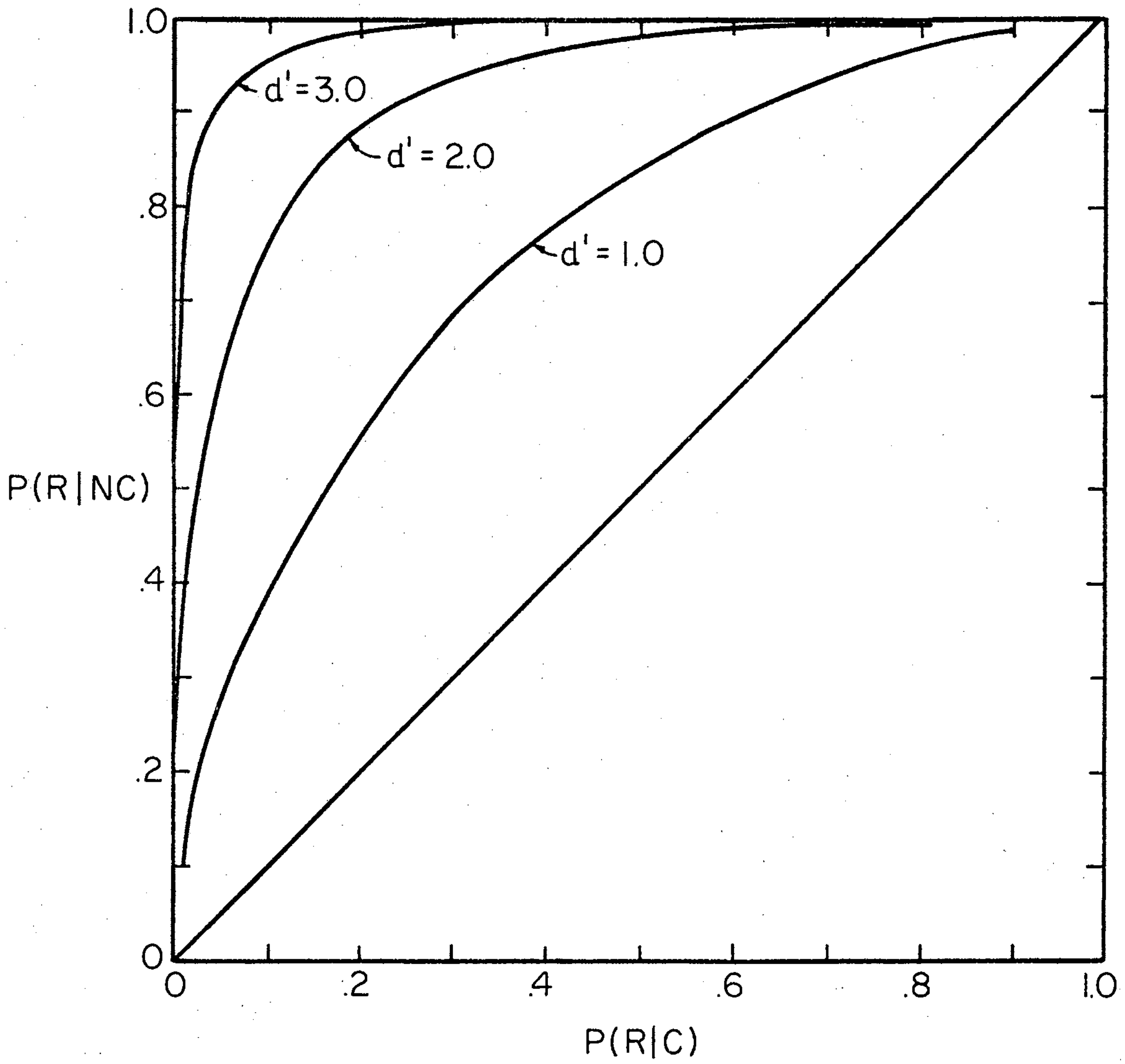


Figure 5. Receiver Operating Characteristic Curve

d′ for the ideal inspector in a given task is 2.0. As the costs, K, and benefits, V, are changed, the optimal criterion for the real inspector would change so that P(R|NC) and P(R|C) will vary along the ROC curve, d′ = 2.0. In the laboratory, trained observers have demonstrated their ability to adjust their criterion to conform to changes in the payoff (26). Adjusting the payoff and, therefore, the optimal criterion provides a means of maximizing either A_2 or A_3 depending on the inspection objective. Application of this concept to industrial inspectors has not been attempted. In this experiment, a payoff will be specified for the inspection task so that a first assessment of the problem of application can be made.

In the signal detection problem, the separation between the means of the two density functions is a function of signal amplitude, therefore d′ is an index of the detectability of a given signal for a given observer. Thus, it is a measure of the observer's sensory capabilities, relative to the performance expected by an ideal observer, as well as a measure of the effective signal strength. Relative to the inspection task, d′ serves as an index of the detectability of a defect for a given inspector. For the task used in this experiment, it is not possible to quantify d′ for the ideal observer as the signal energy and noise power required for determining the signal to noise ratio cannot be determined. However, the use of d′ as the index of detectability of a population of defects enables comparisons to be made of inspector performance relative to that population. By considering the test sample populations as standards, inspector performance can be expressed in terms of d′ relative to those standards.

The Decision Matrix

To relate fully the signal detection problem to McCornack's criteria, it is necessary only to convert the probability matrix of Chapter I to a decision matrix by adding the payoff elements, V and K, to each cell. This is shown in Figure 6.

		Decision	
		Accept	Reject
Product Mix	Conforming	$P(C \cdot A) =$ $(1-p_1)q_0$	$P(C \cdot R) =$ $p_1 q_0$
		$V_{NC \cdot A}$	$K_{C \cdot R}$
	Nonconforming	$P(NC \cdot A) =$ $p_2 p_0$	$P(NC \cdot R) =$ $(1 - p_2)p_0$
		$K_{NC \cdot A}$	$V_{NC \cdot R}$

Figure 6. Inspection Decision Matrix

Summary

In this chapter the inspection task has been described in terms of signal detection theory. The basic concepts of the theory are:

1. The representation of conforming and nonconforming product by overlapping statistical distributions.
2. The representation of a decision criterion and associated payoff matrix.
3. The use of an ideal observer as the basis for evaluating real observers.

From the empirical determinations of $P(R|NC)$ and $P(R|C)$, it is possible to compute d' for each inspector so that this performance measure can be utilized in performance evaluation. Ideally, the experimenter should fix the V and K for each decision outcome and so control β and, therefore, x_c. In the laboratory situation with trained observers this has been accomplished, but the possibility of doing so with industrial inspectors is open to question.

The determination of d' and x_c will be made for the data of this experiment in order to consider in an exploratory sense their application to industrial inspection tasks.

The following chapter deals with the question of performance criteria for the inspection task.

CHAPTER V

PERFORMANCE CRITERIA

The identification of the magnitude of p_1 enables inspector performance measures more meaningful than that of p_2 alone to be developed. These measures are constructed using both p_1 and p_2. In this chapter, other measures, new in the context of inspector performance evaluation, are described. The various measures will be applied to the results of the research and their potential application to the industrial inspection process considered.

McCornack's Criteria

The research reported in Chapter II in all cases used an accuracy criterion corresponding to McCornack's A_3. In addition, Harris used a measure for false detections which related to McCornack's A_2. Harris' false detection measure, however, is simply the percentage of nondefectives in the total number of defectives reported. For a report of false detections in terms of A_2, it is necessary to know the true fraction defective. Unfortunately, Harris' studies did not report this information, so it is not possible to express A_2 for his data exactly. Nor is it possible to express Harris' data in terms of A_1 inasmuch as the total number of inspections is not known.

In vigilance research the measure of effectiveness used corresponds to McCornack's A_3; however, there is no reason to assume that the A_2 type of error is not affected as well in the vigilance situation.

It is the view of the experimenter that the factor A_1 should be used in preference to the other of McCornack's factors because it appears to be even more difficult to identify defects caused by the inspector than it is to identify those caused by the production mechanic. However, the major consideration should be based on the view to be taken of the product in the light of inspection results. It was stated in Chapter I that an inspector with a high error rate can cause substantial damage by passing defective hardware. This implies an error rate of the A_3 type and does not consider the expense associated with the rejection of good product which involves an error rate of the A_2 type. When defects are reported, they must either be repaired or eliminated through product replacement. Either course of action is expensive not only from the standpoint of material and labor costs, but also from the view of a missed schedule and loss of quality reputation. Therefore, the factor A_1 seems most appropriate for use as it relates both inspection errors to the total situation.

However, these criteria, A_1, A_2, and A_3, are not really expressions of confidence in the decision to accept or reject an item once that decision has been made. They are simply expressions of past performance. What is needed is a statement of confidence as to the validity of the inspection decision given that it has been made under a particular set of conditions. To write this kind of criterion, it is necessary to resort to the rule of Bayes and conditional probabilities.

Bayesian Criteria

The rule of Bayes is stated as follows (23):

If $B_1, B_2, \ldots, B_n$ are mutually exclusive events of which one must occur; that is $\sum_{i=1}^{n} P(B_i) = 1$, then

$$P(B_r|A) = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{n} P(B_i)\cdot P(A|B_i)} \tag{12}$$

for $r = 1, 2, \ldots,$ or n.

Application of Bayes' rule to the decision matrix for the inspection task enables more meaningful statements concerning performance to be written with respect to the validity of each of the decisions made. One can now state the probability that an item is a conforming one given the decision to accept it, in the following manner:

Substituting in Equation (2) as follows:

$$P(C|A) = \frac{P(C)\ P(A|C)}{P(C)\ P(A|C) + P(NC)\ P(A|NC)} \qquad (13)$$

where

C = conforming item
NC = nonconforming item
A = accept decision
R = reject decision
P(A|C) = probability of a correct decision on C
P(C) = a priori probability of C
P(NC) = a priori probability of NC
P(A|NC) = probability of a Type II error on NC.

Further substitution using appropriate values from the probability matrix of Chapter I gives:

$$P(C|A) = \frac{q_0(1-p_1)}{q_0(1-p_1) + p_2 p_0} \qquad (14)$$

which is the conditional probability of a conforming item given the decision to accept it by the inspector.

Likewise,

$$P(NC|R) = \frac{(1-p_1)(p_0)}{q_0 p_1 + p_0(1-p_2)} \qquad (15)$$

which is the conditional probability of a nonconforming item given that a decision to reject has been made by the inspector. Thus are established two criteria for evaluating performance relative to either of the decisions indicated by the probability matrix.

Bayesian Criteria Versus McCornack's Criteria

McCornack's criteria, particularly A_3, are usually used in evaluating inspection processes. Using the decision matrix, these can be defined as follows:

$$A_1 = (1 - p_1)q_0 + (1 - p_2)p_0 \tag{16}$$

$$A_2 = (1 - p_1)q_0/q_0 = (1 - p_1) \tag{17}$$

$$A_3 = (1 - p_2)p_0/p_0 = (1 - p_2). \tag{18}$$

It can be seen that the Bayesian criteria use all the information available relative to each decision, whereas McCornack's criteria are relative to each level of the product mix rather than to the decision process. That is, A_2 and A_3 each consider only one cell of the matrix whereas Bayesian criteria consider two cells. A_1 is the same for both McCornack's and Bayesian measures.

Referring to Equation (14), the conditional probability of a conforming item given an acceptance decision is a function of the number of conforming items and the number of nonconforming items accepted. Thus, $P(C|A)$ is a measure of the goodness of the acceptance decision. Also, the conditional probability of a nonconforming item given a reject decision, Equation (15), is a function of the number of nonconforming items and number of conforming items rejected. Thus, $P(NC|R)$ is a measure of the goodness of the rejection decision. Neither of these probabilities measure performance in the traditional sense of A_2 and A_3; that is, the ability to identify good product or bad product. The measure of the goodness of the inspection decision is an indication of the liklihood of defective product passed on to the next station or to the customer - $P(C|A)$, or an indication of the amount of good product reworked or scrapped - $P(NC|R)$. As costs may be associated with both measures, the value of the application of Bayes' rule becomes apparent.

In order that adequate comparison of these criteria can be made, the experimental data will be analyzed both ways, so that the utility of the criteria can be evaluated.

Receiver Operating Characteristic Curve

A further criterion not heretofore used in industrial inspection is the index of detectability of a given signal for a given observer, d'. This parameter has merit in evaluating the inspection process as it provides a basis for comparison of inspector's performance in terms of his receiver operating characteristic curve. As the data collected in this experiment are readily amenable to its calculation, d' will be calculated for each inspector over the range of the study and its utility evaluated along with the criteria already discussed.

Apparent Fraction Defective

If there were no inspection errors ($p_1 = p_2 = 0$), the fraction defective of an inspection lot reported by the inspector would equal the a priori fraction defective, p_o. However, if either or both kinds of errors have been committed by the inspector, then the fraction defective reported, p_3, will be

$$p_3 = q_o p_1 + p_o(1 - p_2) \tag{19}$$

for

$$(0 < p_1 \leq 1) \text{ and } (0 < p_2 \leq 1).$$

This criterion for performance has been proposed in order to provide a means of studying the effect of inspector error on lot disposition in a single sampling situation. Having determined values of p_3 relative to corresponding values of p_o, it is possible to evaluate the protection actually obtained in a sampling plan against that which the plan was designed to provide. Sampling plans are designed with great mathematical rigor to provide a certain level of protection against acceptance of defective product. This design is based on the assumption that there is no error in the inspection process or that whatever error exists is small and can be ignored. This assumption can be tested by seeing what decision for lot disposition would be made if p_3 is perceived by the inspector rather than p_o.

Assume that a single sampling plan is designed to accept some value of p_o with a probability of acceptance of 0.50. If the inspector's perceived value of product defectiveness, p_3, is close to the true value, p_0, the inspection decision made will be very close to that expected under the specific sampling plan employed. That is, lots will be accepted or rejected approximately fifty percent of the time. Unfortunately, they will likely be accepted or rejected for the wrong reasons when p_1 and p_2 errors are being made. An accepted lot may have been accepted as the result of inspecting a sample wherein a number of defectives were missed, or a rejected lot may have been rejected because conforming units were rejected. One could argue that these kinds of errors either tend to cancel out, will be caught at the next station in the case of an improper acceptance, or will be caught by the material review board in the case of improper rejection. This argument cannot be maintained because the costs associated with each alternative are not considered. For instance, if the recipient of the improperly accepted lot is the customer, then a loss of quality reputation and perhaps a loss of business results. In the other case, at a minimum, a loss of schedule position can occur, together with the possibility of unnecessary rework or a scrap disposition by a material review board.

Summary

In this chapter a number of criteria have been proposed as possible measures of the quality of performance of an inspection task. These measures will be applied to the results of the experiment and their potential usefulness indicated.

The following chapter deals with the design and technique of the experiment.

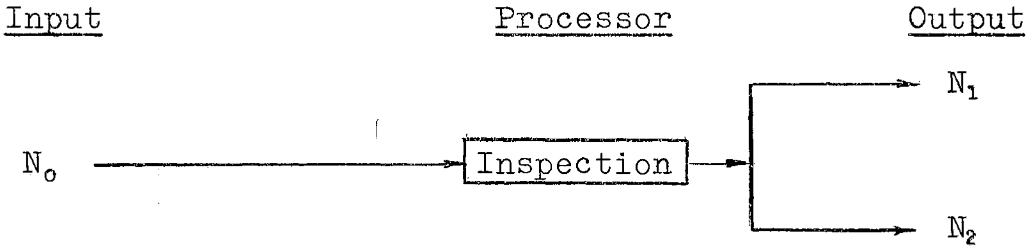
CHAPTER VI

EXPERIMENTAL DESIGN AND TECHNIQUE

This research is concerned with the accuracy of performance of a class of industrial inspectors. This restriction as to the population being tested automatically extends the research effort beyond the laboratory to a pseudo-industrial environment. As a result, the experimenter is immediately faced with the problem of realism in the experimental task. At one extreme of the continuum there exists the possibility of a wholly realistic test environment in which it may not be possible to establish and maintain experimental controls, whereas at the other extreme there exists a totally experimenter-controlled laboratory environment wherein the aspect of realism may be lost. In this research it was desired to obtain as great a degree of realism as possible so that the results would represent, at least to this degree, the real world environment. At the same time the necessity of presenting standardized situations to the testees was recognized, including the assurance that the same problem must be administered to each of them. Thus, a compromise between the two extremes was reached. The details of the design are contained in the balance of this chapter.

Data Collection Model

The first step in the experimental design is to establish a model for data collection. The inspection model of Figure 1 may be restructured as follows:



where:

N_0 = number of pieces in the inspection lot prior to inspection.

N_1 = number of pieces accepted as conforming to specification.

N_2 = number of pieces rejected as nonconforming to specification.

The division of N_0 into N_1 and N_2 reflects the decisions of the inspector relative to the existence of one or more defects in the inspection lot. To gather data concerning N_1 and N_2 and to compute the performance criteria discussed earlier, it is necessary to make further definition as follows. Let:

n_0 = number of conforming units in N_0.

n_1 = number of conforming units in N_1.

n_2 = number of conforming units in N_2.

d_0 = number of nonconforming units in N_0.

d_1 = number of nonconforming units in N_1.

d_2 = number of nonconforming units in N_2.

p_o = true fraction defective of N_o. $(0 \leq p_o \leq 1)$.

p_1 = probability of Type I error in the inspection process. $(0 \leq p_1 \leq 1)$

p_2 = probability of Type II error in the inspection process. $(0 \leq p_2 \leq 1)$

p_3 = apparent fraction defective after the i^{th} inspection. $(0 \leq p_3 \leq 1)$

Then the following relationships exist:

$$N_o = n_o + d_o \tag{20}$$

$$n_o = n_1 + n_2 \tag{21}$$

$$d_o = d_1 + d_2 \tag{22}$$

$$p_o = \frac{d_o}{n_o + d_o} = \frac{d_o}{N_o} \tag{23}$$

$$p_1 = \frac{n_2}{n_o} \tag{24}$$

$$p_2 = \frac{d_1}{d_o} \tag{25}$$

$$p_3 = (1 - p_o)\, p_1 + p_o(1 - p_2)$$

$$= \frac{n_2 + d_2}{N_o} \tag{26}$$

$$P(C|A) = \frac{n_1}{n_1 + d_1} \tag{27}$$

$$P(NC|R) = \frac{n_2}{n_2 + d_2}. \tag{28}$$

By fixing n_o and d_o for a given inspection lot and observing the n_2 and d_2 resulting from the inspection, each of the criteria of concern can be readily calculated using simple addition and subtraction. Further, if each inspection item is uniquely identified such that it is readily discernible to an observer, then the observer can record which item the inspector is examining at any desired point in time and, thus, compute these same criteria relative to the desired time measure.

Based on examination of vigilance performance studies (see Figure 2), a time measure of ten minutes was chosen within which to identify the vigilance decrement should it occur.

Test Materials - Kind

In order to approach the objective of realism as closely as possible, a wire preparation inspection task was chosen with minor modifications made in the usual method of inspection so as to assure control of the experiment.

Normally, when conductors are assembled into a wire harness and the ends prepared for soldering or crimping, approximately one inch of insulation is mechanically stripped from the ends and the exposed strands are inspected for cuts, nicks, and missing strands. Unacceptable preparation is cause for rejection and requires appropriate corrective action. When one considers the inspection of successive bundles of stripped conductors for a few difficult-to-detect defects as a repetitive task with monotonous aspects, it may be supposed that this task may well meet McGrath's criteria for vigilance performance and, hence, should be evaluated in these terms. However, the use of wire bundles does not enable the experimenter to identify decision errors individually by conductor and so does not permit the identification of a specific wrong decision at the time of its occurrence. To overcome this constraint, four-inch lengths of thermally stripped, 22-gauge, Teflon-coated conductor were assembled in groups of ten by affixing each of the ten conductors to a 12-inch strip of 1^1/2 inch masking tape about one inch apart. (See Figure 7.) The first strip of tape was then covered by a second strip which sealed the assembly. Pairs of strips were given a letter designation, A, B, C, ..., and each conductor in the pair was numbered from one through twenty so that each conductor now had a unique alphanumeric identification.

Randomly selected conductors were made defective by nicking or scraping them with a sharp knife. This operation left a defect like those left by a mechanical stripping process.

Test Materials - Quantity

As it was desired to evaluate performance under different levels of defectiveness as well as over time, four values of p_0 were chosen to be .05, .15, .25, and .35 for four inspection lots where N_0 equaled 260. In a pilot

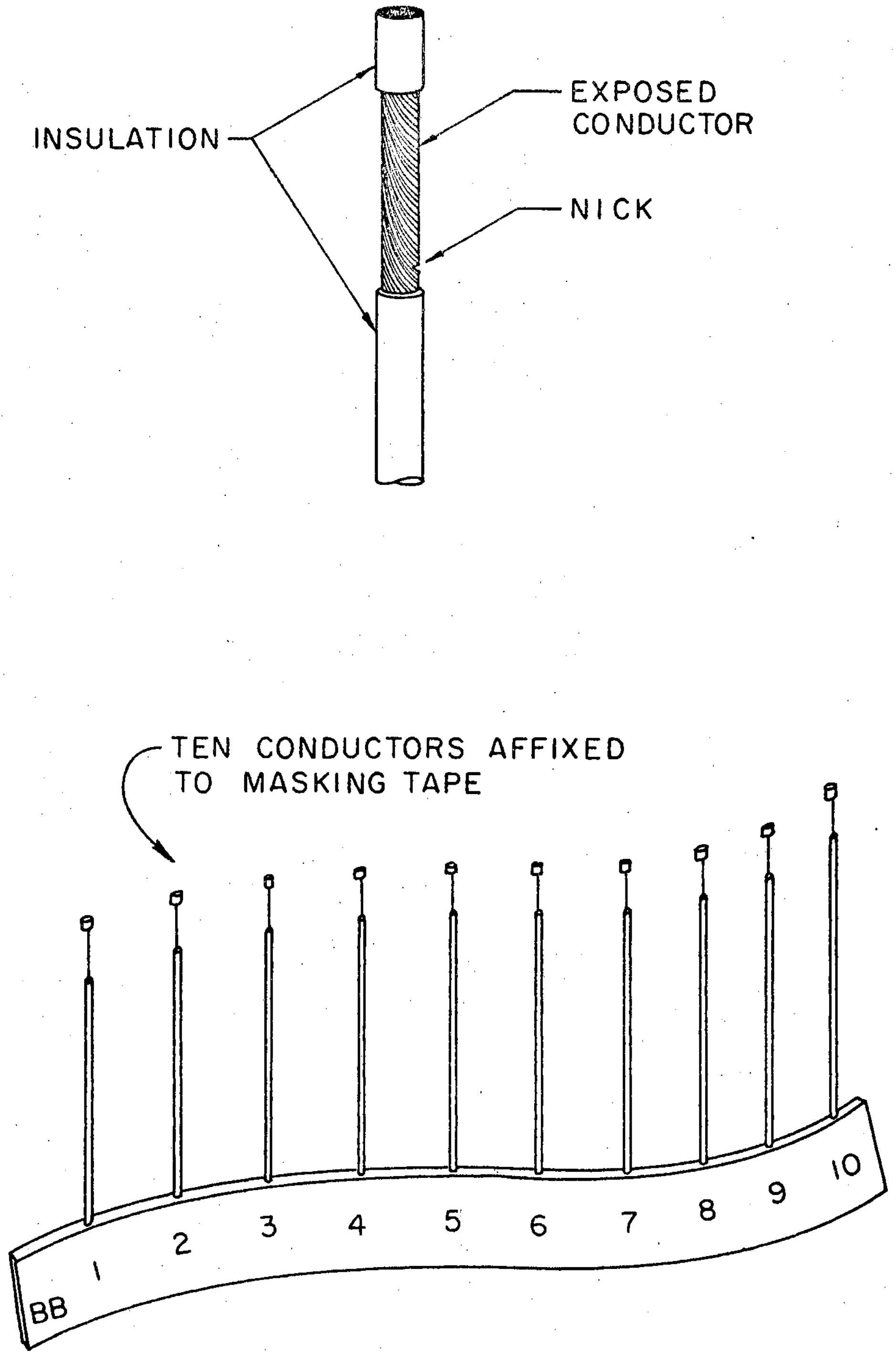


Figure 7. Test Materials

study preceding this experiment, it was found that the average inspection rate was about four conductors a minute; hence, 260 was chosen as being sufficient for approximately one hour's inspection time.

In an effort to avoid biasing any one set of test materials by accidentally making the defects different from those in any other set, the technique of making the defects with the knife was practiced until a reasonable degree of consistency was obtained. Then, the required 208 defects for the four samples, along with some spares, were made at one time. These defective conductors were then distributed randomly over the four sets of materials until all were allocated. The defective items were then distributed randomly within the estimated time blocks per set of materials so that the fraction defective, p_o, for each ten minute period would be about the same as that for the total sample. For example, assume that an inspection is performed at the rate of four conductors per minute, or forty for each ten minute period, and the desired p_o for this particular lot is .10. This would mean that four defectives would be distributed randomly throughout each forty items. In this way, p_o is held more or less constant relative to time throughout the test period. It is not possible to maintain the same level exactly throughout as the inspection task is operator-paced and inspectors will work at different rates.

In this manner the four sets of test materials were developed. The selection of the values 5, 15, 25, and 35 for p_o was arbitrary in that the purpose was to test over a broad range of defectiveness that is nevertheless consistent with process averages experienced in industry. The philosophy was that the results of a gross cut at this problem could be expected to indicate where the next finer cut should be taken. Such was the case here. As will be noted in the conclusions section of this dissertation, recommendations are made as to the range of proposed follow-on experiments.

The four sets (designated by p_o) of materials were identified by assigning the numbers one through twenty to each letter of the alphabet as follows:

5% - A(1-20) through M(1-20)
15% - N(1-20) through Z(1-20)
25% - AA(1-20) through MM(1-20)
35% - NN(1-20) through ZZ(1-20).

Thus, for each letter or double letter of the alphabet, there were two groups of ten conductors each.

Inspectors Tested

Seven electronic inspectors were chosen to participate in the experiment. All were senior electronics inspectors with at least twenty-four months experience. The inspectors were both male and female ranging in age from 28 to approximately 62 and carrying the same job classification. Their vision exceeded the minimum requirement of their employer for the kinds of tasks performed. Their usual work consists of inspecting to standards, drawings, and specifications materials such as the prepared conductors used in this study and includes performing visual checks of complex electronic equipment. It is felt that this group should be a representative sample of this class of inspector.

At the outset of the study, the inspectors were informed of the experimental nature of the study and the kind of inspection task to be performed. In particular, they were told that while they would be informed of the results of their performance on an individual basis, that they would be identified only as a group in the final report, and that the purpose of the study was to learn about performance without identifying those individuals who might not perform as well as others, thus protecting them from potential supervisory criticism. While each subject is known to the experimenter, he is identified only by letter in this dissertation.

Test Schedule

To avoid test compromising effects such as time of day and sample order, the schedule utilized a counterbalanced design such that each inspector inspected the four samples in different order and at different times of day. Because of the work requirements placed on the inspectors throughout the test program, the original schedule was lengthened; however, the sample order and time of day were maintained as planned. The test schedule is shown in Table I. Time period 1 began at 7:15 A. M. and ended just prior to the morning break. Period 2 began at 9:15 A. M. and ended at lunch. Period 3 began at 11:45 A. M. and ended just prior to the afternoon break. The last period began at 1:45 P. M. and ended ten minutes before quitting time. Each period was approximately 105 minutes in duration. This provided sufficient time for an instruction period, a pretest, and the main test.

Pretraining Schedule

In Chapter III ten elements of a vigilance task were defined. Item five of that list stated that the signal to be detected must be specified for the observer. Although the experimental task is one which is routine to the inspectors' work environment, a training program relative to the test materials was established to assure that the requirement of item five was met and that, in fact, the inspectors understood what was required of them. In addition, an attempt was made to establish the concept of payoff by assigning a relative value to each possible decision outcome and then discussing training performance in terms of this score.

A copy of the instructions given to each inspector is included in Appendix A. A copy of the data sheet is included in Appendix B. The instruction samples consisted

TABLE I

TESTING SCHEDULE

Period	Test Day						
	1 Insp/p_o	2 Insp/p_o	3 Insp/p_o	4 Insp/p_o	5 Insp/p_o	6 Insp/p_o	7 Insp/p_o
1	x	A-15 D-15	G-25 x	F-35 H-15	B-35 x	I-25 x	x
2	x	I-5 x	A-15 H-25	B-5 G-35	D-25 x	x	F-25
3	A-35 x	B-15 F-25	D-35 I-15	x	H-5 x	G-15 x	x
4	D-5 H-35	G-5 x	x	I-35 x	F-15 x	B-25 x	A-25 x

of two groups of ten conductors and two groups of twenty conductors with levels of p_o as noted in the flow chart of Figure 8. As the purpose of the procedure was to assure that the inspectors were able to detect the existing defects, provision was made for repeated instruction as necessary. The training was concluded with a payoff greater than twelve in the last step. This was equivalent to two errors in classification.

Test Procedure

Following the satisfactory completion of the training by all inspectors, the testing for record was undertaken according to the schedule noted in Table I.

The administration of the tests was accomplished in a controlled work area immediately adjacent to the inspector area. Each subject was seated at a table on which twenty-six sets of ten conductors were laid out so that they could be examined in alphanumerical order. At the signal to begin, the subject examined the first set of ten conductors and recorded the number of each one of them that appeared to be defective on his data sheet. He continued in this manner until all test items were inspected or until the end of 75 minutes. At the end of each ten minute period, the observer noted the number of the last item completed so that the rate of inspection and number of decision errors could be computed.

Prior to each test a pretest of 20 items was given

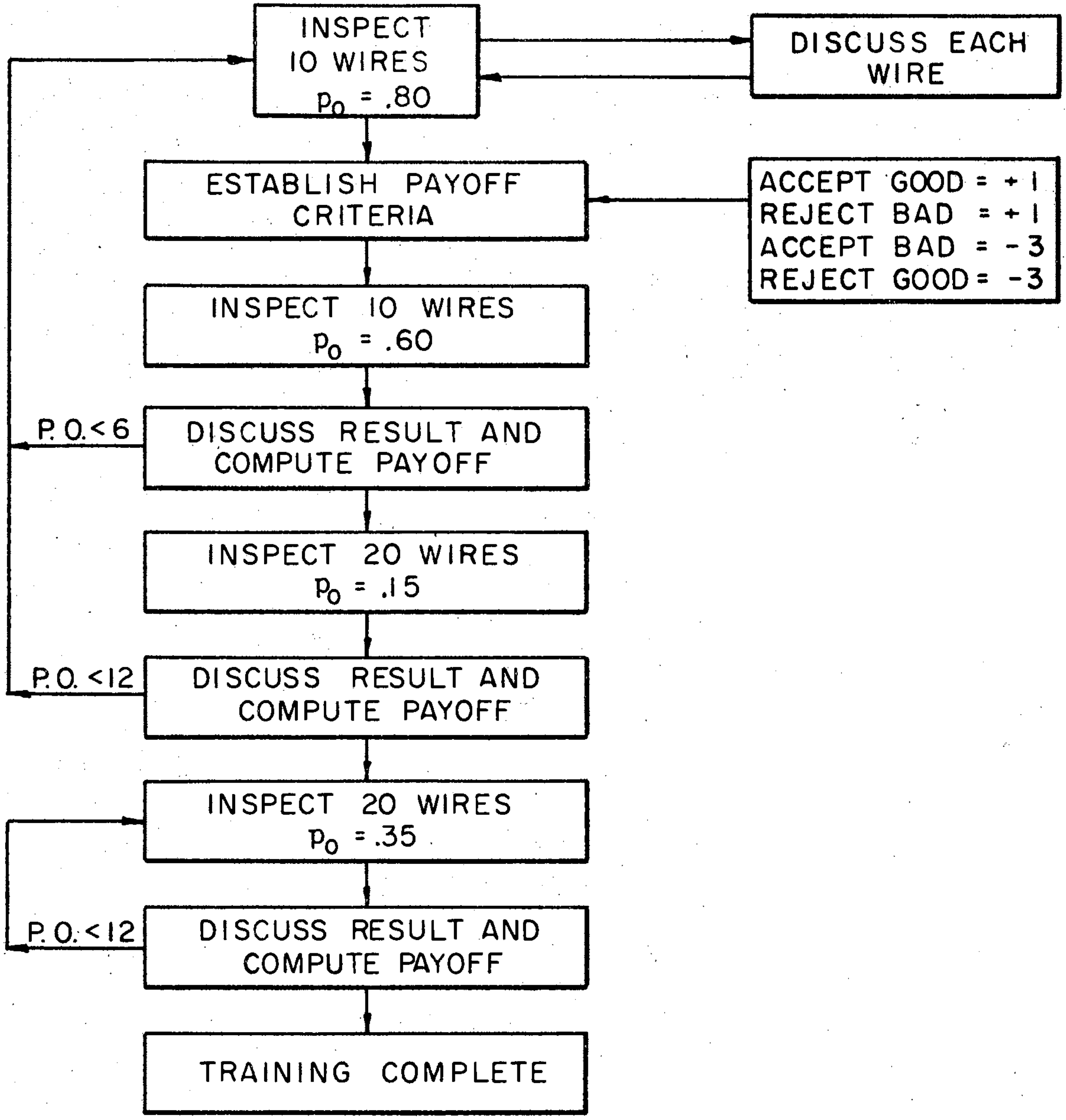


Figure 8. Wire Preparation Inspection Training

the inspector. The twofold purpose of the pretest was to reinforce the training already given and to establish without verbalization the level of defectiveness to be experienced in the test for record. The reason for this latter point is that an inspector working regularly with a particular manufacturing group gradually develops an expectancy or set for the level of defectiveness produced by the group (27). He may have no way of knowing, a priori, what the level is but nevertheless he tends to work to his concept of it. Thus, in the pretest, an attempt was made to establish that expectancy without actually verbalizing the exact value of p_o.

At the conclusion of each test, the data sheet was taken by the observer, the inspector was thanked and sent back to his work area. At the conclusion of all testing, the results were reported to each inspector individually and to supervision for the group. The performance criteria used in the report were McCornack's A_2 and A_3.

Data Analysis

As the experimental data are not measured on a continuous scale but are instead classified as "success" or "failure" for each performance measure, the statistical analysis will be undertaken using a χ^2 statistic for proportions. The basis for this selection is that testing is to determine whether two or more binomial populations have the same parameter P*. Identifying k parameters as $P_1 = P_2 = \ldots = P_k$, the null hypothesis to be tested is

$$H_0 : P_1 = P_2 = \ldots = P_k$$

against the alternative that at least two of these proportions are unequal. Since four binomial populations are utilized in the experiment, the value of k is four. The null hypothesis then is rewritten as

$$H_0 : P_5 = P_{1\,5} = P_{2\,5} = P_{3\,5}$$

where P is the proportion of interest (the failures) for each of the performance measures. Thus, for the measure A_1, P is the proportion of incorrect decisions or $(1 - A_1)$ for each p_0. P for A_2 represents the proportion of failures to correctly identify conforming product, and P for A_3 is the proportion of failures to detect defects.

The test used is based on the fact that

> (1) for large samples the sampling distribution of
>
> $$z_i = \frac{x_i - n_i p_i}{\left(n_i p_i (1 - p_i)\right)^{1/2}} \qquad (29)$$
>
> is approximately the standard normal distribution, (2) the square of a random variable having the standard normal distribution is a random variable having a chi-square distribution with one degree of freedom, and (3) the sum of k indpendent random variables having

---

*An uppercase P is used to differentiate between the test proportion and the lower case p_0, p_1, p_2, and p_3.

> chi-square distributions with one degree of freedom is a random variable having the chi-square distribution having k degrees of freedom (23).

Therefore, the sampling distribution of the statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i p_i)^2}{n_i P(1 - P)} \qquad (30)$$

is approximately the chi-square distribution with k degrees of freedom. Thus, if the null hypothesis is true, and

$$P_5 = P_{15} = P_{25} = P_{35} = P$$

the chi-square statistic becomes

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i P)^2}{n_i P(1 - P)}. \qquad (31)$$

As P is unknown, there is substituted for P the pooled estimate

$$\hat{P} = \frac{x_1 + x_2 + \ldots + x_k}{n_1 + n_2 + \ldots + n_k}. \qquad (32)$$

If the differences betwen the x_i and the $n_i P$ are large, the null hypothesis should be rejected, so that the critical region for the test is $\chi^2 \geq \chi_\alpha^2$, with (k - 1) degrees of freedom. In this experiment, α is taken as 0.05.

Applying the chi-square criterion to the problem of comparing several sample proportions in a 2 by k contingency table, the data may be arranged in the following manner:

	Sample 1	Sample 2	...	Sample k	Total
Success	$n_1 - x_1$	$n_2 - x_2$		$n_k - x_k$	$n - x$
Failure	x_1	x_2	...	x_k	x
Total	n_1	n_2	...	n_k	n

In this table, n is total trials, and x is the total failures for n trials.

To simplify the determination of χ^2, the computing method used is that attributed to Brandt and Snedecor by Cochran and Cox (5). It is

$$\chi^2 = \frac{\sum x_i P_i - x\bar{P}}{\bar{P}\,\bar{Q}} \qquad (33)$$

where

x_i = the number of failures in the i^{th} sample.
$\bar{P}$ = over-all proportion of failures (P).
$\bar{Q}$ = $(1 - \bar{P})$.

The contingency table can now be rearranged as follows:

	$p_o(1)$	$p_o(2)$	...	$p_o(k)$	Total
Failures	x_1	x_2	...	x_k	x
Total	n_1	n_2	...	n_k	n
P	x_1/n_1	x_2/n_2	...	x_k/n_k	x/n

where P is $(1 - A_1)$, $(1 - A_2)$ or $(1 - A_3)$ as appropriate, and x/n is $\bar{P}$.

While contingency tables and the χ^2 statistic are also appropriate for testing for time effects, they must be reinforced by comparing the plotted pattern to the configuration criteria described in Chapter III. As interest is not only in a difference but also in the direction of the difference, the z statistic will be used to test the null hypothesis of no difference. If there is a performance decrement typical of vigilance performance then the maximum value of A_3 should occur during the first or second time block and the minimum value should occur during the third or fourth time block with the difference between them significant at $\alpha = 0.05$. In this situation, the null and alternate hypotheses are

$$H_0: \quad p_1(\max) = p_2(\min)$$

$$H_1: \quad p_1(\max) > p_2(\min)$$

with the critical region being $z \geq z_{.05}$. As $z_{.05} = 1.645$, the calculated value of z must be greater than or equal to 1.645. The z statistic is computed by

$$z = \frac{x_1/n_1 - x_2/n_2}{((\hat{p})(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2}))^{1/2}} \qquad (34)$$

where x equals the number of signals detected in n trials for each of the two populations and with

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}. \qquad (35)$$

As vigilance studies are concerned with the percentage of signals detected, p_1 and p_2 in the equation for the z statistic above, become A_3(max) and A_3(min), respectively, and x is the number of signals detected in n trials per time block. The $\hat{p}$ then becomes $\hat{A}_3$.

Summary

In this chapter there has been discussed the experimental procedure used in the data collection phase of this research. The emphasis in the testing was to be as realistic as possible while maintaining control of the experiment. A training procedure was established to assure that all inspectors worked to the same criteria. Through the use of standard instructions and test materials essentially the same problem was administered to all testees.

The next chapter presents the results of the data collection and their analysis according to the criteria defined earlier.

CHAPTER VII

RESULTS AND ANALYSIS

The results of the analysis of the data collected during the course of the experiment are presented in this chapter with the order of discussion following that of the basic questions posed in Chapter I.

Data summaries along with the contingency tables used in testing hypotheses are included in Appendices C through E.

The first question dealt with the effect of the a priori probability of defectiveness of the test materials on inspector performance. The second question was concerned with evidence of a vigilance performance decrement over the time of the test, and the third question related to the use of possible performance measures other than the traditional one of the percentage of defects detected (A_3) in evaluating inspector performance.

Effect of p_0: McCornack's Criteria

The computed values of A_1, A_2, and A_3 for each p_0 treatment are plotted in Figure 9. For each of the three measures, the null hypothesis of no difference in performance over the given range of p_0 was tested against the

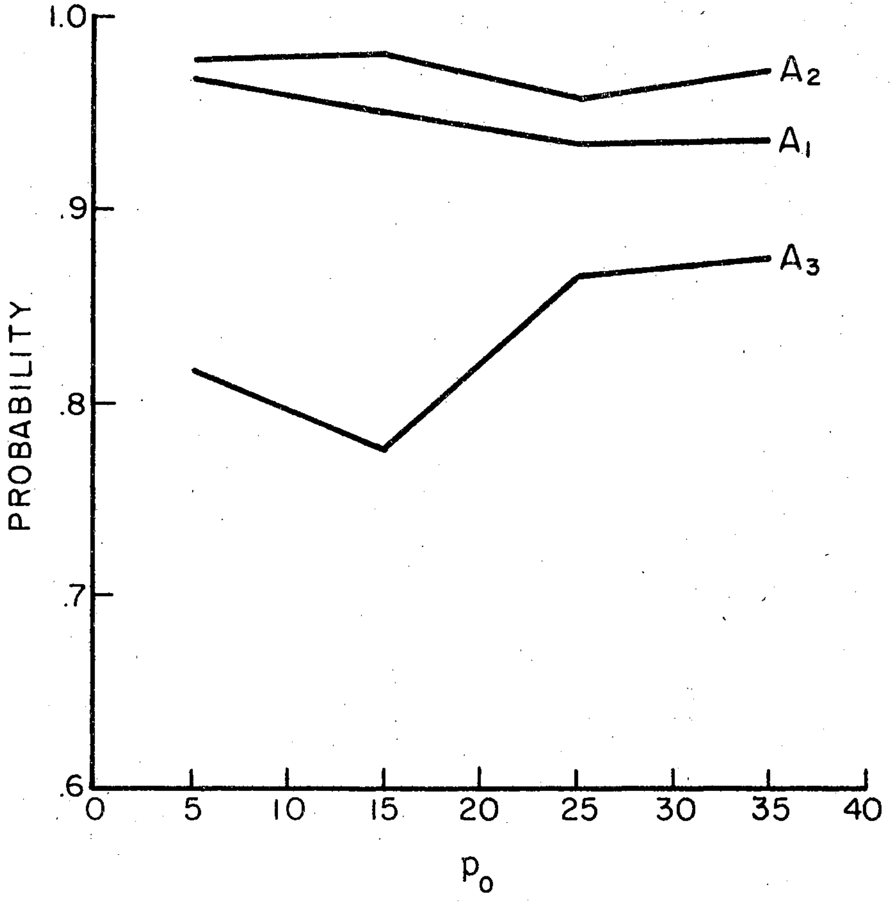


Figure 9. McCornack's Criteria Versus p_0

alternative hypothesis of unequality of performance using the X^2 statistic and the method described in the previous chapter. The results are presented opposite question one in Table II. The contingency tables supporting Table II are in Appendix E.

Because of the lower values of A_3 in the 5%-15% range of p_0 as compared to the 25%-35% range, question two of Table II was formulated and the data combined in a 2 x 2 contingency table for testing. Because the null hypothesis of question two was rejected, and because there appears to be an upward trend in A_3 as p_0 increases, question three was formulated and tested using the method described in Appendix F.

The results of these several tests indicate that while the differences in A_3 over the range of p_0 are due to a linear regression effect, the differences in the A_2 measure are not.

In a further attempt to explain the variation in A_2 which led to the rejection of the null hypothesis of question one, questions four and five were formulated and tested. As a result, it was determined that the value of A_2 at $p_0 = 25\%$ was contributing most of the variation in A_2. Because this phenomenon cannot be explained at this time, further investigation of inspector performance relative to the correct identification of conforming items should be considered. Were it not for this disparity, the conclusion would have been that correct identification of

TABLE II

TEST RESULTS: McCORNACK'S CRITERIA

	Test Questions	Measure	χ^2	$\chi^2_{.05;3}$	Decision
1.	Are there differences in performance among the p_o treatments?	A_1	17.44	7.81	Reject H_o
		A_2	13.51	7.81	Reject H_o
		A_3	13.67	7.81	Reject H_o
2.	Are there differences in performance between the combined treatment p_o equal to 5%-15% and the combined treatment p_o equal to 25%-35%?	A_2	4.17	$\chi^2_{.05;1}$ 3.84	Reject H_o
		A_3	12.82	3.84	Reject H_o
3.	Are the performance differences due to a linear regression effect over the range of p_o?	A_2	2.22	$\chi^2_{.05;1}$ 3.84	Cannot Reject H_o
		A_3	9.25	3.84	Reject H_o
4.	Is the value of A_2 at p_o = 25% different from the other values of A_2?	A_2	14.66	$\chi^2_{.05;1}$ 3.84	Reject H_o
5.	Are there differences among the values of A_2 for p_o = 5%, 15%, and 35%?	A_2	1.77	$\chi^2_{.05;2}$ 5.99	Cannot Reject H_o

conforming items was not affected by the a priori probability of conformance.

Effect of p_o: Bayesian Criteria

The computed values of P(C|A) and P(NC|R) for each treatment of p_o are plotted in Figure 10. Inspection of the patterns suggests that a plateau exists in the midrange of p_o (15%-25%) with the extreme treatments different from the midrange.

Testing for these differences using the z statistic for proportions supports this hypothesis. The results of the several tests required are summarized in Table III.

Relative to these performance measures, it is concluded that the decision process is affected by the a priori probability of nonconforming product at the extremes whereas the midrange is stable but at a different level on the ordinate probability scale. At the 5% treatment, it can be seen that few defective items are accepted but that many conforming items are rejected. These tendencies change significantly at the 15% treatment and even dramatically in the case of P(NC|R). More defective items are accepted while a great many less conforming items are rejected. There is not a significant difference in either measure from the 15% treatment to the 25% treatment while the trend downward in P(C|A) and the trend upward in P(NC|R) continues in the transition from 25% to 35%.

The difference between the plot of P(C|A) and of

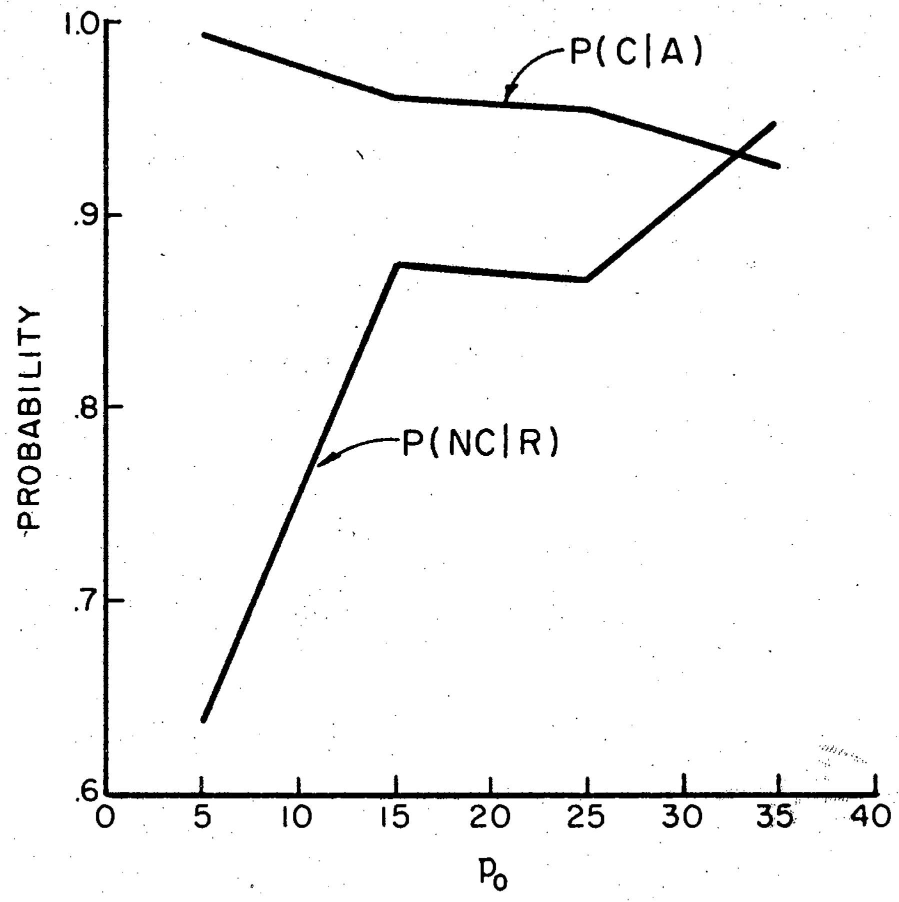


Figure 10. Bayesian Criteria Versus p_o

TABLE III

TESTS OF HYPOTHESES FOR BAYESIAN CRITERIA

Measure	H_0	H_1	Critical Region	Decision
P(C|A)	$p_o5 = p_o15$	$p_o5 > p_o15$	$z > z_{.05}$	Reject H_0
	$p_o15 = p_o25$	$p_o15 > p_o25$	$z > z_{.05}$	Cannot reject H_0
	$p_o25 = p_o35$	$p_o25 > p_o35$	$z > z_{.05}$	Reject H_0
P(NC|R)	$p_o5 = p_o15$	$p_o5 < p_o15$	$z < -z_{.05}$	Reject H_0
	$p_o15 = p_o25$	$p_o15 > p_o25$	$z > z_{.05}$	Cannot reject H_0
	$p_o25 = p_o35$	$p_o25 < p_o35$	$z < -z_{.05}$	Reject H_0

$P(NC|R)$ is of interest. Traditionally, quality control management regards inspectors as being equally able to identify either conforming product or nonconforming product equally well. This view, if valid, would mean that both patterns would have occurred at about the same level of the ordinate scale. This is the case only in the 35% treatment of p_o. At this point, it appears that both decisions are equally good. However, testing the null hypothesis of no difference as indicated below results in a significant difference in the opposite direction to that already observed; that is, the decision to reject is of better quality than the acceptance decision. The null and alternate hypotheses tested at $p_o = 35\%$ were

$$H_0: \; P(NC|R) = P(C|A)$$

$$H_1: \; P(NC|R) > P(C|A)$$

$$\text{Critical region:} \quad z > z_\alpha$$

with the result that calculated z is 1.790. As $z_{.05}$ is 1.645, the null hypothesis is rejected in favor of the alternative.

The implication of changing performance level relative to p_o is significant for quality control management. It means that attention must be given to measuring performance quality for combinations of p_o and inspection tasks and then seeking ways and means for improving unacceptable performance.

Vigilance Performance: McCornack's Criteria

The measures of A_1, A_2, and A_3 versus the five time blocks are plotted in Figure 11 in order to determine whether there exists a performance decrement similar to that observed in vigilance tasks.

Recalling that a vigilance performance decrement is a degradation in performance during the first thirty minutes of the task after which performance stabilizes at some lower level, only the pattern of A_3 for p_0 of 15% is observed to meet the criterion. The decrement during the first thirty minutes is approximately 19% although it is followed by a recovery of 10% during the next twenty minutes. While the magnitude of the fluctuations in the A_3 pattern for p_0 at 5% are considerably greater than the minimum requirement of 10%, the observed pattern does not conform to the usual configuration of the vigilance decrement pattern.

Because of the apparent recovery during the last twenty minutes in the A_3 pattern of performance at 15%, the question of whether this pattern of performance is representative of the typical vigilance performance pattern needs testing. Two tests are required. The first is to determine whether the decrement is significant and the second (assuming H_0 is rejected) is to determine whether the recovery is significant when compared to the lowest level observed.

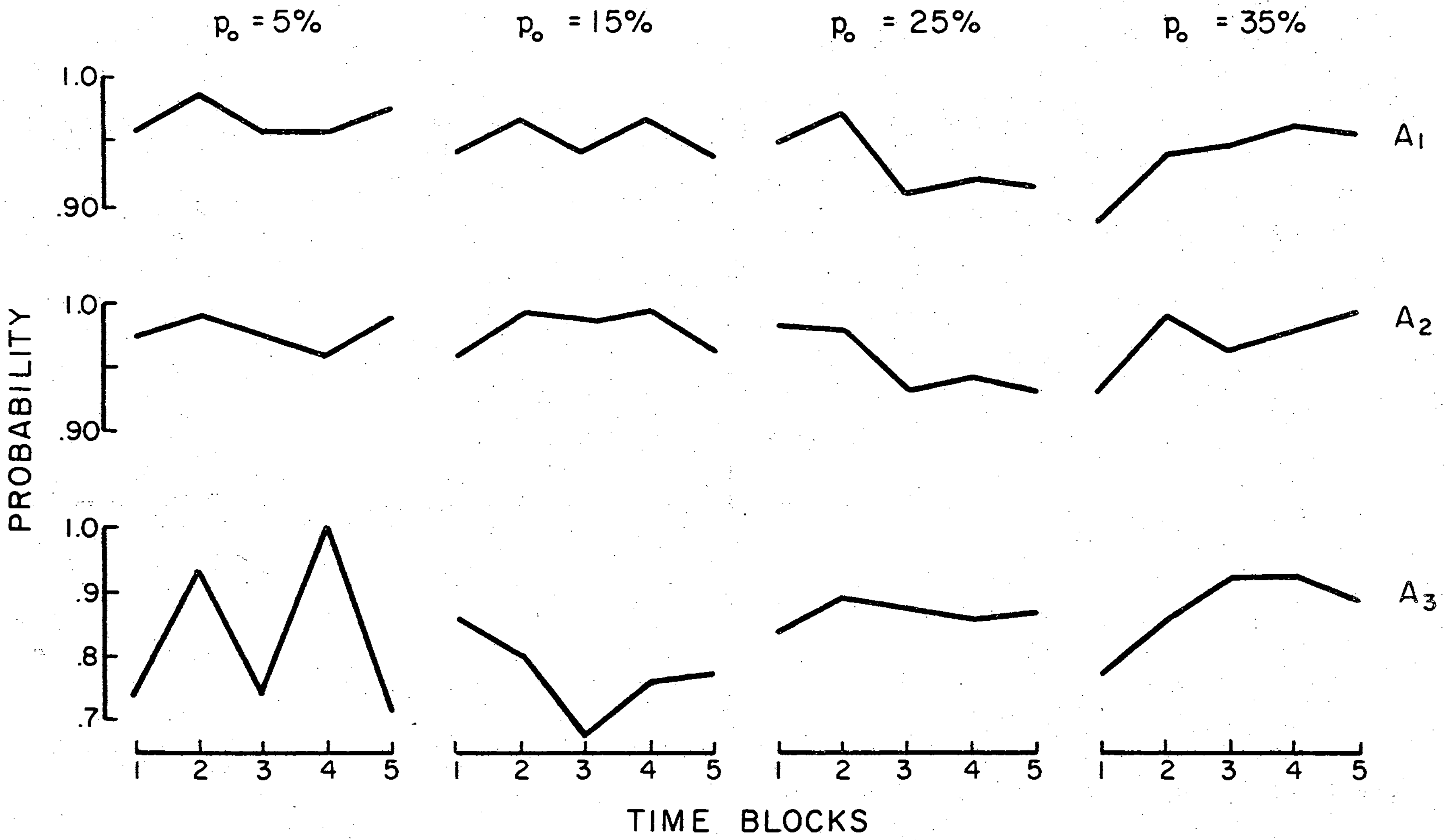


Figure 11. McCornack's Criteria Versus Time Blocks

In the first test the null and alternate hypotheses are

$$H_0: \ A_3\,\text{max} = A_3\,\text{min}$$
$$H_1: \ A_3\,\text{max} > A_3\,\text{min}$$

where the maximum value of A_3 is that observed for time block one and the minimum value is that observed for time block three. The calculated value of z was 2.44. As $z_{.05}$ is 1.645, the null hypothesis was rejected in favor of the alternate hypothesis. In the second test, the null and alternate hypotheses are

$$H_0: \ A_3\,\text{min} = A_3\,\text{max}$$
$$H_1: \ A_3\,\text{min} < A_3\,\text{max}$$

where A_3 min is for time block three and A_3 max is for time block four. The critical region is $z < -z_{.05}$. The calculated value of z was -1.020. As $-z_{.05}$ is -1.645, the null hypothesis cannot be rejected. Therefore, it is concluded that the pattern of A_3 over time at p_0 of 15% is representative of a vigilance performance decrement pattern.

If performance at the 15% treatment is typical of vigilance performance, then performance at the 5% treatment should also exhibit evidence of vigilance performance as the stimulus provided by defective items is even less; however, the observed results were contrary to those predicted. A problem in detection performance over time at the 5% level persists. The only significant difference is between time blocks three and four (z = 2.080), but the pattern of fluctuation in the control chart sense is wide with a saw tooth effect which suggests that the sample may have been drawn from two different distributions.

Two possible explanations for this situation occur. The first is that some bias has accidently been built into the five percent sample of test materials, and the second is that as the inspectors may not have been accustomed to such a low fraction defective, several of them had trouble adjusting to the detection requirements. As the test samples were carefully constructed and as the experiment was subject paced in that not everyone would inspect exactly the same items is a given time block, the second possibility seems more likely.

The question of two populations of inspectors will be considered in the following chapter.

Vigilance Performance: Bayesian Criteria

As shown in Figure 12, plots were constructed for the change in $P(C|A)$ and $P(NC|R)$ over time to determine whether there was a performance decrement similar to that expected in a vigilance task. To make this determination, it was assumed that the operational definition of a vigilance performance decrement presented in Chapter III is applicable to the Bayesian criteria.

Two patterns conform to the requirements of the operational definition. Both are for the measure $P(NC|R)$ and are for the 5% and 25% treatments of p_o.

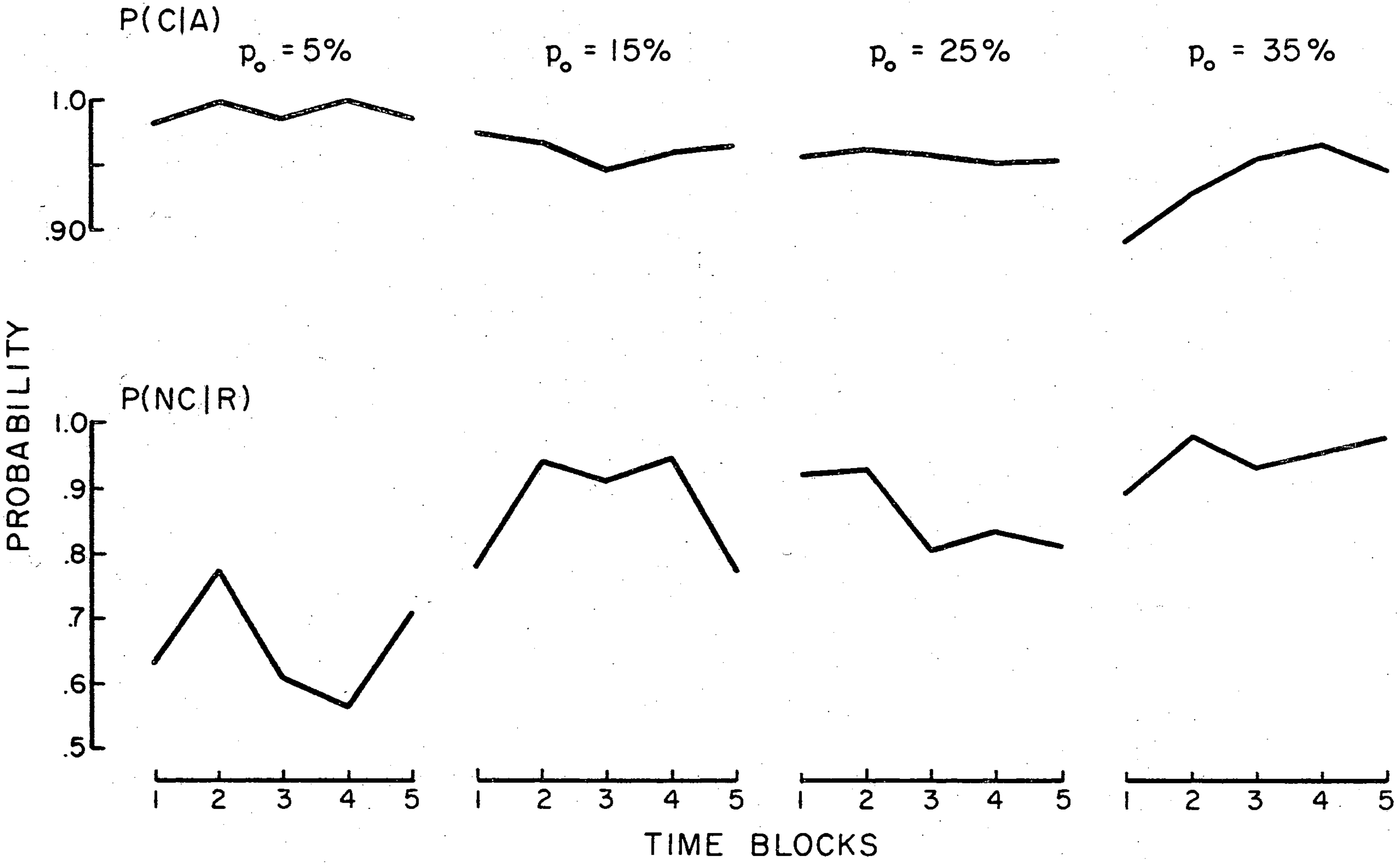


Figure 12. Bayesian Criteria Versus Time Blocks

This additional suggestion of a vigilance performance decrement is of interest as it indicates that more attention should be given to the examination of inspector performance over time if desirable performance levels are to be achieved.

Index of Detectability: d′

Figure 13 shows the intersection of the two probability density functions with the criteria x_c and d′ identified for each p_o treatment.

As d′ is defined as an index of detectability for a given signal for a given observer, it is used here as an index of the detectability of the test samples for the population of inspectors. The relatively close agreement suggests that the sample materials were uniform and the inspectors, as a group, did not have more difficulty from one sample to another in making their inspection.

Also of interest are the differences in x_c for each treatment. It can be seen in Figure 13 that the criterion for p_o of 25% is the lowest of the four treatments. For some reason, the inspectors used a different criterion for this treatment. Whatever the reason for the change, it does not appear to be the result of application of the V and K values provided them prior to the experiment. Application of these values to their selection of a criterion would have resulted in substantially different x_c and, therefore, A_2 over the given range of p_o. This may

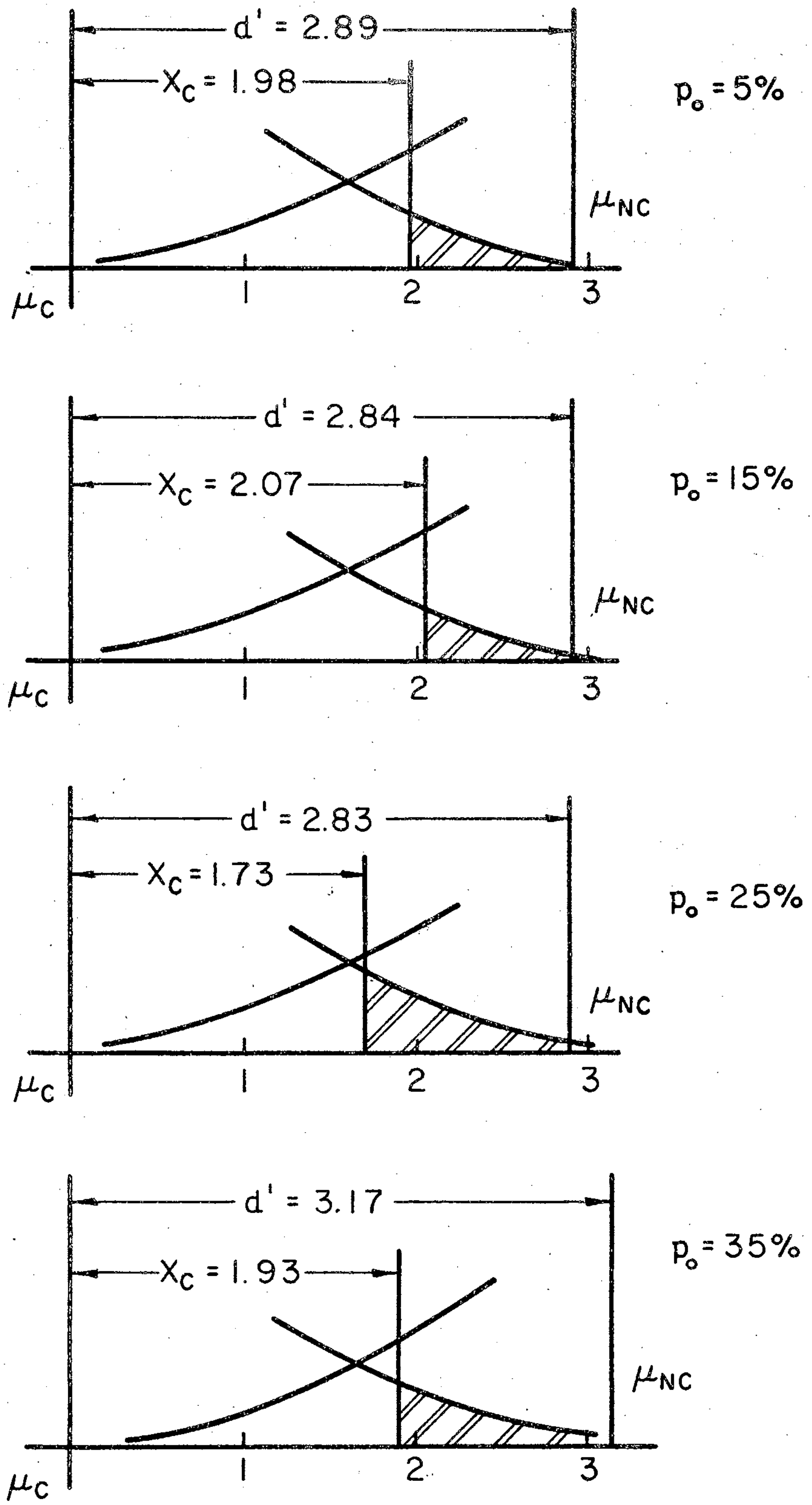


Figure 13. d′ and x_c for Each p_o

be seen by computing β for each treatment and comparing it to the ratio of the ordinate of $f_{NC}(x)$ to $f_C(x)$ at x_c. If the inspector's criterion were optimal, β should equal this ratio.

As V for correct decisions was one and K for incorrect decisions was three, V plus K equals four in both cases and so β is simply the ratio of P(C) to P(NC). The differences between the ratio and β are tabulated below.

$f_{NC}(x)$	β	$f_{NC}(x)/f_C(x)$	Difference
.05	19.00	4.74	-14.26 ✓
.15	5.67	6.36	+ 0.69
.25	3.00	2.44	- 0.56
.35	1.86	3.31	+ 1.45

The major difference between the criterion used and the optimal criterion was for the 5% treatment. To attain a β of 19.00 for the given V and K, x_c would have to be 2.465. This value was obtained by a trial and error comparison of ratios of ordinates of the normal curve with the constraint that d' be 2.89. This would mean that the observed A_2 would be approximately 0.993 and the observed A_3 would be 0.665. In other words, defect detection would suffer in order to improve the correct identification of conforming items. This view of the inspection task is diametrically opposed to the view of the inspection objective cited earlier which is to maximize A_3.

From the above it is concluded that whatever payoff the inspectors used, it was not that specified by the experimenter. One reason for this may be that the concept of payoff is not usually applied in this manner to inspection tasks. Generally, only a penalty is applied in the form of disciplinary action by supervision in the event that nonconforming items are missed by the inspector and then detected at some subsequent point in the process. A question in the application of the payoff concept to industrial inspection is the difficulty in determining realistic benefits and costs for the decision outcomes. This question is discussed briefly in the final chapter.

Apparent Fraction Defective

The apparent fraction defective p_3 was defined as the value of product defectiveness perceived by the inspector as the result of occurrence of both kinds of inspection errors, p_1 and p_2.

Figure 14 is a plot of the normalized deviations of p_3 from p_0 for each time block. It can be seen in this plot how the inspectors adjusted their level of performance relative to the true fraction defective of the product as they gained experience with the sample. With p_0 taken as zero, it can be seen that at $p_0 = 5\%$ the average performance level was 0.95 standard deviations above p_0; whereas, for the 15% and 35% treatments, as the average performance level was 0.93 and 1.0 standard deviations, respectively, below the true p_0. The average performance

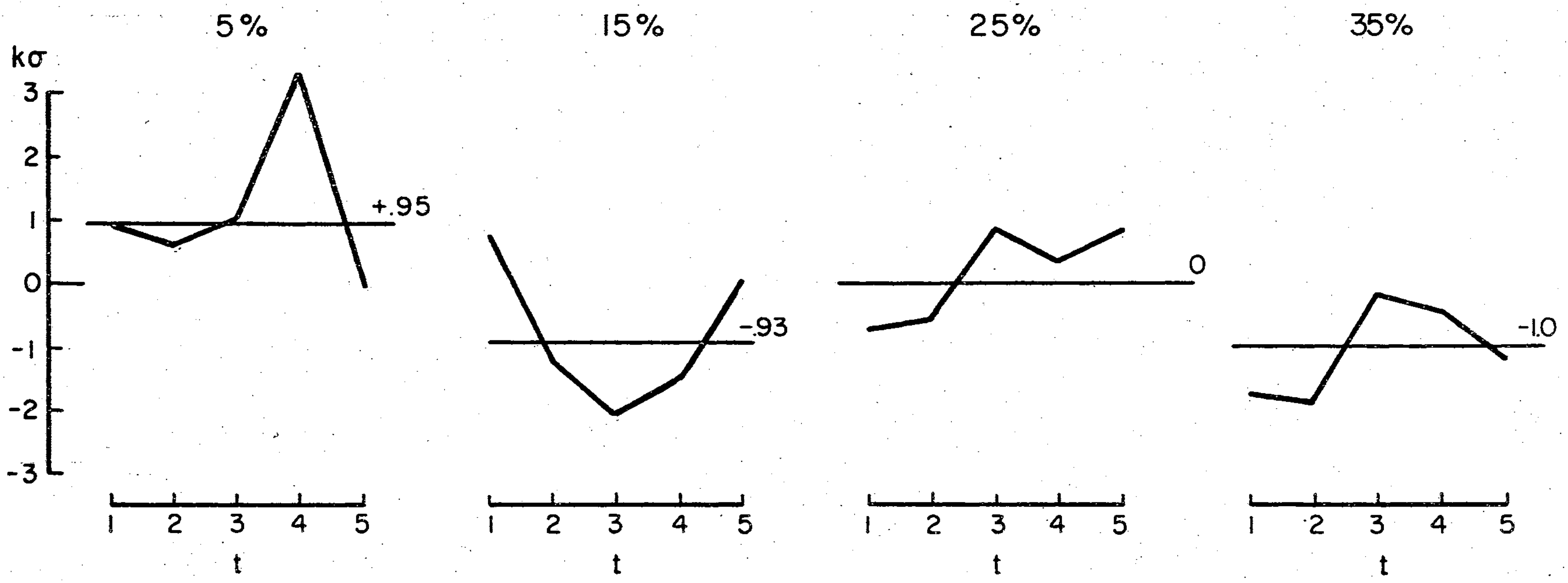


Figure 14. Normalized Deviation From p_o for Time Blocks

at the 25% treatment was at 25%. The data supporting Figure 14 are in Appendix G.

The pattern at 15% is similar to that of the vigilance performance decrement pattern and encompasses a net shift of 2.86 standard deviations from the maximum point at period one to the minimum point at period three. The conclusion is that p_3 shows evidence of a vigilance decrement in the same manner as A_3 and $P(NC|R)$.

The major point of interest is the inspector's error in estimating the true fraction defective of the inspection lot. At p_0 of 5%, the tendency was to overestimate, while at 15% and 35%, the tendency was to underestimate. Even if the estimates were, on the average, correct as at 25%, the decisions made would be for the wrong reasons because of the effect of p_1 and p_2 on the decision process.

A detailed example of these error effects in a sampling situation is included in Appendix H using the results of this experiment. In the example, the inspector's operating characteristic curve for the sampling plan used gives more protection against poor quality than the design operating characteristic curve. The use of any sampling plan implies that management has determined a satisfactory breakeven point for sampling error and cost of inspection. The effect of inspector error, which is not taken into account in the design of sampling plans, is, in this instance to increase costs of inspection above those expected. The reason is that the higher rejection rate requires that a greater than expected number of lots must be screened.

Summary

The criteria A_2 and A_3 deal with correct decisions relative to Type I and Type II error, respectively, but do not relate these errors to the total situation. The conditional probability statements $P(C|A)$ and $P(NC|R)$ treat the acceptance decision and the rejection decision separately and account for the errors made relative to each decision. No single one of these criteria is best for measuring performance per se. Each is good relative only to the objective for which a performance evaluation is being made. And certainly none of these criteria say how much better or worse performance is unless some utility measure be applied. To accomplish this one must consider the total decision process along with the benefits incurred for correct decisions and costs incurred when incorrect decisions are made. While the criterion A_1 is based on a probability statement which defines the probability of a correct decision given that one has been made, it in itself does not have any utility value assigned. While p_3 is an indication of the fraction defective perceived by the inspector and permits, in terms of the average total inspected, an indirect assessment of costs, it does not provide information about the magnitude of either p_1 or p_2.

The best criteria are determined to be d′, x_c and E(V) for the decision process, as not only are the kinds and costs of error included and related in the performance evaluation, a reference point, x_c, is provided which, when compared to the optimal criterion indicates both the direction and magnitude of the improvement needed.

It was suggested earlier that some of the difficulties observed with the 5% test sample may have been because part of the inspector group had more difficulty with this treatment than with the others. In other words, for p_0 equal to five percent there were two populations of inspectors, one able to deal effectively with the treatment, the other not. This question is analyzed in the following chapter.

CHAPTER VIII

TWO POPULATION ANALYSIS

In the previous chapter a question concerning the possibility of two inspection groups relative to performance at the 5% treatment. Two approaches to the analysis of the data are available. The first, or classical method, utilizes the chi-square statistic while the second method utilizes signal detection criteria. The classical method requires analysis with reference to the usual inspection objective, maximizing detection of defects whereas the signal detection method relates to the utility of the process in terms of the payoff matrix. Both approaches will be described below.

Classical Method

The question of two populations of inspectors with respect to the 5% sample may be examined by testing for differences among the inspectors using the χ^2 statistic and a 2 x 7 contingency table. The criterion used is A_1 as it accounts for both correct decisions.

Inspector	A	B	D	F	G	H	I	Total
Failures	21	3	10	5	6	4	4	53
Total	260	260	205	260	260	260	260	1765
P	.0808	.0115	.0488	.0192	.0231	.0154	.0154	.0300

The calculated χ^2 is 34.0. Since $\chi^2_{.05;6} = 12.84$, the null hypothesis is rejected and it is concluded that there are significant individual differences among the inspectors for the 5% treatment.

Inspection of the raw data reveals that three inspectors - A, D, and F, made the greatest number of A_2 type errors while making the least number of A_3 type errors.

	A_2	A_3
Total Errors	38	15
ADF Share	35	1
ADF %	92.2	6.67
BGHI %	7.8	93.33

With reference to the objective of maximizing the detection of defects, group ADF performs better than group BGHI at the 5% treatment, but at the expense of more errors in correctly identifying conforming items. The differences between the two groups are significant when tested using the chi-square statistic. The results are summarized below and the supporting contingency tables are in Appendix E.

Criterion	H_0	H_1	^{2}	$^2_{.05;1}$	Decision
A_2	ADF=BGHI	ADF≠BGHI	5.96	3.84	Reject H_0
A_3	ADF=BGHI	ADF≠BGHI	8.53	3.84	Reject H_0

Signal Detection Method

In the traditional sense, group ADF would be thought to be the better group of inspectors. However, with respect to the payoff matrix prescribed for this experiment a different conclusion is possible. In Figure 15 are plotted d′ and x_c for each of these groups for the 5% treatment. The index of detectability, d′, is approximately the same for both groups; whereas, the difference in criterion, x_c, is 1.136 standard deviations.

Now, the question of performance quality at p_0 equal to five percent may be answered by saying that while both groups were equally able to detect defects (approximately the same d′), each group used a different criterion and this difference accounts for the performance differences observed.

The optimal criterion was defined earlier as that value of x_c where $\beta = 19.0$. The ratio of the ordinates at the observed x_c for each group are

	$f_{NC}(x)/f_C(x)$
ADF	0.570
BHGI	34.50

Group ADF is less than β while group BHGI is greater than β. By trial and error the optimal criterion was

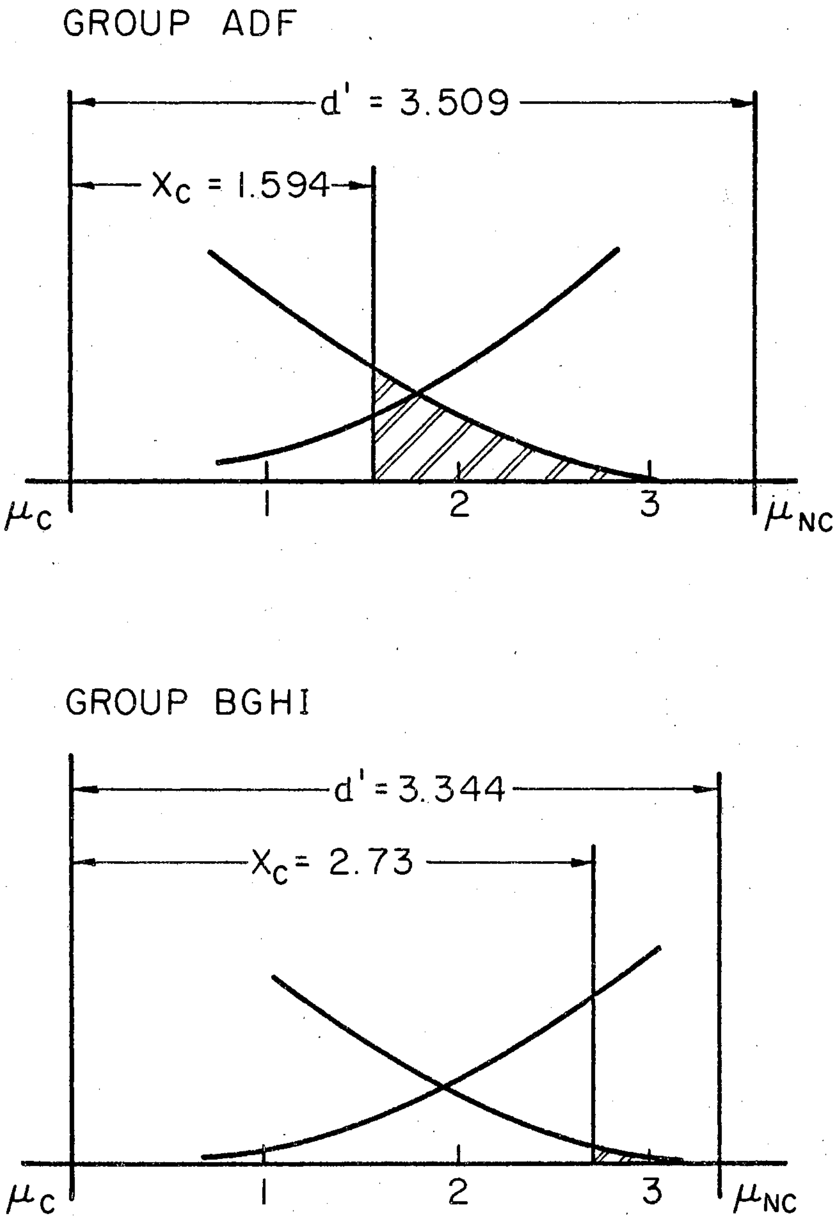


Figure 15. d′ and x_c for Two Inspector Groups

determined to be approximately 2.57 for d′ = 3.425. Group ADF needs to adjust its criterion to the right by 0.98 standard deviations, while group BGHI needs to adjust its criterion to the left by 0.16 standard deviations. Thus, it would appear that group BGHI used a criterion much closer to that desired than the other group.

Analysis of individual differences for the other treatments gave the following results:

P_o	χ^2	$\chi^2_{.05;6}$	Decision
15	8.90	12.84	Cannot reject H_o
25	7.69	12.84	Cannot reject H_o
35	60.60	12.84	Reject H_o

Examination of the data for the 35% treatment revealed that inspectors F and H performed significantly different from the rest of the group, while differences among the rest of the group were not significant. As inspector F performed well, failing to identify correctly only four conforming items and inspector H performed poorly, failing to identify eighteen conforming items and nineteen defective items, there was no further attempt at subgroup comparisons. The values of x_c and d′ for each inspector for each treatment are included in Appendix C.

Inspector H inspected the 35% sample during the last period of the day. He looked at his watch a number of times and appeared to be concerned about quitting on time. This loss of attention may be the cause of his poor performance. The difference in performance cannot be explained except to say that his index of detectability was high for both this sample and the 5% sample (5.98 and 6.06, respectively), thus his capability for this particular defect is better than the rest of the group.

Summary

This chapter has presented an analysis of inspector performance using both the classical and signal detection approaches. For the 5% treatment, two groups of inspectors were identified. By relating performance and payoff using the signal detection approach, a different conclusion as to desirable performance was reached than with the classical method.

The following chapter sums up the research and recommends several areas for additional study.

CHAPTER IX

SUMMARY AND CONCLUSIONS

This final chapter is composed of three sections. First is presented a summary of the research, its applicability, and the major conclusions. The second section describes some of the problems in the design and conduct of the experiment. The final section recommends potentially worthwhile areas for additional study.

Summary

This dissertation reported an experimental study of inspector accuracy using industrial electronics inspectors as subjects. A visual, subject-paced task was used in which the subject gave a "yes-no" response to a stimulus presented for evaluation. The experimental task simulated a real-world task in that the stimuli used were the same as those with which the subjects had daily contact. The performance measures were expressed in terms of success in giving the correct yes-no responses.

The threefold objective of the research was to

1. Determine the effect of the a priori probability of defectiveness on an inspector's decision to accept conforming items and to

reject nonconforming items.

2. Determine whether a vigilance decrement could be observed for a task meeting vigilance task requirements.
3. Determine which of several available measures was most effective in describing inspector performance.

These objectives were achieved as evidenced by the conclusions delineated below. As a particular class of inspectors was used in a visual, subject-paced task, the results may not be applicable in all cases. For instance, they should not be applied to a task that is externally paced, as by a production line, nor should they be applied to a task requiring measurement, such as machine parts inspection. The results are thought to be generally applicable to visual inspection tasks meeting the requirements of a vigilance task description as presented in Chapter III.

The major conclusions which follow are expressed in terms of the several performance measures described in the earlier chapters. These conclusions are:

1. Inspector performance relative to detection of defects (A_3) varies linearly with p_0.
2. Inspector performance relative to correct identification of good product (A_2) may be unaffected by p_0. In three treatments - 5%, 15%, and 35% - this was true. The

observed difference in the fourth treatment, 25%, is in part explained by the choice of a different criterion, x_c.

3. The vigilance decrement for p_o of 15% using A_3 and for the p_o of 5% and 25% using $P(NC|R)$, indicates that the consideration of the possibility of vigilance performance for low values of p_o in industrial inspection tasks is worthwhile. A short break or interruption after about fifteen minutes of a vigilance type task would likely reduce the probability of a performance decrement.
4. Inspectors had difficulty in estimating the true fraction defective (p_3 versus p_o) as evidenced by overestimation in one case and underestimation in two others. In a sampling situation, this variation caused by decision error will affect the protection offered by the sampling plan.
5. Bayesian measures are more useful in describing the quality of either decision than are the measures A_2 and A_3, since they account for all the information relative to the decision for which they are written.

6. The signal detection measures d′ and x_c are more useful than the other measures in performance evaluation as they not only provide for relating performance to payoff, but also indicate the magnitude and direction of improvement required.
7. A major difference in choice of x_c was observed between two inspection groups for the 5% treatment. This difference was not in evidence at the 15% and 25% treatments. At the 35% treatment, two inspectors performed differently from the rest of the group, one scoring high and the other low.

Problems in Experimentation

A number of problems existed in the conduct of this experiment which need to be identified.

The first problem dealt with the availability of subjects. The participating inspectors were taken off the job either one or two at a time to perform the tests. Since their availability was a function of the work load in the department, it was not always possible to obtain the particular one required for the experimental design. As a result, to maintain the constraints of order and time of day it was necessary to stretch the experiment over a longer calendar period than was anticipated.

The second problem was the establishment of the payoff values V and K. The values established in this experiment were the result of a consensus of the views of three quality control engineers, three inspection supervisors, and a quality control manager. If possible, a more precise determination of payoff should be made in terms of actual dollars, although this may be very difficult because of the intangible costs which can apply. Also, consideration of payoff requires that a clear statement of inspection objectives be made for the task along with a specification of the risks (Type I and Type II error) acceptable for the decision process. Ideally, the criterion (x_c) should be one which permits these risks and no more. By varying the payoff over the course of the experiment, the experimenter could learn by comparison of x_c's whether his specification of payoff was being followed in choice of criterion by the inspectors.

The third problem relates to the control of the introduction of the stimulus. All vigilance research reviewed by the experimenter used externally paced tasks; whereas, most visual inspection tasks are subject paced. For the purpose of this research, it was assumed that this difference would not be a problem since in the pilot study it was demonstrated that rate of presentation did not affect performance. The experimental task was designed based on a predicted average rate of four wires inspected per minute. Although one subject inspected at better than twice that rate, the quality of her performance did not differ from the rest of the group.

Knowledge of the a priori probabilities of defectiveness is the subject of the fourth problem. Signal detection theory presumes a knowledge of this value by the observer prior to making the observations. However, the inspector does not usually know the fraction defective of the inspection lot prior to his inspection. His only cue may be his past experience with a given production department or individual operator for the task at hand. In this experiment a cue was provided in the form of a pretest sample with the same fraction defective as the test sample in an effort to minimize uncertainty during the first few minutes of the task.

The fifth problem relates to the recovery observed during the last twenty minutes of the experiment. As the inspector was able to see that the task would soon end as he inspected the 200th item, it is assumed that this stimulated him to become more attentive during the final few minutes. As this effect was greater than anticipated, it is recommended that future experimenters furnish enough test materials to last twenty or thirty minutes beyond the time span of interest.

The final problem deals with the lack of ability to specify d′ for an ideal observer for the test materials. A possibility might be to use the ratio of the area of the defect to the total area inspected as being comparable to the signal to noise ratio. This is essentially the same approach as that taken using electrical energy in the usual signal detection task.

Recommendations for Additional Study

A number of open questions requiring additional study have been generated during the course of this research. Also the research itself was not broad enough in scope to consider every pertinent aspect in great detail. The following areas of additional research are suggested as being of interest in extending knowledge of the accuracy of industrial inspectors.

1. Evaluation of performance using the measure A_2. Specifically, why was not a vigilance decrement observed, and why was there a difference in criterion (x_c) for the treatment at p_0 equal 25%?
2. Evaluation of the Bayesian criteria. Either specification of an operating definition for vigilance performance using these criteria or experimental justification of the definition used in this experiment would be of interest. With reference to the question of the performance plateau at the midrange of the given p_0, the question might be asked whether this was peculiar to this

experiment rather than a true performance characteristic.

3. <u>Evaluation of the signal detection criteria</u>. Knowledge of the distribution of d′ for a group of inspectors relative to a particular task would be of interest in deciding whether or not differences were significant between indices obtained for other inspectors and a standard population. However, the availability of a method for defining d′ for the ideal observer as noted in problem six above, would eliminate this requirement.

4. <u>Evaluation of sampling plan design with respect to error possibilities and location of the criterion x_c</u>. Sampling plans should be designed by taking into account the risks of decision error as well as those of sampling error. In some inspection tasks, error probabilities may be small while in others they may be so large as to be prohibitive. Development of a design methodology to account for decision errors would be highly useful.

BIBLIOGRAPHY

1. Adams, C. K. Reliability of Subject Judgement in the Inspection of Hard Red Winter Wheat. Unpublished Master's thesis, Kansas State University, Department of Psychology, 1960.

2. Baker, C. H. "Maintaining the Level of Vigilance by Means of Artifical Signals," Journal of Applied Psychology, Vol. 44 (1960), p.336.

3. Buckner, D. N., and J. J. McGrath. Vigilance: A Symposium. New York: McGraw-Hill, 1963.

4. Carter, C. W. "Quality Control of Visual Characteristics," ASQC Convention Transactions, 1957, 623-634.

5. Cochran, W. G., and G. M. Cox. Experimental Designs. 2nd Ed. New York: John Wiley and Sons, p. 105.

6. Colquhoun, W. P. "The Effect of a Short Rest Pause on Inspection Efficiency," Ergonomics, Vol. 2, No. 4 (1959), p. 367.

7. ______________. "Temperament, Inspection Efficiency and Time of Day," Ergonomics, Vol. 3, No. 4 (1960), p. 377.

8. ______________. "The Effect of Unwanted Signals on Performance in a Vigilance Task," Ergonomics, Vol. 4, No. 1 (1961), p. 41.

9. Frankmann, J. P., and J. A. Adams. Theories of Vigilance. USAF CCDD Tech. 60-25, 1960.

10. Harris, D. Evaluation of an Experimental Chassis Inspection Method TM-31-040-22-34, Autonetics Division, North American Aviation, Inc., Anaheim, California, 1963.

11. _________. Exploratory Study of Module Assembly Inspection, TM-31-040-22-35. Autonetics Division, North American Aviation, Inc., Anaheim, California, 1963.

12. ___________. *Research on Etched Circuit Board Inspection Performance*, T4-577/3111 Autonetics Division, North American Aviation, Inc., Anaheim, California, 1964.

13. ___________. *Development and Validation of an Aptitude Test for Inspectors of Electronic Equipment*, T4-623/3111, Autonetics Division, North American Aviation, Inc., Anaheim, California, 1964.

14. ___________. *Study of Human Factors in Ceramic Printed Circuit Inspection*, T4-1059/3111, Autonetics Division, North American Aviation, Inc., Anaheim, California, 1964.

15. ___________. *ESSD Inspector Workmanship Program: I. Inspection Proficiency and Training Needs*, T4-1471/3111, Autonetics Division, North American Aviation, Inc., Anaheim, California, 1964.

16. ___________. *The Effect of Equipment Complexity on Inspection Performance*, T4-1698/3111, Autonetics Division, North American Aviation, Inc., Anaheim, California, 1964.

17. Hayes, A. S. "Control of Visual Inspection," *Industrial Quality Control*, 6, 1950, 73-76.

18. Jackson, J. E. "Effect of Inspection Errors on Waste and Quality of Outgoing Product," *Industrial Quality Control*, Vol. 14, No. 6 (1957), p. 5.

19. Jacobson, H. J. "A Study of Inspector Accuracy," Industrial Quality Control, 9, 1952, 16-25.

20. ______________. "Management's Method of Insuring Accuracy of Inspection," Transactions of the *Seventeenth National Convention of the American Society for Quality Control*, 1963, p. 277.

21. Kennedy, C. W. *Inspection and Gaging*. New York: The Industrial Press, 1957, p. 428.

22. McCornack, R. L. *Inspector Accuracy: A Study of the Literature*. SCTM 53-61 (14), Sandia Corporation, Albuquerque, New Mexico, 1961.

23. Miller, I., and J. E. Freund. *Probability and Statistics for Engineers*. Englewood Cliffs, New Jersey: Prentice-Hall, p. 31.

24. Seaborne, A. E. M. "Social Effects on Standards in Gauging Tasks," _Ergonomics_, Vol. 6, No. 2 (1963), p. 205.

25. Schilling, E. G. "The Challenge of Visual Inspection in the Electronics Industry," _Industrial Quality Control_, Vol. 18, No. 2 (1961), p. 12.

26. Swets, John, Editor. _Signal Detection and Recognition by Human Observers: Contemporary Readings_. New York: John Wiley and Sons, 1964, p. 3.

27. Thomas, L. F. "Perceptual Organization in Industrial Inspectors," _Ergonomics_, Vol. 5, No. 3 (1962), p. 429.

28. Tiffin, J., and E. J. McCormick. _Industrial Psychology_. 4th Ed. Englewood Cliffs, New Jersey: Prentice-Hall, 1958, p. 373.

29. Wald, A. _Statistical Decision Functions_. New York, N. Y.: John Wiley and Sons, Inc., 1950.

APPENDICES

APPENDIX A

TEST INSTRUCTIONS

A. Pretest Instructions

This is the first phase of a test program which is being conducted in your area during the next few weeks.

The purpose of the program is to establish some guidelines for establishing values for the accuracy of the inspection function. Your group has been selected as the pilot group in the program. Results of the testing will be reported as group results with no attempt made to identify individual performance.

Today, you will be taking some pencil and paper tests relating to visual inspection. The balance of the test program will involve a wire preparation inspection task.

Instructions for each of today's tests will be provided by the test instructor.

B. Training Instructions

This second phase of our test program involves a wire preparation inspection task. In this phase you will be asked to inspect a series of wire ends that have been thermally stripped, and to determine whether any of these three types of defects exist:

1. Scraped strands such that the copper is showing
2. Nicked or ringed strands
3. Cut (missing) strands.

You do not need to squawk other types of defects such as birdcaging, improper lay, insulation defects, etc.

It is required that you do not use a magnifying glass in making your inspection.

The test conductor will give you instructions prior to your inspection of each set of wires.

B.1 1st Wire Group

In the first task you are required to inspect each of ten wires. After inspecting the first wire, tell the test conductor whether or not you think it is defective before going on to wire Number 2. Repeat this procedure on each wire until inspection of all ten is completed.

B.2 Payoff Criteria

Before proceeding further we will establish a system of keeping score on your performance. Whenever you accept a good wire or reject a defective wire, you will receive one point. But, whenever you accept a defective wire or reject a good wire, you will have three points subtracted from your score. At the conclusion of the following test groups the test conductor will compute your score and tell you how you did. The best possible score will be 10, the worst possible score will be a minus 30.

B.3 2nd Wire Group

In this task you are required to inspect each of ten wires. However, instead of telling the test conductor whether or not the wire is defective, write the letter/number of any defective wire on a sheet of paper. After you have completed your inspection of all ten wires, the test conductor will compute your score and tell you what it is.

B.4 3rd Wire Group

In this inspection task you are required to inspect each of 20 wires. Record the letter/number of each defective wire on the sheet of paper as before. When you have completed your inspection, the test conductor will compute your score as before.

B.5 4th Wire Group

In this last inspection task, you are again required to inspect each of 20 wires. Record the letter/number of each defective wire on the sheet of paper as before. When you have completed your inspection, the test conductor will compute your score as before.

C. General Instructions: Wire Preparation Inspection

The third phase of our test program involves a wire preparation inspection task. During the next few days you will be asked to inspect four samples of wire ends that have been thermally stripped. Each sample contains a different number of defective wires. For each sample you are to determine whether any of these three types of defects exist:

1. Scraped strands such that copper is showing.
2. Nicked or ringed strands.
3. Cut (missing) strands.

You do not need to squawk other types of defects such as birdcaging, improper lay, insulation, burned strands, etc. Be sure to inspect the total length of the exposed wire.

It is required that you do not use a magnifying glass in making your inspection.

You will notice that each wire has a small piece of insulation left on the end in an effort to prevent the lay from being disturbed. Please try not to remove or otherwise disturb this protective covering.

Your score for the four samples will be determined as before. Whenever you accept a good wire or reject a bad wire, you will receive one point. But, whenever you accept a defective wire or reject a good wire, you will have three points subtracted from your score. After you have inspected all four samples, the test conductor will compute your score and tell you how well you did. Please do not discuss your performance with the other inspectors until after the test program is over. Do not talk during the test.

The test conductor will give you instructions prior to the inspection of each sample.

C.1 Pretest Sample

Before beginning your inspection of the test sample you will be given a practice inspection sample of 20 wires. Record the number of each defective wire on the squawk sheet. After you have completed your inspection of the 20 wires, the test conductor will compute your score.

You will have five minutes to complete the practice inspection.

Do you have any questions?

When the test conductor gives you the signal, you may begin your inspection.

C.2 Wire Sample A-M

In this task you are required to inspect 260 wires to determine whether or not there are scraped, nicked, ringed or cut strands.

The sample consists of thirteen groups of wires lettered A through M. Each group contains twenty wires numbered 1 through 20. When you have determined that a defect exists, write the appropriate letter and number of the defective wire on the squawk sheet furnished you by the test conductor. For example, if wire Number 14 in group B is defective, you will write B14 on the squawk sheet.

You will have seventy-five minutes to make your inspection. Work steadily and accurately.

Do you have any questions?

When the test conductor gives you the signal, you may begin.

C.3 Wire Sample N-Z

In this task you are required to inspect 260 wires to determine whether or not there are scraped, nicked, ringed or cut strands.

The sample consists of thirteen groups of wires letter N through Z. Each group contains twenty wires numbered 1 through 20. When you have determined that a defect exists, write the appropriate letter and number of the defective wire on the squawk sheet furnished you by the test conductor. For example, if wire Number 14 in group R is defective, you will write R14 on the squawk sheet.

You will have seventy-five minutes to make your inspection. Work steadily and accurately

Do you have any questions?

When the test conductor gives you the signal, you may begin.

C.4 Wire Sample AA-MM

In this task you are required to inspect 260 wires to determine whether or not there are scraped, nicked, ringed, or cut strands.

The sample consists of thirteen groups of wires lettered AA through MM. Each group contains twenty wires numbered 1 through 20. When you have determined that a defect exists, write the appropriate letter and number of the defective wire on the squawk sheet furnished you by the test conductor. For example, if wire Number 14 in group BB is defective, you will write BB14 on the squawk sheet.

You will have seventy-five minutes to make your inspection. Work steadily and accurately.

Do you have any questions?

When the test conductor give you the signal, you may begin.

C.5 Wire Sample NN-ZZ

In this task you are required to inspect 260 wires to determine whether or not there are scraped, nicked, ringed, or cut strands.

The sample consists of thirteen groups of wires lettered NN through ZZ. Each group contains twenty wires numbered 1 through 20. When you have determined that a defect exists, write the appropriate letter and number of the defective wire on the squawk sheet furnished you by the test conductor. For example, if wire Number 14 in group RR is defective, you will write RR14 on the squawk sheet.

You will have seventy-five minutes to make your inspection. Work steadily and accurately.

Do you have any questions?

When the test conductor gives you the signal, you may begin.

APPENDIX B

INSPECTION DATA SHEET

Inspector ______________________________ Date ___________

Test No. _____________ Time Began ______ Time Ended ______

Write the number of any nonconforming item in the space provided below. In the event that you need additional space, you may use the reverse side of this sheet.

APPENDIX C

SUMMARY OF PERFORMANCE MEASURES
FOR SEVEN INSPECTORS

Inspector	p_o	A_1*	A_2*	A_3*	P(C|A)*	P(NC|R)*	x_c	d′
A	5	919	919	923	996	375	1.40	2.83
	15	942	959	846	974	785	1.74	2.76
	25	963	975	925	975	923	1.96	3.40
	35	965	973	951	975	951	1.43	3.08
B	5	988	1000	769	988	1000	4.00	4.74
	15	973	995	846	974	972	2.58	3.60
	25	942	943	938	980	848	1.58	3.12
	35	962	983	923	961	965	2.12	3.55
D	5	951	948	1000	1000	500	1.63	5.63
	15	950	972	820	969	843	1.91	2.83
	25	912	939	830	945	818	1.55	2.51
	35	938	988	846	923	975	2.26	3.28
F	5	980	980	1000	1000	722	2.06	6.06
	15	962	982	846	975	893	2.10	3.12
	25	923	933	892	965	818	1.50	2.74
	35	984	976	1000	1000	958	1.98	5.98
G	5	976	991	692	985	818	2.37	2.87
	15	959	978	846	975	869	2.02	3.04
	25	934	984	784	932	945	2.14	2.93
	35	926	970	846	922	940	1.88	2.90
H	5	985	999	770	989	910	3.09	3.83
	15	988	1000	590	935	1000	4.00	4.23
	25	961	984	892	965	952	2.14	3.38
	35	820	865	736	860	748	1.10	1.73
I	5	984	1000	692	985	1000	4.00	4.50
	15	923	977	615	936	828	2.00	2.29
	25	928	966	807	942	885	1.83	2.70
	35	926	994	802	903	986	2.54	3.39

*Values in these columns should be divided by 1000 to obtain probabilities.

APPENDIX D-I

DATA SUMMARY FOR SEVEN INSPECTORS
SAMPLE A = 5%

Period	N_0	n_0	n_1	n_2	d_0	d_1	d_2	$n_1 + d_2$	$n_2 + d_2$	$n_1 + d_1$
1	374	351	341	10	23	6	17	358	27	347
2	410	395	391	4	15	1	14	405	18	392
3	449	426	415	11	23	6	17	432	28	421
4	279	265	254	11	14	0	14	268	25	254
5	158	150	148	2	7	2	5	153	7	150
Total	1670	1587	1549	38	82	15	67	1616	105	1564

Period	A_1	A_2	A_3	p_3	p_0	$P(C\|A)$	$P(NC\|R)$	$1-A_2$	d'
1	.959	.973	.740	.0722	.0615	.983	.630	.027	2.57
2	.989	.990	.935	.0439	.0368	.999	.779	.010	3.83
3	.960	.975	.740	.0625	.0514	.985	.608	.025	2.61
4	.961	.960	1.000	.0896	.0502	1.000	.560	.040	5.24
5	.975	.988	.715	.0444	.0444	.987	.715	.012	2.83
Total	.968	.976	.817	.063	.0491	.991	.638	.024	2.89

APPENDIX D-II

DATA SUMMARY FOR SEVEN INSPECTORS
SAMPLE N = 15%

Period	N_o	n_o	n_1	n_2	d_o	d_1	d_2	$n_1 + d_2$	$n_2 + d_2$	$n_1 + d_1$
1	377	319	305	14	58	8	50	355	64	313
2	401	340	337	3	61	12	49	386	52	349
3	403	342	338	4	61	20	41	379	45	358
4	342	294	292	2	50	12	38	330	40	304
5	162	140	135	5	22	5	17	152	22	140
Total	1685	1435	1407	28	252	57	195	1602	223	1464

Period	A_1	A_2	A_3	p_3	p_o	$P(C \mid A)$	$P(NC \mid R)$	$1-A_2$	d'
1	.942	.956	.862	.169	.154	.975	.782	.044	2.80
2	.964	.992	.804	.129	.152	.965	.943	.008	3.27
3	.941	.989	.672	.112	.151	.945	.911	.011	2.74
4	.965	.995	.760	.117	.146	.961	.950	.005	3.28
5	.940	.965	.773	.136	.136	.965	.773	.035	2.51
Total	.952	.981	.775	.132	.1495	.960	.875	.019	2.84

APPENDIX D-III

DATA SUMMARY FOR SEVEN INSPECTORS
SAMPLE AA = 25%

Period	N_o	n_o	n_1	n_2	d_o	d_1	d_2	n_1+d_2	n_2+d_2	n_1+d_1
1	384	299	293	6	85	14	71	364	77	307
2	449	333	326	7	116	13	103	436	110	339
3	387	292	272	20	95	12	83	355	103	284
4	273	203	191	12	70	10	60	251	72	201
5	172	127	118	9	45	6	39	157	48	124
Total	1665	1254	1200	54	411	55	356	1563	410	1255

Period	A_1	A_2	A_3	p_3	p_o	P(C|A)	P(NC|R)	$1-A_2$	d′
1	.948	.980	.836	.203	.221	.955	.923	.020	3.03
2	.973	.977	.889	.245	.259	.962	.936	.023	3.22
3	.919	.932	.874	.266	.246	.958	.806	.068	2.64
4	.920	.942	.858	.264	.256	.951	.834	.058	2.64
5	.915	.930	.868	.279	.262	.953	.813	.070	2.59
Total	.935	.958	.865	.247	.247	.956	.865	.042	2.83

APPENDIX D-IV

DATA SUMMARY FOR SEVEN INSPECTORS
SAMPLE NN = 35%

Period	N_o	n_o	n_1	n_2	d_o	d_1	d_2	$A_1 + d_2$	$A_2 + d_2$	$n_1 + d_1$
1	290	189	180	9	101	23	78	258	87	203
2	318	204	201	3	114	16	98	299	101	217
3	327	207	199	8	120	10	110	309	118	209
4	316	212	208	4	104	8	96	304	100	216
5	313	207	205	2	106	12	94	299	96	217
Total	1564	1019	993	26	545	69	476	1469	502	1062

Period	A_1	A_2	A_3	p_3	p_o	$P(C\mid A)$	$P(NC\mid R)$	$1-A_2$	d′
1	.890	.953	.772	.300	.348	.887	.898	.047	2.42
2	.940	.992	.859	.318	.369	.926	.970	.015	3.27
3	.946	.962	.918	.361	.367	.952	.934	.055	2.99
4	.963	.980	.923	.317	.329	.965	.960	.038	3.20
5	.955	.992	.886	.307	.339	.945	.980	.044	2.92
Total	.939	.973	.874	.321	.348	.925	.948	.027	3.17

APPENDIX E

CONTINGENCY TABLES

E.1 Contingency Tables for A_1 , A_2 and A_3

p_o	5	15	25	35	Total
A_1 Failures	54	83	109	95	341
Total	1670	1685	1665	1564	6584
$P = (1 - A_1)$	.032	.049	.065	.061	.052
A_2 Failures	38	28	54	26	146
Total	1588	1435	1254	1019	5296
$P = (1 - A_2)$	.024	.0195	.043	.0255	.276
A_3 Failures	15	57	55	69	196
Total	82	252	411	545	1290
$P = (1 - A_3)$	.183	.225	.135	.126	.152

Sample calculation:

$$\begin{aligned} A_1 \quad \chi^2 &= [(54)(.032) + (83)(.049) + (109)(.065) + \\ &\qquad (95)(.061) - (341)(.052)] \div (.948)(.052) \end{aligned}$$

$$\chi^2 = 17.44$$

$\chi^2_{.05;3} = 7.81$, hence reject H_o

E.2 Contingency Tables for A_2 and A_3 Combined Treatments

p_o	5-15	25-35	Total	χ^2
A_3 Failures	72	124	196	
Total	334	956	1290	
$P = (1 - A_3)$	.2155	.1295	.01152	12.82
A_2 Failures	66	80	146	
Total	3023	2273	5296	
$P = (1 - A_2)$	.0218	.0352	.0276	4.17

E.3 Contingency Tables for A_2 Combined Treatments

p_o	5, 15, 35	25	Total	χ^2
A_2 Failures	92	54	146	
Total	4042	1254	5296	
$P = (1 - A_2)$	.0228	.043	.0276	14.66

p_o	5	15	35	Total	χ^2
A_2 Failures	38	28	26	92	
Total	1588	1435	1019	4042	
$P = (1 - A_2)$	.024	.0195	.0225	.0228	1.77

E.4 Contingency Tables for Two Groups of Inspectors

	Group	ADF	BGHI	Total	χ^2
A_2	Failure	35	3	38	
	Total	631	957	1588	
	P	.0555	.00314	.024	5.92

	Group	ADF	BGHI	Total	
A_3	Failure	1	14	15	
	Total	36	52	88	
	P	.0278	.269	.1705	8.53

APPENDIX F

METHOD FOR CALCULATING SIGNIFICANCE OF REGRESSION

The single degree of freedom χ^2 for regression is found by first determining the quantities

$$N = \Sigma x_1 p_0 - \frac{(\Sigma x_1)(\Sigma n_1 x_1)}{(\Sigma n_1)}$$

$$D = \Sigma n_1 x_1^2 - \frac{(\Sigma n_1 x_1)^2}{(\Sigma n_1)}.$$

Where N and D are, respectively, the numerator and denominator of the regression coefficient b of P on p_0. Chi-square for linear regression then is

$$\chi^2_\ell = \frac{N^2}{\bar{P}\ \bar{Q}\ D}$$

with one degree of freedom. If a significant relationship exists, then it may be concluded that the true relation is approximately linear.

APPENDIX G

NORMALIZED DEVIATION OF p_3 FROM p_o FOR EACH TIME BLOCK

	t	p_3	p_o	$p_3 - p_o$	$\frac{p_3 - p_o}{\sigma_o} = k\sigma$
	1	.0722	.0615	+.0107	+.905
	2	.0439	.0368	+.0071	+.601
5%	3	.0625	.0514	+.0111	+.994
	4	.0896	.0502	+.0394	+3.34
	5	.0444	.0444	.0000	.0
	mean	.0603	.0491	.0112	+.95
	$\bar{n}$ = 334.0	$\bar{p}_o$ = .0491		$\bar{p}_3$ = .0603	σ_o = .0118
	1	.169	.154	+.015	+.799
	2	.129	.152	-.023	-1.22
15%	3	.112	.151	-.039	-2.06
	4	.117	.146	-.029	-1.54
	5	.136	.136	.0	.0
	mean	.132	.1495	-.0175	-.932
	$\bar{n}$ = 361.0	$\bar{p}_o$ = .1495		$\bar{p}_3$ = .132	σ_o = .0188
	1	.203	.221	-.018	-.76
	2	.245	.259	-.014	-.59
25%	3	.266	.246	+.020	+.845
	4	.264	.256	+.008	+.338
	5	.279	.262	+.017	+.718
	mean	.247	.217	.0	.0
	$\bar{n}$ = 333.0	$\bar{p}_o$ = .247		$\bar{p}_3$ = .247	σ_o = .0237
	1	.300	.348	-.048	-1.78
	2	.318	.369	-.051	-1.90
35%	3	.361	.367	-.006	-.223
	4	.317	.329	-.012	-.445
	5	.307	.339	-.032	-1.190
	mean	.321	.348	.027	-1.0
	$\bar{n}$ = 313.0	$\bar{p}_o$ = .318		$\bar{p}_3$ = .321	σ_o = .0269

APPENDIX H

INSPECTOR ERROR EFFECTS ON SAMPLING DECISIONS

A. For the purpose of this example, the single sampling plan n = 200, c = 14 was chosen to study the effect of inspection error with respect to the protection offered by the sampling plan. Table IV summarizes the computations for the operating characteristic curve of n = 200, c = 14 when no errors are made and when p_1 and p_2 errors of the magnitude indicated occur. In Figure 16 are plotted the two OC curves. Examination of these curves shows that the inspectors as a group operate to a somewhat tighter plan than is desired.

TABLE IV

O.C. CURVE COMPUTATIONS
$p_1 = 0.024$, $p_2 = 0.183$, $n = 200$, $c = 14$

p_0	$P_{a:o}$	$1-p_0$	$p_1(1-p_0)$	$p_0(1-p_2)$	p_3	np_3	$P_{a:3}$
2	1.0	98	2.36	1.634	3.99	7.98	.983
4	.983	96	2.31	3.270	5.58	11.16	.842
6	.772	94	2.26	4.904	7.16	14.32	.542
8	.368	92	2.21	6.540	8.75	17.50	.245
10	.105	90	2.16	8.174	10.33	20.66	.083
12	.070	88	2.11	9.808	11.92	23.94	.020

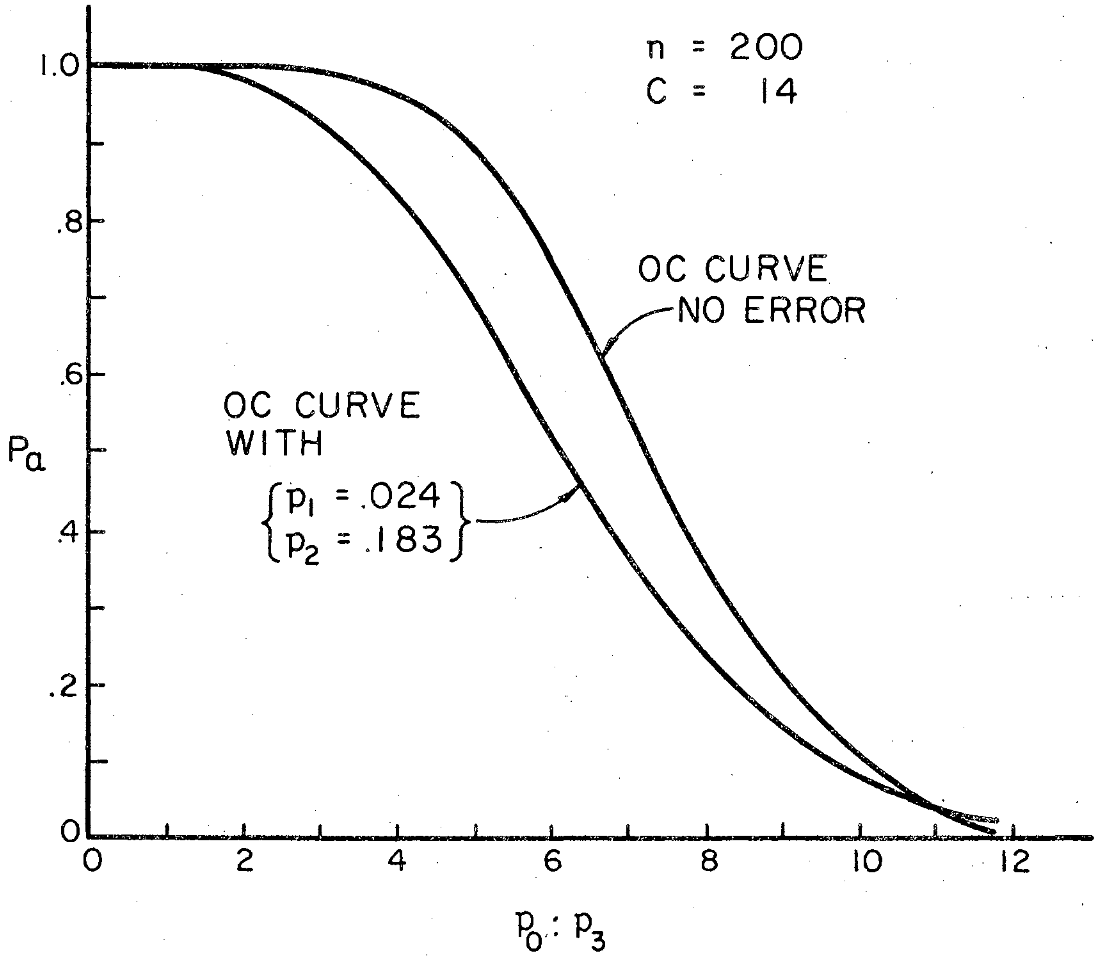


Figure 16. Effect of Inspection Error on O.C. Curves for a Single Sampling Plan

B. As there was identified two groups of inspectors at the 5% level, each with different orders of p_1 and p_2 error, their performance relative to sampling will be shown also. Table V summarizes the computations for the O.C. curves for each group and Figure 17 contains the plots of the O.C. curves.

Inspection of Figure 17 shows that group ADF has a much tighter O.C. curve than group BGHI. The latter group's curve is close to that of the design curve up to p_0 of 9% where it crosses the design curve. For an inspection lot that was, say, 10% defective the protection obtained would be slightly less than that expected. The magnitude of the difference in this example is small, but under different amounts of error it could be sufficiently large to reduce the protection offered by the plan. That is, the inspectors' curve would be to the right of the design curve.

C. The effect of the p_1 error is to shift the inspector's O.C. curve to the right when p_1 is small and to the left of the design curve when p_1 is large. In Tables IV and V, inspection of the column headed $p_1(1 - p_0)$ reveals little variation in that value compared to the other component of p_3, $p_0(1 - p_2)$. Thus, if there were no p_1 error, the effect of p_2 error would be to move the O.C. curve to the right of the design O.C. curve. As p_1 error increases, the O.C. curve moves back to the left, eventually becoming less than the design curve. Depending on the particular

TABLE V

CALCULATIONS OF PROBABILITY OF ACCEPTANCE FOR p_3 FOR TWO INSPECTOR GROUPS

Group ADF

For $p_o = 5\%$: $1 - A_2 = p_1 = .0555$

$A_3 = 1 - p_2 = .972$

p_o	$(1-p_o)$	$p_1(1-p_o)$	$p_o(1-p_2)$	p_3	np_3	$P_{a:3}$
2	98	5.44	1.945	7.385	14.79	.503
4	96	5.33	3.890	9.22	18.44	.182
6	94	5.22	5.835	11.055	22.10	.046
8	92	5.11	7.780	12.89	25.78	.009
10	90	5.00	9.725	14.725	-	-
12	88	4.89	11.670	16.56	-	-

Group BGHI

For $p_o = 5\%$: $1 - A_2 = p_1 = .0278$

$A_3 = 1 - p_2 = .731$

p_o	$(1-p_o)$	$p_1(1-p_o)$	$p_o(1-p_2)$	p_3	np_3	$P_{a:3}$
2	98	2.720	1.462	4.182	8.364	.979
4	96	2.665	3.924	5.589	11.178	.846
6	94	2.610	4.386	6.996	13.992	.570
8	92	2.555	5.848	8.403	16.806	.296
10	90	2.500	7.310	9.81	19.62	.123
12	88	2.446	8.770	11.216	22.432	.041

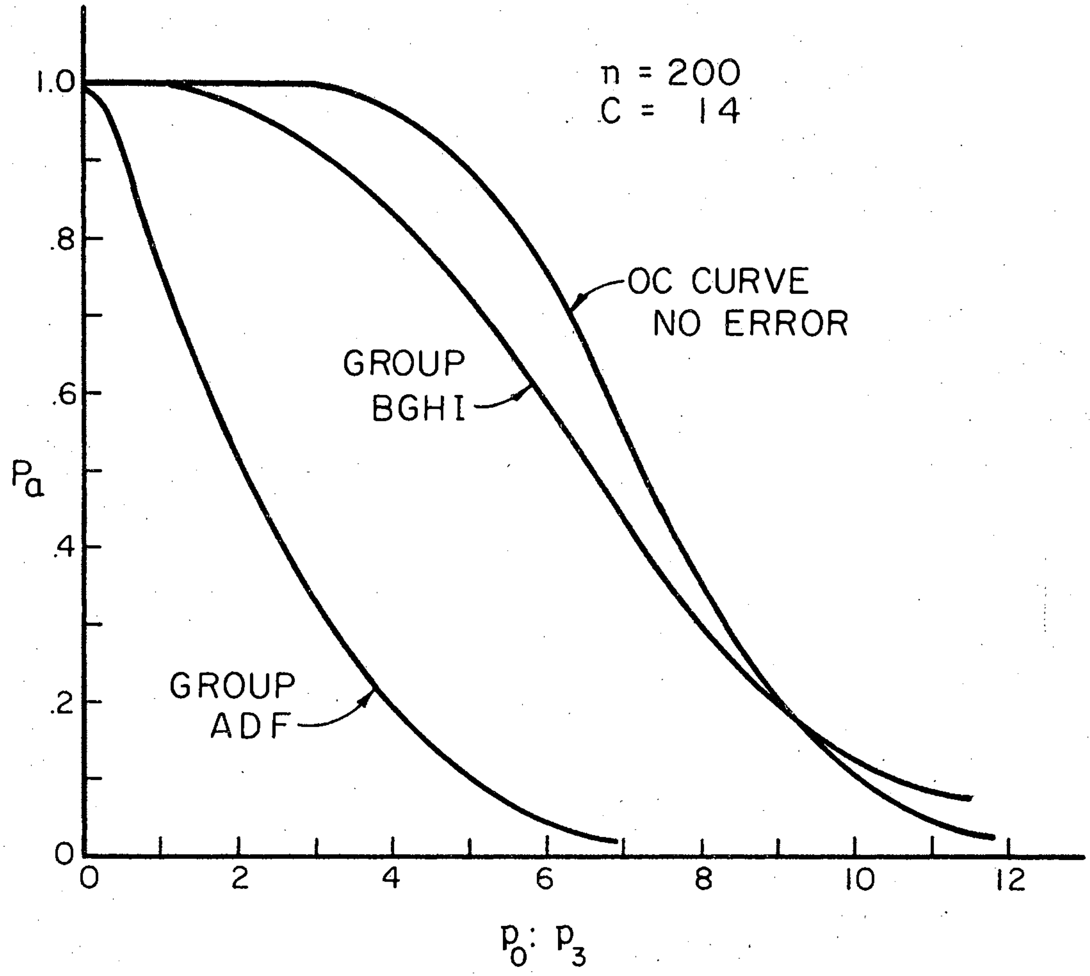


Figure 17. O.C. Curves for Two Groups of Inspectors

error combination, the probability of acceptance of the perceived fraction defective, p_3 tends to be less than that expected for p_0 to some crossover point and then becomes greater than that expected for p_0. In either event, the decision to accept or reject a lot is made for the wrong reasons under conditions of inspector error.

VITA

Paul Mark Wallack

Candidate for the Degree of

Doctor of Philosophy

Thesis: AN EXPERIMENTAL INVESTIGATION OF INDUSTRIAL INSPECTOR ACCURACY UNDER VARYING LEVELS OF PRODUCT DEFECTIVENESS

Major Field: Industrial Engineering and Management

Biographical:

Personal Data: Born August 3, 1927, in Girard, Kansas, the son of Walter Mark and Faye Verma Wallack of Girard, Kansas.

Education: Attended high school in Pittsburg, Kansas, and matriculated at Kansas State Teachers College during last semester of high school term. Upon graduation from high school in 1944, entered Arkansas City, Kansas, Junior College for one semester and then entered the University of Tulsa in 1945. Received the Bachelor of Science degree in Petroleum Refinery Engineering in May, 1950. Entered the Oklahoma State University in September 1955, and received the degree of Master of Science in August 1956. Attended Arizona State University (part time) during 1959-60, 1960-61. Reentered Oklahoma State University in 1962 and completed requirements for the Degree of Doctor of Philosophy in July 1967.

Professional Experience: Employed by the Pure Oil Co., as a petroleum engineer in 1950, and from 1952-55. Employed as a manufacturing and quality control engineer by Sandia Corporation 1956-59; served as Assistant Professor in the Industrial Engineering Department at Arizona State University 1959-62; and served as Assistant and later Associate Professor in the Department of Industrial Engineering at Kansas State University, 1962-64. Employed by North American Aviation, Space and Information Systems Division, as Specialist/Research, Assistant to Manager of Manufacturing Quality Control, and Quality Engineering Supervisor, 1964-66. Currently employed by North American Aviation's Autonetics Division as a Specialist/Research in Computer Applications Technology, Research and Engineering Division.

Professional Membership: American Institute of Industrial Engineers, American Society for Quality Control, Registered Engineer, Alpha Pi Mu, Pi Epsilon Tau, Phi Kappa Phi.