1959

Perception of work performance by workers and supervisors.

Mahoney, James Taylor

Boston University

http://hdl.handle.net/2144/14182

Boston University
Perception of Work Performance

By

Workers and Supervisors

A Dissertation
Presented to
The Faculty of the Graduate School
Boston University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
James Taylor Mahoney

April, 1959
As-sociate Professor of Psychology, Chairman

Professor of Psychology

Associate Professor of Psychology
The writer has attempted through the present study to demonstrate the theoretical and pragmatic implications of value judgments in the area of work performance. A theory of value judgments or value expectations is advanced and several hypotheses derived from this theory are empirically tested. It is believed that the frame of reference advocated holds promise for the greater understanding of human behavior in the work setting.

I would like to express my appreciation to the following organizations for their generous cooperation in this project:

Avery and Saul, Inc.
Badger Manufacturing Co., Inc.
Blanchard Machine Corp.
Edwin Case Shoe, Inc.
Fanny Farmer, Inc.
General Radio Corp.
Gilchrist's, Inc.
Gillette Safety Razor Co., Inc.
Jim's Place, Inc.
John Hancock Life Insurance, Inc.
McAlister-Bicknell, Inc.
Minneapolis-Honeywell, Inc.
Raytheon Manufacturing Co., Inc.
Simplex Wire and Cable, Inc.
United States Testing Co., Inc.
Veterans Administration Hospital, Brockton, Mass.

I would also like to express my gratitude to Dr. Henry Weinberg for his invaluable assistance and encouragement throughout the research project, to Dr. Austin Berkeley for his penetrating criticism and counsel, and to Dr. John Gilmore for his constructive guidance in the organization of the final draft of this report.
author, of course, assumes full responsibility for any defects or errors in the text.

I would also like to thank the following friends and colleagues for their support and criticism:
Dr. J. Edward Conners, Dr. Anthony Davids, Dr. Don Haefner, and Dr. Bernard Stotsky.

To Miss Alice Connors who typed the manuscript go my thanks for a demanding job well done.

Finally, to my wife who has patiently withstood the onslaught for the past four years go my special thanks.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WORK AS AN AREA FOR PSYCHOLOGICAL STUDY</td>
<td>1</td>
</tr>
<tr>
<td>Performance Prediction</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>3</td>
</tr>
<tr>
<td>Values and the Criterion of Work Performance</td>
<td>4</td>
</tr>
<tr>
<td>Phase 1 - Supervision and Occupation as Determiners of Supervisors' Value Systems</td>
<td>8</td>
</tr>
<tr>
<td>Phase II - The Effect of Congruence of Supervisor - Worker Value Systems Upon Supervisor Perception of Worker Effectiveness</td>
<td>10</td>
</tr>
<tr>
<td>11. BACKGROUND OF THE PROBLEM</td>
<td>12</td>
</tr>
<tr>
<td>Phase I - The Effect of Reference Group Upon the Consistency of Value Systems</td>
<td>12</td>
</tr>
<tr>
<td>The Criterion as a Problem for Psychology</td>
<td>13</td>
</tr>
<tr>
<td>Theoretical Approaches to the Criterion Problem in Industrial Psychology</td>
<td>16</td>
</tr>
<tr>
<td>General Factors Affecting the Adequacy of Criterion Measures</td>
<td>27</td>
</tr>
<tr>
<td>The Effect of Function Upon the Criterion Measure</td>
<td>34</td>
</tr>
<tr>
<td>Measurement Methods for the Assessment of Work Performance</td>
<td>37</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Comments on the Current Status of the Criterion</td>
<td>56</td>
</tr>
<tr>
<td>Phase II - Congruence and Supervisor Perception of Work Performance</td>
<td>59</td>
</tr>
<tr>
<td>Values and Behavior</td>
<td>63</td>
</tr>
<tr>
<td>111 STATEMENT OF THE PROBLEM</td>
<td>69</td>
</tr>
<tr>
<td>Phase I - Reference Group and Performance Value Systems</td>
<td>69</td>
</tr>
<tr>
<td>Other Possible Variables</td>
<td>78</td>
</tr>
<tr>
<td>Phase II - Congruence and the Perception of Work Performance</td>
<td>80</td>
</tr>
<tr>
<td>Derivation of the Congruence Hypothesis</td>
<td>82</td>
</tr>
<tr>
<td>An Addendum to the Derivation of the Congruence Hypothesis</td>
<td>85</td>
</tr>
<tr>
<td>Summary</td>
<td>93</td>
</tr>
<tr>
<td>1IV METHOD</td>
<td>95</td>
</tr>
<tr>
<td>Instruments Employed</td>
<td>96</td>
</tr>
<tr>
<td>Statistics Utilized</td>
<td>107</td>
</tr>
<tr>
<td>Phase II - Congruence and the Perception of Work Performance</td>
<td>112</td>
</tr>
<tr>
<td>Subjects</td>
<td>113</td>
</tr>
<tr>
<td>Administration</td>
<td>114</td>
</tr>
</tbody>
</table>
Scoring ............................................. 115
Criterion of Work Performance ............. 116
Statistics ............................................. 116

V RESULTS ............................................. 119

Phase I - Reference Group and Performance

Value Systems ..................................... 119

Table 1 - Pentad Coefficients of Concordance of
Heterogeneous Employees and Heterogeneous
Supervisors ........................................... 119

Table II - Hypothesis No. 1 - Statistical
Analysis of Relative Agreement of
Employees Versus Supervisors .............. 120

Table III - Coefficients of Concordance of Fifteen
Groups of Homogeneous Supervisors and One
Group of Heterogeneous Supervisors ........ 122

Table IV - Hypothesis # 2 - Comparisons of
Fifteen Groups of Homogeneous Occupations
with Heterogeneous Group of Occupations ... 123

Table V - Spearman Rank Order Correlations
Between Successive Administrations of the PAS

N= 10 - Pentad Reliability of the PAS ....... 126
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results - Phase II - Congruence and the Perception of Work Performance</td>
<td>128</td>
</tr>
<tr>
<td>Table VI - Tau Correlations Between the Variable of Worker - Supervisor Congruence and Supervisor Ranking of Worker</td>
<td>128</td>
</tr>
<tr>
<td>Table VII - Combining of CHI Square to Obtain Overall Probability</td>
<td>129</td>
</tr>
<tr>
<td>Intelligence and Supervisor Ranking of Individual Worker Performance</td>
<td>130</td>
</tr>
<tr>
<td>Table VIII - Tau Correlations Between the Variable of Worker - Intelligence and Supervisor Ranking of Worker Associated Values of Chi Square</td>
<td>131</td>
</tr>
<tr>
<td>Table IX - Tau Correlations Between the Variable of Worker - Supervisor Congruence and Worker Intelligence - Associated Values of Chi Square</td>
<td>132</td>
</tr>
<tr>
<td>Table X - Relationship of Worker - Supervisor Congruence to Supervisor Rank of Worker - Intelligence Partialled Out</td>
<td>133</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>VI. CONCLUSIONS AND DISCUSSION</td>
<td>135</td>
</tr>
<tr>
<td>Phase I - Reference Group and Performance</td>
<td></td>
</tr>
<tr>
<td>Value Systems</td>
<td>135</td>
</tr>
<tr>
<td>Phase II - Congruence and the Perception of Work Performance</td>
<td>140</td>
</tr>
<tr>
<td>Summary</td>
<td>144</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>145</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>159</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

WORK AS AN AREA FOR PSYCHOLOGICAL STUDY

One of the areas of interest for psychologists in the study of human behavior is that of worker performance. Since work represents such an important aspect of an industrial society and because it has seemed reasonable that the insights and methods of modern psychology might profitably be utilized towards its greater understanding, psychologists have devoted a great deal of effort in the area. Of this effort, a large portion has been directed towards the prediction of worker performance. Many studies have been reported in the literature which demonstrate relationships between psychological variables and the quality of individual performance on the job. There are, however, a great many unsolved problems in predicting work behavior. These problems are encountered not only in the conceptualization and measurement of the significant psychological predictor variables, but also in the conceptualization and measurement of job behavior.

PERFORMANCE PREDICTION

Performance prediction may be separated most conveniently into two aspects. The first aspect concerns the isolation and refinement of the predictor variables. It is this aspect which has received the most attention in the psychological literature. In this respect, it has now been established that various types of tests may be used to predict the future
ordering of workers on the performance dimension. This predictive accuracy has held true for virtually any type of method used to measure performance. The rationale for the effectiveness of many of these instruments is fairly well understood. It is assumed that certain types of abilities are differentially important for various kinds of jobs. Tests, which in fact measure these abilities should therefore have relatively good predictive qualities. For example, it is well known that intelligence tests predict future learning. In jobs where learning ability is of great importance, intelligence tests should then be of value in the prediction of how an individual will perform.

The second major aspect in the prediction of work performance centers around the performance itself. This aspect, however, has received comparatively little attention from psychologists. Nevertheless, problems in the conceptualization and measurement of work performance are just as crucial as those related to the psychological variables believed relevant to it. The conceptualization and measurement of work behavior together with the difficulties involved are commonly subsumed under what is called the 'criterion problem'. A criterion may be defined as a measure of success or failure in an activity. In some areas of psychology this definition presents no particular problem. For example, in a learning experiment success may be decided on the basis of the number of trials required by a subject before he is able to produce three successive perfect repetitions of a list of nonsense syllables. In other areas of
psychology, however, a great many obstacles arise in the construction of adequate criteria. This is especially true in the area of work performance. The behavior under study is much more complex than in a learning situation. Moreover, the decision as to what constitutes failure or success is much more troublesome.

PURPOSE OF THE STUDY

The general purpose of this study is to demonstrate the feasibility of viewing the area of worker performance as representing a problem in the study of values. This more general purpose will be accomplished through the experimental testing of three hypotheses which rest upon the foregoing assumption.

The study is divided into two phases. The reason for this division is primarily that of clarity of presentation. Nevertheless, these is one important theoretical consideration which recommends this division. This consideration rests upon the previously mentioned division of performance prediction with one category encompassing the predictor variables and the other predicted or criterion variable. The three hypotheses under test in this study may be separated from each other in a similar manner because two hypotheses involve the predicted or criterion variable and the other a predictor variable. Each phase of the study deals with the development, theoretical rationale, method and test of the hypotheses. Unlike the usual personnel prediction study, however, the hypotheses are related at the theoretical level and for the following reasons.
1. All involve the investigation of value systems.

2. The first two hypotheses are derived from the same theoretical rationale.

3. The hypothesis under test in the second phase of the study is to some extent dependent upon the empirical validity of the two hypotheses of the first phase.

VALUES AND THE CRITERION OF WORK PERFORMANCE

Criteria of work performance have been constructed in most instances upon judgments. These judgments have been, in the main, based upon some form of supervisory ratings. There has been a great deal of criticism, however, of procedures employing supervisors¹ ratings. This criticism has been directed at rating procedures because of their necessarily subjective nature. It is well known that the subjectivity inherent in ratings introduces various kinds of bias. ¹

Nevertheless, the crucial question in the matter is whether or not this bias is serious enough to critically invalidate the ratings. The various ramifications of this and other problems in the construction of criteria will be discussed in Chapter 11. It should be pointed out, however, that when the general question of the validity of ratings is raised, a comparison is implied, i.e., ratings are invalid as compared to some other procedure. For many theorists this other procedure involves

some objective criterion such as production record. ² For others, the alternate criterion is a hypothetical true or ultimate measure. ³, ⁴ It is one of the theses of this study that there is no procedure as valid as a good rating when considering work performance from a general standpoint. This should not be construed as a denial of the value of objective measures. Rather, it is suggested that the scope of these measures is limited. If, however, an investigator is only interested in production and an objective measure is available, this should most certainly be used in preference to a rating of productivity. When an overall measure of performance is required, however, there is no procedure which will replace a rating.

This conclusion is based upon the following analysis of the nature of work performance. The analysis deals with the problem at both the abstract theoretical and measurement levels. It will be recalled that a criterion was defined in terms of success or failure in a given activity. It is suggested that these terms directly imply value judgments in virtually any context. In the area of work performance, however, they


are especially value laden. Value may be defined as the degree of worth or excellence ascribed to an object or activity.\(^5\) The writer believes that the words 'success' and 'failure' may be considered two anchor points approaching the extremes on a continuum of worth or excellence. It is thus seen, that from a linguistic view, any definition of the criterion which includes or directly implies success or failure by necessity also implies values.

There is another aspect to the criterion problem at the theoretical level which seems to imply values. Previously, it was mentioned that production record has been suggested by some investigators as the most valid measure of work performance.\(^6\) Nevertheless, it seems clear that production as such does not completely describe performance. There are many behaviors on the job which clearly relate to performance but are not related to production, except in the most tangential fashion. For example, the quality of a worker's interpersonal relationships, the degree to which he can assume responsibility, his attitude, etc. are pertinent to performance. Moreover, in the opinion of the writer, the assumption that production record as such is the most important aspect of performance represents a pre-judgment based upon


\(^{6}\)A. Abruzzi, op. cit.
the value system of a particular judge - and may or may not be representative of the value systems of other judges.

It is in the translation, however, from the theoretical to the practical level that problems of work performance become most acute. While it may be possible to argue for an idealized, ultimate criterion as Thorndike and others do, it is quite another matter to construct a criterion incorporating the "properly weighted embodiment of all the elements making for success in an activity" - Thorndike's definition of the ultimate criterion. Pragmatic decisions must be made as to the element to be included and those not to be included. Thereupon, weights must be assigned to the included elements. Of great importance in the settling of this matter is the question of the identity of the individual who is to make these decisions. From the standpoint of this research, the question may not be settled on the pragmatic level until the values implicit on the theoretical level are recognized. It is believed that the value system of an individual will significantly influence that individual in his decision to include certain elements and in the assignment of weights to these elements. A value system is defined as that set of values overtly accepted by a person or social group. This research, however, is only concerned with

7 R.L. Thorndike, op. cit.
8 H.E. Brogden and E.K. Taylor, op. cit.
values as they are reflected in the area of work performance. A performance value then is defined as the degree of worth or excellence ascribed to a particular segment of work behavior by a particular judge, or group of judges. A performance value system, therefore, is that set of performance values overtly accepted by a particular judge or group of judges. *

PHASE 1 - SUPERVISION AND OCCUPATION AS DETERMINERS OF SUPERVISORS' VALUE SYSTEMS.

It is well known in social psychology that value systems tend to differ from individual to individual. It is also accepted, however, that value systems are not completely idiosyncratic, but that the worth or excellence ascribed to certain behavior or objects tend to be shared in much the same manner by individuals throughout a particular society. Moreover, the social reference groups of an individual or group of individuals often have a crucial effect in determining the content of value systems. **

In much the same manner, it is theorized that performance value systems will tend to differ from individual to individual, but that there will also be certain consistencies as a result of general cultural factors, and factors associated with significant social reference groups.

Investigated in this study will be the effect upon performance value systems of two social reference groups. Thus it is hypothesized that membership in a supervisory reference group will have a sig-

* See Page 8A.
** See Page 8A
* The term value as it is used in this study refers to the process of valuing work behavior. As such it is more restricted than in the usual social psychological sense. It is quite possible that judgments of the worth of work behavior by a particular individual depends, in turn, upon general values within that individual. The point of the study, however, is to demonstrate that while individuals vary in the way that they assign value to segments of work behavior, this variability may be to some extent predicted. The performance value system of an individual refers to the manner in which he ascribes worth to a relatively large group of statements describing work behavior. Some readers may prefer to substitute the term, performance expectations, for performance value system. The reasoning involved in arriving at the hypotheses would not be altered appreciably if it is granted that the performance expectations are related to value judgments within the individual.

** The common distinction between reference and membership groups is somewhat difficult to draw in this study because while it is clear that the makeup of the experimental groups has been accomplished by a division of the subjects according to class membership, it is also possible that reference groups may play a part in determining a subject's responses. Reference group in this study, however, refers primarily to class inclusion.
significant effect upon performance value systems. In like manner, it is hypothesized that membership in a particular occupation will be an important determiner of performance value systems. These two effects will manifest themselves in such a way that increased homogeneity of social reference group will tend towards greater homogeneity of performance value systems. Thus, supervisors could be expected to show greater homogeneity of performance value systems than non-supervisors. Following this same line of reasoning, supervisors in similar occupations could be expected to show greater homogeneity of performance value systems than supervisors drawn from widely disparate occupations.

As previously stated, supervisor judgments of individual worker have formed the basis for by far the most studies in the prediction of performance level. It seems probable that this trend will continue, for the supervisor is in a uniquely advantageous position to observe worker behavior. It is most certainly true, however, that when a supervisor is utilized as a judge one faces the problem of human bias. It is also true, according to the argument advanced here, that the use of any judge of worker behavior introduces bias. Bias in this sense, however, merely means that the way any judge perceives and values work behavior must reflect his own performance value system. If value systems are completely
idiosyncratic, then of course prediction to these systems would be impossible without a comprehensive study of the personality dynamics of each judge. It is the hypothesis of this study, however, that performance value systems are not completely idiosyncratic, but may be predicted with some degree of confidence if the significant social reference groups of a judge are known. If the value systems can be to some extent predicted, it would seem to follow that the prediction of performance from psychological variables will be enhanced.

For example, if it is known that a particular group of judges highly value effective interpersonal behavior, it should be possible to predict that individual workers who possess high interpersonal skills will, from the standpoint of the same group of judges, perform better than those who do not - all other variables being equal. It will be the task of the first phase of the study to demonstrate that performance value systems do, in fact, tend to be consistent when the reference groups of the judges are taken into account.

PHASE II - THE EFFECT OF CONGRUENCE OF SUPERVISOR - WORKER VALUE SYSTEMS UPON SUPERVISOR PERCEPTION OF WORKER EFFECTIVENESS.

It is assumed that virtually all members of an industrial society must possess some sort of value system as related to performance whether well articulated or not. It is therefore clear that the workers themselves must hold values which relate to performance.
It would seem probable, however, that the content of worker value systems would differ systematically from that of supervisors. If the concept of significant social reference group is to hold true, then this would have to be the case, for certainly supervisors and workers do belong to significantly different social reference groups. Nevertheless, it would seem likely that among workers there would be differences in value systems just as with supervisors. Now if it is postulated that there is a general tendency for individuals to behave in ways consistent with their own value systems, then workers whose value systems are most congruent with the value system of their supervisor should exhibit behavior which will be most highly valued by that supervisor. If this is in fact the case, then it would follow that workers whose value systems were highly congruent with the supervisor's value system should be judged by him to be more effective on the job than workers whose value systems are highly incongruent.

The second phase of the study then, will test the hypothesis that congruence of value systems is significantly related to worker performance as judged by the supervisor.
CHAPTER 11
BACKGROUND OF THE PROBLEM

PHASE I - THE EFFECT OF REFERENCE GROUP UPON THE CONSISTENCY OF VALUE SYSTEMS

Values and the Measurement of Work Performance.

There has been some theoretical recognition of the part played by values in the area of work performance. This recognition, however, has not extended very far on the pragmatic level, i.e., in the construction of criteria. This seems to be especially the case when considering the ultimate criterion as a useful model. An ultimate criterion does not have much significance when placed in the context of value judgments.

As an introduction to the first phase of the study, a review of the various facets of the criterion problem will now be undertaken. It should be recognized at the outset, however, that there is an important difference between behavior and its measurement. The criterion problem refers specifically to measurement. It is believed that failure to draw this distinction is responsible for

9 D. Fiske, "Values, Theory and the Criterion Problem" Personnel Psychology, (1951) V4, pp. 93-98
some of the confusion in the area of work performance. When theoretical measurement problems become interchanged with problems connected with the behavior which is being measured, there is apt to be a good deal of cloudiness in the conceptualization process. While the present study has implications for the measurement of work behavior, these implications are secondary to the main purpose of the study. This purpose is concerned with theoretical issues revolving around the effect of values upon the perception of work behavior. The study does not assume that any aspect of work performance behavior is intrinsically good or bad, Rather, the degree of worth must be referred to a performance value system of an individual or group of individuals.

THE CRITERION AS A PROBLEM FOR PSYCHOLOGY

The criterion has been defined in many ways. Nagle quoting Horst, wrote that "the measure of success or failure in an activity is what is technically known as a criterion". Bechtoldt states that the criterion is "a means of describing the performance of individuals on a success continuum".


Despite the seeming simplicity of these definitions, the criterion is an ever present problem in many areas of psychology. The difficulty lies not only in the fact that adequate measures are often not available, but also that what constitutes success has not been fully conceptualized. Clinical Psychology, for example, is faced with such questions as the nature and extent of psychopathology, the measurement of improvement in mental status, etc. Counseling Psychology, to give another example, is vitally concerned with what sort of measures should be used to evaluate the effectiveness of counseling. Similarly, examples could be found from other areas of psychology. They would all serve to point up the problems of how the degree of success is to be measured and the difficulties in the conceptualization of success. The problem, therefore, is almost as broad as the discipline itself since very few areas can escape dealing with it. In the area of Industrial Psychology, however, the criterion is of particular importance.

The Importance of the Criterion in Industrial Psychology.

The importance of arriving at adequate measures of worker performance has recently received more attention in the psychological literature.

Wherry stresses that adequate criteria are essential for the evaluation of training, selection, suggestion systems, and rating
Many other writers have emphasized the crucial significance of adequate resolution of the criterion problem before real progress can be made in the field of personnel assessment. Wallace and Weitz, in the "Annual Review of Psychology" as recently as 1955 claimed that "the criterion problem continues to lead all other topics in lip service and trail most in terms of work completed". Ghiselli in 1956 emphasized the same point by stating "it is certainly true that far more attention has been devoted to the development of predictive devices than to the understanding and evaluation of criteria".

Strong writing in the 1958 "American Psychologist", although primarily concerned about another problem, states -

"The correlations between job satisfaction and production are low not only because measures of job satisfaction are inadequate but because measures of success on the job are also inadequate. It is well recognized that production


is not a complete measure of success on the job.
There is possibly no more difficult problem in
industrial psychology than the determination of
adequate criteria of success. 16

Nevertheless, Nagle reviewing the literature in 1953 could
report only thirty-one titles on the subject for the previous ten
years. 17

It seems quite evident that future advances in the area of
personnel assessment must depend, in part, upon more
sophisticated approaches to the criterion. From this, it hardly
seems necessary to note the necessity for many more studies
and concomitant advances in theory before our understanding of
the subject begins to approximate the ideal.

THEORETICAL APPROACHES TO THE CRITERION PROBLEM
IN INDUSTRIAL PSYCHOLOGY

There have been some thoughtful articles by psychologists
about the criterion as applied to work performance. It should be
noted, however, that psychologists are not the only professional
group interested or working on it. Time and motion analysts,
job evaluators, and personnel people in general are all vitally

16 E. K. Strong, "Satisfactions and Interests" American
Psychologist V13, pp. 449-456

17 B. F. Nagle, op. cit.
concerned. At present, it is sufficient to note that the different professional groups bring different orientations to the problem and that what might prove to be satisfactory for one group might not be so for another; by the very nature of the purpose for which the criterion measure is to be used.

An early article was written by Bellows 18 who advocated that the adequacy of criteria be evaluated along the following dimensions:

1. Reliability
2. Accessibility and Cost
3. Acceptability - to the sponsor
4. Predictability
5. Agreement with criteria

These rules while useful as a first step in criterion measurement do not fully settle the matter. While few would dispute the fact that criteria should not significantly lack any of these qualities, it seems quite clear that more theory needs to be formulated in order to arrive at better criterion measures.

The Objective - Subjective Controversy

In order to point up the enormous complexity of the problem of what criteria should be utilized as the most adequate measure of worker performance and also to introduce one of the major areas of dispute, it is necessary only to turn to two of the early pioneers in the industrial psychology movement. In 1926 Bingham and Freyd compared the relative merits of personality and aptitude measure in the prediction of work performance. They concluded that personality tests tended to predict best to supervisor ratings while aptitude tests were more effective in the prediction of production figures. 19 This very significant finding was recently confirmed. 20 As a corollary, it has also been demonstrated that supervisory ratings correlate to production record only to the extent of plus .48 to plus .55. 21 Many psychologists have interpreted the relative lack of correspondence of ratings and production figures as an indication of an inherent deficiency in the

---

19 W. V. Bingham and M. Freyd, Procedures in Employment Psychology, (New York: Shaw, 1926)


validity of ratings. This deficiency is assumed to be the result of the subjectivity inherent in ratings. Subjectivity, in turn, is presumed to introduce the possibility of bias.

As a result, a number of objective measures have been proposed. These have included work output, spoilage, sales, absenteeism, turnover, and wages. 22,23 None of these has found unqualified acceptance.

Toops in 1944 while generally in favor of the use of objective methods, pointed out some of the factors which tend to invalidate them. 24 For example:

Wages - favoritism may be operative just as it is in ratings. Further, in situations where unions are strong, the wage scale may mean no more than another measure of seniority.

Production - The units of production may be unequal, opportunity bias may be present, i.e., one worker may have a better machine than another or if workers are in teams, one worker may


have better team mates than another. Moreover, as with wages, in some union and non-union shops, there may be artificial elements operating in production figures. One of these elements has been described in the Hawthorne studies. 25 Strong social pressure may be brought to bear on the better workers. This pressure is aimed at restricting production so that the poorer workers will not be placed in an unfavorable light. These and other objections may be raised against the foregoing objective procedures. In fact, valid criticisms may be made of any of the objective measures so far proposed.

Meanwhile ratings, despite their many inadequacies, and while enjoying little theoretical favor have continued to be the most popular measure for assessing work performance. Some investigators, however, seem apologetic about the use of ratings and sometimes point out that no objective measure was available or none could be conveniently obtained. 26 It is apparently the opinion of these investigators that somehow objective measures are greatly preferable to subjective measures. It is

25 F. Roethlisberger and W. J. Dickson, Management and the Worker (Cambridge: Harvard Univ. Press, 1939)

the opinion of this investigator that a rigid dichotomy along an objective versus subjective dimension is artificial. Such a dichotomy has a tendency to obscure theoretical questions about the behavior under study. Further, it seems evident that any objective measure must have important subjective components.

The observations by previously cited investigators 27, 28 relative to the differential predictive merits of personality and aptitude measures is of theoretical importance. If, however, an objective measure is considered the more valid criterion, this implication is to a great extent lost. If, on the other hand, both objective and subjective measures are considered valid, it becomes clear that personality and aptitude tests are predicting to different elements in job behavior. Productivity as measured by production figures, thus is seen to be only one element in performance behavior. The worth of the behavior must be referred to the value system of a particular individual and cannot be settled a priori by the psychologist.

27 W. V. Bingham and M. Freyd, op. cit.

28 G. H. Rush, op. cit.
The Concept of the Ultimate Criterion

Wherry \(^{29}\) recently has called the whole subjective versus objective controversy "stupid" and regards the entire era as the "dark ages of criterion development". He states that more sophisticated thinking about the problem has led to a classification of criteria along the dimension of proximal to ultimate and to some searching discussion of the problem of criterion equivalence. He believes also that increased attention to the idea of the ultimate criterion can help to further increase our sophistication. He predicts for the next decade the following trends:

1. Continued study of the ultimate criterion with emphasis upon the classification of purpose (of criteria), refinement of definition, and the improvement in measurement methods.

2. Increased interest in the field of criteria equivalence: i.e., attempts to prove that more easily or cheaply obtainable methods can be substituted for more expensive ones.

3. Improvement in methods of job and situational analysis.

4. Studies to validate rational (clinical) approaches empirically.

5. Further attention to the selection of proper criteria.

Wherry also claims that the ultimate criterion must refer to the "dollar criterion. Just what he means by this is difficult to state exactly, but evidently he implies that the worker's contribution to an organization be measured in terms of money earned for the organization. This is reminiscent of Brogden and Taylor's suggestion of applying the principles of cost accounting in constructing a criterion.  

This is a method adopted by job evaluation experts in which the estimated value of certain operations or behavior are translated into monetary terms. Unfortunately, however, the method is not too practical for performance evaluation.

It was Thorndike who first introduced the idea of the ultimate criterion into the area of industrial psychology. He thinks of it as "the properly weighted embodiment of all the elements making for success in an activity."

The idea of the ultimate criterion seems to represent an attempt to conceptualize the criterion in much the same manner as

\[ \text{Ibid.} \]


Thurstone and Guilford view the problem of reliability.\textsuperscript{33, 34} In the case of reliability, however, this is conceived to be the relationship of true variance to obtained variance. The true variance may be estimated. When one speaks of an ultimate criterion, however, there is no way of estimating this from the data. Further, one immediately runs into the problem of the adequacy of practical criteria in representing the ultimate which has been the question right along.

Nevertheless, virtually all writers on the criterion problem since Thorndike's work in 1949 have utilized the concept of the ultimate criterion when discussing criterion problems. It seems to have value as an aid to conceptualization. Nevertheless, difficulties arise on a conceptual level with such a definition of the ultimate criterion just as they do with any definition which employs the word 'success'. These problems, as previously mentioned, refer to the value judgments involved in defining success for a particular activity. The proper weighting of the elements making for success may be accomplished on the prag-

\textsuperscript{33} L. L. Thurstone, "Ability, Motivation and Speed", \textit{Psychometrica} \textcopyright\textsuperscript{1937} V2, pp. 299-354

matic level only if the configuration of the elements is known on a theoretical level. The introduction of a concept such as an ultimate criterion may confuse theoretical issues not simply because the ultimate criterion is impossible to construct in reality, but rather because it implies that there is some absolute standard of performance behavior if psychologists are only clever enough to discover it.

**Dimensional Problems of Criteria**

Ghiselli recently pointed out that the criterion problem is many faceted. He adds that it is necessary to look not only at the specific techniques used to measure performance but also along the dimension upon which measurement is being made. He notes that there are at least three dimensions along which a particular criterion may be conceived primarily to fall. The static dimension is the one most often employed in psychological research so far. A measurement at one particular point in time is utilized. The variable or variables used in the criterion measure are summed so that each worker is placed along a continuum from good to poor. The predictive device or devices are then evaluated to see how well they have ordered the workers. Ghiselli remarks that there are

---

35 *E. E. Ghiselli, op. cit.*
many more problems in predicting to multivariate criteria than to a single dimension. Good workers must be located in multi-dimensional space not merely on a single continuum in this case.

The second dimension which criteria may have is called dynamic. This is differentiated from static mainly by virtue of the fact that measurements are made over time and rate of improvement in performance is taken into account. This has direct relevance to predictions which are to be made to different situations.

It has been found, for example, that the best predictors to success in training were not necessarily the best predictors to later actual performance outside of training. This may be another way of saying that abilities important in training are not necessarily important on the job or abilities or qualities valuable initially on a job may become less important over a period of time. This may have direct implications for predicting work performance from personality tests.

The third dimension Ghiselli calls the calls "the dimension of the individual". To illustrate this he states that different individuals may contribute different things to an organization yet still have the same job description and be considered equally.
valuable. One college professor may prove to be an excellent
teacher, another an excellent researcher and so forth. Ghiselli
is clearly thinking of ipsative types of measurement in this
respect for he specifically mentions the Q technique developed
by Stephenson. He hoped that further advances in this type of
criterion research could point specifically to important aspects
of the job and how this could be related to the individual worker.
Humphries recently has noted the problems involved in translating
ipsative measurements into normative types of scales. It
remains to be seen whether this can be successfully accomplished
on such a complex problem as work performance.

GENERAL FACTORS AFFECTING THE ADEQUACY OF CRITERION
MEASURES

Bias and the Criterion

Brogden and Taylor have contributed a valuable article on
some general methodological problems involved in the criterion
problem. They bring up the question of the effect of biasing
factors upon the criterion as well as where the various types of

36 W. Stephenson, The Study of Behavior, (Chicago:
Univ. of Chicago Press, 1953.)

37 L.G. Humphries, "Characteristics of Type Concepts with
Special Reference to Sheldon's Typology," Psychological
Bulletin, (1957) V54

38 H.E. Brogden and E.K. Taylor, "Theory and Classification
of Criterion Bias", Education and Psychological Measurement
(1950) V10, pp. 159-186
bias are particularly likely to appear, and how their effects can be to some extent controlled.

A biasing factor may be defined as any variable except errors of measurement and sampling error, producing a deviation of obtained criterion scores from a hypothetical 'true' criterion score.11

They then go on to list three ways in which a criterion may be biased in terms of the ultimate criterion.

1. Criterion deficiency - Some important variable actually included in the ultimate criterion is not included in the criterion. In statistical terms - the criterion does not include variance present in the ultimate criterion.

2. Criterion contamination - Extraneous variance is included in the criterion.

3. Criterion distortion - Although the criterion contains only the elements in the ultimate criterion these are combined in such a way as to cause distortion.

Of crucial significance to the researcher is whether these various biasing factors are correlated to the predictors or not. And if correlated - in what direction? In the case of correlation, there are two possibilities. The predictors either bear a spuriously high relationship to the ultimate criterion or a spuriously low relationship. Either situation is unsatisfactory.
Relevancy and the Ultimate Criterion

Nagle in reviewing the criterion question raised what he considers to be three fundamental problems. 39

1. Relevancy - Refers to the extent of which an index of success (practical criterion) is related to the 'true' order of success in a given activity. Relevancy is the hypothetical correlation coefficient between the criterion used and the ultimate criterion.

This is a key concept, for it theoretically gives one a way to tie the practical criterion to the ultimate criterion. Unfortunately, however, there are at present no ways in which this can be accomplished pragmatically. It will be noted, nonetheless, that Brogden & Taylor's notions about sources of bias in the criterion can be directly related to relevancy. 40

2. Reliability - The reliability of the criterion measure is conceptualized in much the same way as it is conceived for mental tests.

Nagle, as does Adkins, stresses the importance of high reliability for two reasons. 41, 42 (a) Relevancy cannot exceed the square

39 B.F. Nagle, op. cit.
40 H.E. Brogden and E.K. Taylor, op. cit.
41 B.F. Nagle, op. cit.
root of the reliability coefficient of the criterion. It is plain here that Nagle is using the concept of relevancy in just about the same way as that of validity of a test. (b) The obtained validity coefficient between criterion and predictor cannot exceed the square root of the product of the reliability coefficients of the predictor and the criterion. It is clear then, that a necessary condition for high correlation coefficients between predictor and criterion is high reliability of the criterion as well as that of the predictor.

In this respect, Nagle notes the usual sources of criterion unreliability.

1. Size of the sample of performance.
2. Range of ability of the subjects.
3. Ambiguity of instructions.
4. Variation in conditions during the measurement period.
5. The amount of aid provided by the instruments.

He also remarks that criterion contamination can cause spuriously high reliabilities.
The third general problem of criteria with which Nagle deals is:

3. An overall measure versus a combined measure—Nagle believes firmly in the necessity for a complexity of variables in order to maximize the relevancy of the practical criterion to the ultimate criterion. 43

Work performance is a very complex concept and the means to codify this in the form of the criterion must of necessity be complex also. There has been some disagreement about this in the literature especially as this concerns ratings. Jurgenson has found very high intercorrelations between rated traits of worker performance. 44 These intercorrelations may exceed in magnitude the test-retest reliability coefficients of the traits themselves. Jurgenson later concluded that the rank order method may be inherently more reliable than the combination of various trait ratings. 45

43 B.F. Nagle, op. cit.


More recent factor analytic studies, however, would seem to support the idea of the necessity for the combination of multiple sub-criteria into a composite criterion score. Rush found 4 relatively uncorrelated factors which in combination gave significantly higher multiple r's than did an overall measure for sales people. A similar result was reported by Ryans, who found that the behavior of teachers in the classroom could most effectively be described by several major factors.

**Methodological Problems in the Combining of Sub-Criteria**

There are initially two general methodological considerations involved in the combination of sub-criteria. These problems concern the comparability of measurement units within a particular scale and from scale to scale. These problems, however, are troublesome in virtually any area of psychology where scaling devices are employed, and are beyond the scope of this paper.

There are, however, specific problems in the combination of sub-criteria which transcend measurement considerations. Nagle reviews several of the methods which so far have been used.

---

46 C.H. Rush, *op. cit.*

to deal with these problems. 48

1. Weighting of the sub-criteria by experts. This was first proposed by Toops in 1928. 49

2. Weighting in proportion to reliability. Lawshe and Nagle found, however, "that much of the time and effort spent in determining the reliabilities of raters and combining their ratings using differential weights is for naught."50 This, of course, refers only to the combination of different raters judgments of overall work performance.

3. Weighting in proportion to the average correlation with other variables. This, has the effect of exaggerating the influence of the general factor in the sub-criteria.

4. Weighting according to a factor analysis.

5. Weighting criterion variables to the degree that they agree with the predictor. As Nagle points out, this is a

48 B.F. Nagle, op. cit.


perversion of the whole prediction process.

6. Weighting the variables in such a way that the distance between all possible pairs of subjects are as great as possible.

Nagle believes that the only defensible way of combining sub-criteria is weighting in proportion to the sub-criteria's relationship to the ultimate criterion. He also believes that the relevancy must be judged and since the weights of the sub-criteria must be judged, all weightings of sub-criteria on a relevancy basis is subjective.

THE EFFECT OF FUNCTION UPON THE CRITERION MEASURE

Suffice to say that running implicitly throughout many of the quoted writers ideas has been the notion that criterion measures may be very different depending upon the purpose to which they are put. As Wherry has stated, proper criteria must be selected. One would assume conversely that improper criteria must not. Nagle claims that the criterion which is closest to the ultimate criterion in terms of relevancy must be employed. Relevancy, however, he can only see as determined

---

51 B.F. Nagle, _op. cit._

52 R. Wherry, _op. cit._

53 B.F. Nagle, _op. cit._
by expert judgment. The question naturally arises about who are the experts. And if one has differing groups of experts is the ultimate criterion going to be different? These are questions which have not been answered but from available evidence, it would seem that differing groups will come up with different notions as to what the criterion should be. Berkeley, for example, found that supervisors and workers tend to rate different aspects of job performance in different ways with workers emphasizing motivational factors and supervisors production factors. The same general point may be made by considering the text written by Abruzzi on work measurement. It is Abruzzi's opinion that all subjective procedures are worthless and unscientific. It must be noted, however, that Abruzzi is a job evaluation expert concerned with the setting of pay rates in the garment industry. The garment industry has a highly articulate union. It is not surprising, in view of the history of this industry, that union officials would regard subjective ratings of workers by supervisors


55 A. Abruzzi, op. cit.
as suspect. In this case, it would hardly matter how well a subjective procedure was constructed. The psychologist, however, is not faced with this particular situation. Rather, he is much more likely to be concerned with the criterion as a means of validating his predictive devices. As such he is interested in personnel decisions with respect to the workers to whom he has administered his instruments. In order to find out who has been a success in the activity which is being predicted he must have a criterion. It is not, however, usually necessary that he substantiate his criterion to anyone but other psychologists and management. It should be clear then that the degree of 'fairness' of his criterion in terms of whether it looks equally reasonable to employee and employer alike is irrelevant. The purpose for which the criterion is being constructed is crucial in determining what the nature of the criterion will be. Furthermore, the very nature of the predictive device may change depending upon the criterion. The prediction of productivity, for example, may involve quite a different rationale than for turnover.

MEASUREMENT METHODS FOR THE ASSESSMENT OF WORK PERFORMANCE

There are many methods utilized for the measurement of work performance. These methods may be considered to fall into two major classifications - subjective and objective. Subjective procedures depend upon the judgments of individuals while the objective for the most part do not. Within the subjective classification are such specific techniques as rank order, paired comparison, and all the various rating scales. The objective methods include production record, wages, absenteeism, turnover, etc. Each of these methods has usually some advantage to recommend it, but it is also clear that there are various disadvantages connected with each. A critical evaluation of the objective and subjective orientations has already been made. A detailed discussion of the relative merits of the specific techniques is not particularly germane to this study. While the study is concerned with judges and the variables which influence judgments, this is in the abstract. The specific behavior being judged is secondary. While the study certainly has implications for the specific techniques, it does not depend upon any of these techniques except in a rather oblique manner. The interested reader will find relatively complete accounts of these techniques
There are, however, some recent innovations in the construction of specific criteria which do bear directly on the theoretical rationale of the study. Two of these innovations involve rating scale methodology while the third is concerned with the use of factor analysis in the construction of criteria.

**The Critical Incidents Technique**

A recurrent problem in appraising personnel performance has to do with relevancy of a particular criterion to that of a hypothetical ultimate criterion. It is unnecessary, however, to become so abstract because there are problems on a lower level which, if solved, could be quite valuable.

One of these problems involves the relevancy of various aspects of the criterion employed to an overall judgment of worker effectiveness. This is most clearly evident in the case of rating scales. These scales are often made up a priori and a natural question is the relevancy of the items employed to the job under study. For example, rating scales may be made up consisting of traits or abilities which bear differential applicability to a particular job. It is well known that different jobs require

---

57 J.P. Guilford, _op. cit._
It is quite possible that abilities for a particular job may be under-represented by the items in a particular scale (criterion deficiency) while other abilities may be over-represented (criterion contamination).

Flanagan and his colleagues have recognized this problem and have attempted to deal with it by means of a special technique termed Critical Incidents. The procedure was developed for the Air Force and later applied to the industrial setting. The technique grew out of Flanagan's thinking about the critical requirements for a particular job. These he defines as "a requirement which is crucial in the sense that it has been responsible for outstandingly effective or definitely unsatisfactory performance of an important part of the job or activity in question." These critical requirements can only be established by a careful study of the behavior of workers on the job.

Flanagan originally believed that the setting of critical requirements could only be accomplished by job analysis.

---


59 J. C. Flanagan, Personnel Psychology (1949) V2, pp. 419-425
Presumably job analysts would observe actual job behavior. Later, however, he apparently modified this view by stating that supervisors were in the best position to make the observations. The critical incidents of a particular job represent the operational approach to critical requirements. An incident is defined as "any observable human activity that is sufficiently complete in itself to permit inferences and predictions to be made about the person performing it." In order for an incident to be critical it "must occur in a situation where the purpose or intent of the act seems fairly clear to the observer and where it's consequences are sufficiently definite to leave little doubt concerning it's effects." An analysis of the implications of these definitions reveals that there is no necessary relationship to the work setting. Flanagan has stated that the critical incidents technique might be applied to any situation where a criterion must be constructed. He mentions as suitable such diverse problems as the evaluation of the effects of motivation, leadership, and psychotherapy. Moreover, Hobbs used the method in a study conceiving the development of a code of


61 Ibid.
ethics for psychologists. Nevertheless, the most common use has been in the establishment of critical incidents for jobs.

The procedure as described by Flanagan is as follows:

First, the general aims of the study are settled. Then plans and specifications of the study are detailed with instructions given carefully to observers. The incidents are preferably obtained by direct observation but recalled observation will suffice if the former is not possible. Observers are instructed to note discrete pieces of behavior which in their opinion have been examples of outstandingly effective or outstandingly ineffective performance on a particular job.

These observations are then collected, analyzed and the critical incidents themselves made up from the analysis. A list of these critical incidents then can be established for a particular job and rating scales developed which incorporate them. In this way it is felt that extraneous elements for a particular job will be eliminated while insuring adequate representation of the critical requirements.


63 J.G. Flanagan, op. cit.
The method was first utilized in a study of the reasons for pilot failure in training.\footnote{N.E. Miller, Psychological research on pilot training, U.S. Government Printing Office, Army Air Force, Aviation Psychological Program, Research Report, No. 8 (1947)} It was later applied to the study of the factors involved in vertigo in pilots. This proved so successful that it was possible to redesign aircraft equipment to prevent future occurrence of the phenomenon.\footnote{F. Wickert, Psychological research on problems of redistribution. U.S. Government Printing Office, Army Air Force Aviation Psychological Program, Research Report, No. 14, (1947)}

The first application in industry was reported by Miller and Flanagan in 1950.\footnote{R.B. Miller & J.C. Flanagan, "The Performance Record; An Objective Merit Rating Procedure for Industry", American Psychologist U.S. (1950) pp. 331-332.} It has been used since to study the critical requirements for such jobs as dentists, insurance agency heads, bookkeepers, foremen, and psychiatric aides.\footnote{R.F. Wagner, "A Study of the Critical Requirements for Dentists" Univ. of Pittsburgh Bulletin, (1950) (Abs) V46, pp. 331-339}  
There are many advantages to the critical incidents technique and undoubtedly it will become even more popular as a research tool. Its chief disadvantage is partially the result of its principle virtue which is the specificity of the incidents for the particular job under study. It is very difficult, therefore, to make generalizations over classes of even similar types of jobs. Nevertheless, the rationale of the critical incidents technique bears directly upon this study. In investigating the critical incidents, the researcher is studying either behavior or its direct consequence. It is assumed implicitly that the critical incidents will differ from job to job. Therefore, the critical behavior upon which the incidents are based must also differ from job to job to a greater or lesser extent. These incidents are gathered through the report of judges who, for the most part consist of supervisors. It is the thesis of this study that supervisors judge behavior through the medium of value systems. It is known that value systems will differ from individual to individual. Therefore, one would expect the critical incidents for good or poor performance to differ from individual to individual. Flanagan, however, has demonstrated that critical incidents tend to be relatively uniform in the same occupation, regardless of supervision. It is suggested that the reason for this
uniformity lies in the fact that supervisory value systems within the same occupation tend to be similar. By the same line of reasoning, critical incidents in widely different occupations should be different because the value systems of the supervisors are different. Nevertheless, the assumption that critical incidents found for a particular job in one setting can be transposed entirely into the same type of a job in another setting appears unwarranted. This conclusion is based upon the theory that while value systems for the supervisors in the same occupation tend to be similar, there are probably other significant variables operative which would not be reflected in a straight occupational breakdown.

Forced-Choice Rating Scales

The forced-choice method of constructing rating scales owes its development chiefly to the Personnel Research section of the U.S. Army. Travers\(^7\) in a critical review, however, has noted that the idea was originally conceived by Horst and later utilized by Wherry. The forced-choice method was devised largely to meet the problem of leniency errors by raters and to a lesser extent

that of halo.

All of the services during the Second World War were faced with a tremendous problem in appraising the performance of their officers. While merit rating scales had been used for many years, it was apparent that most officers were very reluctant to say anything derogatory about their fellow officers. As a result they were inclined to use only the upper point of the scales. It was well known in military circles that a rating of very good as opposed to that of excellent was very damaging to an officer; for it connoted in effect that he was considered a very poor officer. The natural consequence of this was a very large negative skew in the distribution of the ratings. While this allowed the detection of obviously unfit officers, it left little room for discrimination among the rest of the group. Since there was no reason to believe that ability to function as an officer was distributed in the way that the ratings implied and because there were many reasons to assume that the ability was roughly normally distributed, the Army decided to study methods for improving the distribution. The result of their efforts was the Forced-choice Rating Scale.

The rationale for the method as developed by the Army is as follows: Since raters have a tendency to dislike checking derogatory items, then give them a scale in which they do not have to check
derogatory statements. In the ordinary check list they have this option but problems then arise of skewed distributions. The forced-choice method is essentially a check-list but with quite a different construction and format. Instead of just a list of adjectives or statements which may or may not be checked as applicable by the rater, the statements are grouped and the rater must make certain decisions about each group. In the case of the Army, groups of four statements (tetrads) are established and the rater must choose the statement which is most applicable and the one which is least applicable. A sample tetrad follows:

----------careless
----------serious minded
----------energetic
----------snobbish

It will be noted that there are two favorable items and two unfavorable. The rater therefore can check a favorable item as most characteristic and an unfavorable one as least characteristic. The items, however, are differentially weighted. Each item has associated with it two scores; its discrimination value and its preference value. The discrimination value is obtained by correlating the frequency of its occurrence in high and low criterion groups and essentially is an index of validity (relevance). The preference value of the item is an index of the degree to which the quality is valued by individuals in general. This score essentially
reflects the social desirability of the item. Items are paired on the basis of similar preference values but widely dissimilar discrimination values. In the previously described format two items with relatively high preference value are paired with two items of low preference value.

It is assumed that the effect of the leniency error and halo tendencies of the rater will be counteracted by this pairing. The theory is that the irrelevant items will serve as suppressor variables which will act to decrease the effect of rater bias. It is believed that if rater bias is actually seriously operating, he will tend to check the irrelevant items (which, adds nothing to the rater score) as often as he does the relevant items which does add to the score.

The technique first was reported by Sisson in 1948. Since that time articles have continued to appear describing the applications of the technique to particular situations. Most of the articles have reported favorable results. 74, 75 Travers, however,


in 1951 called the basic assumptions and findings of the method into question. It was his contention that the form of the distributions reported were not markedly superior to that of previously reported studies using graphic techniques. Since the construction of forced-choice scales involve considerable effort and expense, it was moot whether the amount of gain was worth it. Baier, however, takes issue with Travers.

Taylor and Wherry in a study comparing the distributions of forced-choice and graphic ratings under two experimental conditions found that there was a marked increase in mean scores for the graphic technique in shifting from a situation where the raters knew that the ratings were not to be used for a personnel action to a "for keeps" situation. The forced-choice ratings did not do so.

Highland and Berkshire, however, found that when raters were instructed to make rates appear more favorable, the averages of scores increased 50% to 75% of a standard deviation using a forced

---

76 R.M. Travers, op. cit.


choice scale. Guiford, however, has pointed out the general problem of transforming ipsative scores to normative ones as a limitation to the theoretical and practical value of the forced-choice technique.

Forced-Choice as a Reflection of Reference Group

A further difficulty is the fact that in the forced-choice method, as so far developed, there is no adequate rationale for the reason for differences between the pairs of statements on the discrimination index.

Statements or traits which are highly characteristic of the high criterion groups receive higher discrimination values than statements which are not. Those statements which receive high discrimination values can therefore be considered highly relevant to successful performance. A theoretical question arises, however, as to why these statements are highly relevant to performance in a given activity. This question can not be satisfactorily answered within the framework of present force-choice methodology.

---


80 J.P. Guilford, op. cit.
This lack of theory, however, cannot be considered a reflection on the method itself for forced-choice was not designed to deal with such questions. Rather, forced-choice rating scales are constructed in an effort to minimize the effect of rater bias and concomitantly deal with the knotty problem of social desirability. The high criterion group is accepted as a given, and questions directed to the establishment of why individuals are placed in a high criterion group as opposed to a low criterion group are irrelevant. The forced-choice method is an empirical device designed to help solve methodological problems and has no necessary relationship to psychological theory.

Nevertheless, the empirical findings reported using forced-choice scales do in fact have implications for psychological theory—especially the theory being developed in this study. A very relevant question from this study's standpoint is why some individuals are placed in the high criterion group and others are not. It is theorized that individuals in the high criterion group exhibit behavior which is highly valued by a group of judges.

The differential discrimination values assigned to the statements of a forced-choice scale are a function of the expert judges who are employed. These judgments must go beyond simple social desirability considerations. Otherwise, there would be no
difference between the discrimination and preference values of statements except those which arise on a random basis. In order to make this point clearer, consider this hypothetical example. Suppose that in order to assess the performance of Army officers, a completely random sample of judges was drawn from the general adult population. These judges could be assumed to have little or no experience in this activity. They would perhaps include housewives, students, machinists, butchers, - virtually any member of society. Now assume that they received no instructions about what behavior was considered relevant. On what basis would these individuals judge the behavior of officers? It is suggested the only basis available would be from a general cultural context, i.e. using the general cultural values. This, of course, is just another way of stating that the judges would be valuing behavior utilizing the social desirability of the behavior as a yardstick. Therefore, in this hypothetical situation, there would be no point in arriving at discrimination values of various statements representing behavior or behavioral traits. It should be pointed out, however, that it is not really true that the random sample of judges would be judging using only the value system of the general culture as they perceive it. Actually, of course, all the judges have in addition to the values imparted by the general culture, other values
which are specific to them and the significant social reference
groups from which they are drawn. Nevertheless, these values
would not have any effect upon such a hypothetical study except
in the case where the sample was biased. However, if the sample
was truly random, differences in value systems should cancel
themselves out leaving only the effect of a general cultural factor.

Naturally, the drawing of a random sample of judges from
the population to assess work behavior would be of little value.
There was a very important theoretical point, however, which has
been raised by the foregoing example. This involves the question
of bias; and bias is intimately connected with value systems.
Bias in a group of individuals concerns the consistency found in
value systems. In the completely random sample of judges, one
would find relatively little systematic bias other than that resulting
from the general cultural factor. On the other hand, it is assumed
that with a group of senior Army officers used as judges, there
will be far more systematic bias or consistency in the value
systems. This group hardly represents a random sample. Never-
theless, it is suggested that the psychologist is interested in the
codification of systematic bias above and beyond social desirability
considerations. When the question is viewed in this manner, the
consistency of value systems within non-random samples of judges
is of crucial importance.
Factor Analysis as a Contribution to Performance Criterion Development.

Factor analysis has been applied usefully to the criterion problem - as this concerns both ratings and the combination of ratings with other measures.

This technique measures the various aspects of job performance and assigns weights to these elements on the basis of loadings of the factors discovered. It is well known, especially with ratings, that traits tend to correlate highly with one another. It is also recognized that various types of abilities are often correlated with one another.

When this is the case, it is clear that the various conceived discrete traits and abilities are really measuring the same thing - at least in part. Factor analysis may be employed to isolate the elements which are common to the intercorrelated measures. It is therefore possible to examine carefully the makeup of criterion measures and make an evaluation of the various factors which are contributing to them.

Rush, noting the fact that overall sales success could not be adequately described only in terms of sales volume, did a factor analytic study of sales criteria. 81 He used three different classes

---

81 C.H. Rush, op. cit.
of criteria. These were supervisory ratings of a number of performance characteristics, sales records and grades in a technical school. He found four relatively uncorrelated factors.

Roach in a factor analysis of rated supervisory behavior using a 390 item check list derived some 15 factors. Some of these, however, were intercorrelated. 82 Hausman and Strupp also have found relatively independent factors of a non-technical nature in supervisor ratings of aircraft mechanics. 83

Factor Analysis and Job Evaluation

The previously mentioned studies have been concerned with performance of the individual. A somewhat related area is that of job evaluation. In this procedure it is not the excellence of the individual performing his duties which is the object of study. Rather, it is the job itself which is evaluated and not the person performing it.

Ash did a factor analysis of eight job rating scales using three classes of positions - clerical, sub-professional and professional. 84


Working from detailed job descriptions, analysts rated each job on these rating scales. The scales contained such things as the supervisory control exercised over the position, originality of thinking required and personal relations. Ash was able to isolate one factor which was general for all three classes of jobs and two which were found in two of the other classes. These he identified as supervisory responsibility for all three and general skill demands for two of the classes. However, other factors such as job complexity, non-supervisory authority and technical competence demands were also found but were specific to the different classes.

Job evaluations are primarily aimed at arriving at equitable wages for a particular job and bear no necessary relationship to the evaluation of individual performance. Nevertheless, Ash's results are very suggestive for performance criterion research for they seem to clearly imply that determiners of worth are different for different types of jobs.

It can be seen from these brief examples that factor analysis has contributed to the understanding and development of more adequate criteria. It must be remembered, however, that the method is just as dependent for its results upon the adequacy of the information supplied as other methods. It cannot evaluate significant omissions of information (criterion deficiency).
Furthermore, factor analysis is only another empirical technique. It has no necessary relationship to theory. Some of the cited studies demonstrate that performance standards tend to differ from occupation to occupation. While this may be suggestive to theory building, the data in itself does not present to the investigator a ready-made comprehensive theory.

COMMENTS ON THE CURRENT STATUS OF THE CRITERION

From the foregoing discussion a number of conclusions and hypotheses may be drawn about the current status of the criterion problem.

1. There is marked disagreement among authorities in the field as to what constitutes a proper criterion.

2. There are in general two ways in which criterion measurement has traditionally been attempted. One involves the use of objective measures, the most prominent of which have been production figures; the other has depended upon human judgment of the effectiveness of work performance. A third opinion takes the view that both sets of data should be utilized.

3. The purpose to which the criterion is to be constructed vitally affects its nature. With more sophistication, this fact is becoming more and more evident.

This naturally leads to the question of the philosophy behind the purpose of the criterion measurement. Fiske has suggested that the psychologist use the judgments of top management as the
ultimate criterion. 85

In like manner, Stern, Stein and Bloom 86 suggest that the psychologist try to determine the implicit as well as the explicit expectations of management. In this way it is felt that prediction from psychological variables will be far more efficient. All of the authors recognize the part played by values in the construction of a pragmatic criterion measure. They have not, however, attempted a comprehensive theory of the role of values in the criterion problem.

In this connection the question may be raised as to whether this is really all the psychologist should expect from himself. Are the predictions only to be utilized to approximate a future version of the status quo or can the psychologist make a real contribution in the raising of performance throughout an organization by taking an active part in changing the value systems of management?

This is a question which cannot be answered definitely at the present time. Not only are values difficult to change, but there is also a prior problem to be settled. This problem concerns the

---

85 D. W. Fiske, op. cit.

content of values which are to be suggested by the psychologist as more appropriate than the current ones held by management. In other words, whose value system is to be recommended? Even if one were only interested in the pragmatic consequences of different kinds of value systems, there are many difficulties. When philosophical issues especially outside the referral point of management are raised, however, the problem becomes virtually insoluble. It would seem therefore, that the psychologist will make his contribution by pointing out the pragmatic consequences of different value systems. These consequences he may arrive at through his research. The criterion problem has many social ramifications, but the problems connected with these ramifications, however, are not scientific. The present study will attempt to codify the value systems of one group of experts opinion as to what constitutes good work performance. This group consists of first-line supervisors of workers.
PHASE II - CONGRUENCE AND SUPERVISOR PERCEPTION OF WORK PERFORMANCE

Personnel Selection through Psychological Tests

The process of personnel selection has for many years been of concern to psychologists. Gathering impetus from selection and classification problems encountered by the Armed Services in World War I, interest spread to the possibility of utilizing psychological devices, namely tests, in the industrial work setting. During the 1920's and 1930's more and more concerns began to use intelligence and aptitude tests as aids in the assessment procedure. It was during World War II, however, that the psychological testing movement received its greatest boost. Faced with the problems of classifying and effectively assigning millions of individuals to jobs where maximal use of their talents could be obtained, the services instituted comprehensive testing programs. These testing programs while inadequate in many respects were still very valuable in making known to classification officers a great deal about the general intelligence and special abilities of recruits about which very little had been known previously. The Armed Forces also used psychologists in the assessment procedure for special projects

87 L.J. Cronbach, Psychological Testing.
such as OSS, the prediction of air cadet success and the like. Since the war the Army, Navy and Air Force have made considerable outlays for extensive research upon the entire assessment procedure. Also since the war industrial organizations have shown more interest in research on the selection process. Test batteries have been devised which more or less successfully predict performance in such occupations as taxi cab drivers, sales, clerical work, machinist and virtually all of the skilled trades, to name just a few. In addition

88 Office of Strategic Services, Assessment Staff, Assessment of Men (New York: Rinehart, 1948)


the United States Employment Service has carried out extensive research programs on the prediction of performance for vast numbers of jobs using a specially devised aptitude battery, the General Aptitude Test Battery or GATB.

The promise and problems connected with the selection of the right man for the right job has attracted the interest of many psychologists. There is an ever increasing body of literature on the subject. The problem of the selection and construction of adequate criteria has already been discussed. There are also very great obstacles in the way of constructing adequate measuring devices for the abilities, interests and traits presumed important for performance on the job. This is a vast problem area which is beyond the scope of this paper, however.98, 99

95 General Aptitude Test Battery, U.S. Employment Service


98 J.P. Guilford, op. cit.

It is sufficient to note that at this time the general aim of assessment studies using psychological tests is to order individuals on the basis of the selected tests so as to approximate the ordering of the same individuals on the criterion measure. The closer the fit of the sets, the better the power of the instrument to predict to the criterion. There are, in general, two ways which the mechanics of prediction studies are carried out. The first is actually often preliminary step for the second. This involves the attempt to order workers presently employed on the criterion using the variable under investigation. Once the association has been established in the employed population, the variable can be used to predict the success of new workers. The second step is the acid test for any variable presumed to be related to work performance.

It should be noted, however, that with the use of the first step, there are often factors which may attenuate the true relationship between variable and criterion. The poorer workers may, for example, may have already been released and the investigation is probably dealing with workers in the average range and above.
VALUES AND BEHAVIOR

It is accepted in psychology that value systems tend to influence behavior in various predictable ways. Thus, for example, it may be predicted that conservative economic value systems will generally characterize individuals who vote the straight Republican ticket. In a similar manner, it has been observed that high religious values may be related systematically to church attendance. There have been many studies of the effects of values either directly upon behavior or indirectly through the medium of attitudes. These efforts have met mixed success.\textsuperscript{100} It appears that some behavior is related to the value system of the individual while other behavior does not show such a relationship. A possible explanation for the variability in the usefulness of the concept of value may be adduced if it is recognized that the values so far experimentally employed have been relatively circumscribed as compared to the wide range of behavior investigated. The very great majority of studies have utilized the Allport - Vernon Scale as the measure of values. This scale has been throughly

\begin{flushright}
\textsuperscript{100} W.F. Dukes, Psychological Studies of Values, Psychological Bulletin, (1955) V52, pp. 24-50
\end{flushright}

\begin{flushright}
\textsuperscript{101} Ibid.
\end{flushright}
It is designed to measure the relative emphasis within an individual of six major values - social, aesthetic, religious, political, theoretical, and economic.

The general theoretical question being raised at this point is whether or not these six master values cover the range of human values. The answer to this question, of course, is no. However, this issue is not raised for facetious purposes. Rather, it is to point out that the Allport - Vernon scale may be considered not general enough on the one hand and too general on the other. Thus one might expect that the values measured by the scale would not necessarily be related to a wide range of human behavior. Conversely these six values may be conceptually broken down into a number of sub-values. It is possible that the sub-values might be related to specific areas of behavior to a much greater extent than the master values. For example, the religious value might be conceived along a liberal - conservative dimension, the

102 G. W. Allport and P.E. Vernon, A Study of Values (Boston: Houghton Mifflin, 1931) (Revised in 1951 in collaboration with G. Lindzey.)

theoretical value along an applied-pure axis, etc.

Now, it will be recalled that a value was defined as the degree of worth or excellence ascribed to an object or activity. From this definition, it is clear that individuals may hold values over the entire gamut of human experience. It would, of course, not be economic to construct scales to measure values ascribed to every object or activity. On the other hand, it might be well worth while to construct scales to measure values beyond those measured by the Allport-Vernon scale. It is believed that extensions of the theory of values into specific sub-systems and even new major systems holds promise for the greater understanding of the role played by values in determining human behavior. An extension has been proposed in this study which involves work performance value systems.

Work Performance and Achievement

It has already been mentioned that there are many psychological variables which tend to predict behavior. The behavior which is being investigated in this study is work behavior. More specifically it is work behavior as related to performance. Performance behavior, however, may be subsumed under the broader concept of Achievement. This concept has been defined as "the degree or
level of success attained in an activity". The similarity of this definition to that of the criterion may be recognized. The crucial difference between the two concepts is that the criterion directly includes measurement. Achievement, however, is the more generic term. The value judgments inherent to the criterion are equally present in the concept of achievement, in the opinion of the writer.

There is a vast body of literature devoted to the prediction of achievement. Most studies, exclusive of the area of industrial psychology, are concerned with the prediction of academic achievement.

Values and Academic Achievement

There is some support in the literature for the hypothesis that value systems are related to academic achievement. Indirect evidence for this hypothesis has been the demonstration that value systems tend to be related to intelligence. A more direct connection was found by Rothney, who obtained small but significant associations between values and grades in certain subjects.

---

104 H. English and A. C. English, op. cit.


There is one particularly relevant psychological variable which has been found to be related to academic achievement behavior. This variable refers to the orientation towards achievement within an individual and has been called the achievement motive. 107 This motive may, in a sense, be referred to a more general system which could be termed achievement values. When viewed in this manner, it seems clear that achievement must be related to values in some way.

**Congruence of Values and Academic Achievement**

Bills' reports a study which demonstrates a significant relationship between the degree of congruence of value systems of instructor and student and the grades which the students receive in a course. 108 Dukes believes that the foregoing finding points up the fact that the value system of the achiever must be viewed in relationship to the value system of the person judging the achievement. 109 Moreover, value systems may facilitate or retard learning. Bills' study has a direct relationship to the second phase


109 W.F. Dukes, *op. cit.*
of this investigation. Rather than academic achievement, congruence of values will be investigated as it relates to work achievement. Moreover, the theory of value which guides the first phase of the study may be related in much the same manner to the second phase.
CHAPTER 111

STATEMENT OF THE PROBLEM

PHASE 1 - REFERENCE GROUP AND PERFORMANCE VALUE SYSTEMS

The theoretical rationale of the study has already been outlined in a rather general fashion. Prior to a more detailed discussion, this rationale will now be briefly restated. It is believed that values are inherently imbedded in the entire area of work performance and its measurement. With this assumption as a starting point, it has been postulated that virtually all members of an industrial society must hold values with respect to performance. The configuration of these values within an individual has been termed as performance value system. The study proposes to investigate the influence of two major variables upon performance value systems. It is hypothesized that these two variables will be associated with consistency in performance value systems. The two experimental hypotheses, however, may be logically deduced from a higher level hypothesis. This more general hypothesis states that consistency of performance value systems is a function of the homogeneity of the reference groups from which judges are drawn. Thus, the rationale for the two experimental hypotheses may be formulated in the following manner.
Hypothesis #1. Since supervisors may be considered as selected from a more homogeneous reference population, there should be a significant tendency for supervisory value systems to be more consistent than the value systems of non-supervisors.

Hypothesis #2. Since supervisors within the same or similar occupations may be considered as selected from a more homogeneous reference population there should be a significant tendency for supervisors classified as occupationally homogeneous to hold more consistent performance value systems than supervisors who are classified as occupationally heterogeneous.

Determiners of Supervisor - Performance Value Systems

It is well known that differing groups hold different values on various types of psychological dimensions. For example, people from one socio-economic group differ significantly from another in many ways e.g. attitudes, moral values, etc. Indeed membership in a certain group seems to some extent to determine how an individual perceives reality. In considering work performance as measured by supervisor ratings, it seems reasonable to conceptualize the rater (supervisor) as engaging first in perceptual activity (observing the behavior of a group of workers) then making judgments about the adequacy of the worker behavior to some idealized standard of performance which the rater carries around with him. This idealized version of work performance which the supervisor uses as his yardstick for the rating of workers, is subject to many influences in its formation. These will now be reviewed.
Culture as a Variable - Social Desirability and the Perception of Work Performance.

The effect of the general culture upon this hypothetical idealized set of values or expectations has already been suggested. This seems to form the groundwork for almost all supervisor expectations. A supervisor who held values markedly different from that of the cultural milieu would be quite rare. As a matter of fact, it is hardly conceivable that a man could be appointed into a position of leadership over people in industry who was markedly deviant in this respect. It is therefore assumed that the general culture will have a real effect in the molding of performance.

value systems.

In this connection, Uhrbrock reported a study in which 724 statements often made concerning employees were scaled by foremen. High agreement was found among the subjects with respect to the relative merits of the statements for good work performance. An interesting facet of this study was the fact that students were also asked to judge the statements. Their results were almost

---

identical with the foremen. While this constitutes impressive evidence for the reliability of the scale values with judges of widely differing backgrounds, it raises a question as to what is actually being measured. It has often been demonstrated that there is a very large general factor in supervisors' expectations. Wherry has called this bias. Uhrbrock appears to have demonstrated that this general factor is not even specific to supervisors, but extends to other groups in our culture when they are asked to look at job performance. It would appear, therefore, that there are general work expectations which the culture implants in virtually all of its members regardless of station. In the case of Uhrbrock's study it will become clear that the items and methodology employed tended to maximize the calling out of a general set which could be hypothesized to generate out of cultural values.

A few of his items with their scale values follow:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Scale Value</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is outstanding in every way</td>
<td>10.6</td>
<td>1.05</td>
</tr>
<tr>
<td>Has a pleasing personality</td>
<td>8.0</td>
<td>3.60</td>
</tr>
<tr>
<td>Is nearly always well prepared</td>
<td>7.0</td>
<td>2.15</td>
</tr>
<tr>
<td>Is always asking for advice</td>
<td>4.0</td>
<td>1.00</td>
</tr>
<tr>
<td>Is inclined to make trouble</td>
<td>2.0</td>
<td>1.35</td>
</tr>
<tr>
<td>Is a complete failure</td>
<td>1.0</td>
<td>.01</td>
</tr>
</tbody>
</table>

111 R. Wherry, op. cit.
From this list, it seems evident that the last 3 items are very different from the first three. The first three in differing degree represent qualities which are highly valued in the culture while the last three seem to embody traits for which the culture could be expected to disapprove. It becomes moot then to decide whether the items are really specific to the work situation. It would be more defensible and parsimonious to attribute the consonance of the former and students to cultural factors rather than to those which arise directly from the work setting.

The question, however, immediately becomes whether there are specific factors with respect to the work setting and how one goes about measuring them. Regardless of whether or not values exist which are relatively idiosyncratic to the work situation, it seems clear that the general factor assumed due to the culture must be significantly reduced before such measurement becomes possible. When correlations in the .90's are found between two groups of judges one of which is experienced in evaluating work performance and the other not, it is quite obvious that there is little room left to find differences. One of the tasks of this study was to design a scale such that the general cultural factor would be minimized.
Supervision as a Variable

If a scale were devised where the effects of culture had been effectively minimized, it would seem to follow that if a random sample of judges were drawn from the general population there should be relatively little consistency among performance value systems. This latter statement summarizes the theoretical discussion about the rationale of forced-choice methodology.

Although the general cultural factor is not under investigation, if an estimate of the factor were available, this estimate might serve as a baseline for evaluating the effects of other variables. These other variables are presumed specific to the work setting and are derived from the reference group hypothesis.

The most general breakdown by reference group in the work setting is by a division of individuals into management and labor categories. In the case of this study, the breakdown is accomplished by dividing subjects according to whether they exercise supervisory authority or not. From the theoretical framework of the study, there are many reasons why the performance value systems of supervisors should tend to be more consistent than non-supervisors. These reasons include common experience, training, interests, personality patterns, and a common bond to the philosophy of management. The most general test of the supervision hypothesis
would be accomplished if the performance value systems of a random sample of supervisors were compared to the systems of a random sample of persons from the general population. This was not feasible in the present study because the general population sample was not available. As an estimate, of the consistency of the performance value systems resulting from the general cultural factor, a sample of workers was utilized. It should be apparent, however, from the previous discussion, that a sample of workers probably represents an upper-bound estimate. This conclusion is based upon the fact that the working population is more homogeneous than the general population. Nevertheless, supervisors in addition to being members of the working force, have the additional factor of supervision which by definition makes them as a group more homogeneous than the workers.

Hypothesis #1 states, therefore, that supervisory performance value systems should tend to more consistency than the performance value systems of workers.

Occupation as a Variable

It is the second hypothesis of this study that occupation in itself is a significant variable affecting supervisory performance value systems. It is a natural consequent then that the more heterogeneous the occupations under study, the more inconsistent the value systems
of the supervisors. There are important qualifications to this, however, for some of the other variables mentioned will exercise limits to such inconsistency of performance values. Moreover, the methodology of this study imposes certain restrictions. The converse hypothesis is that the more homogeneous the occupations studied, the more consistent will be value systems. This is the actual hypothesis which will be tested empirically in this study.

There are many reasons which would tend to lend support to this hypothesis. Some of these will be briefly reviewed here.

First of all, the activities being supervised are similar. One would therefore expect that problems arising in the same or a similar occupation would have more in common than problems in other occupations. Methods for dealing with such problems might be similar. Secondly, the people who are being supervised could be expected to come from roughly the same socio-economic backgrounds, have the same amount of education, and hold similar value orientations. Thus one might expect supervisors in the same or similar occupations to develop personnel handling methods along roughly parallel lines.

A third possible determinant of similar work values is the fact that the supervisors in the same or similar occupations are probably homogeneous along many dimensions. These dimensions
could include training and education, abilities, interests, personality etc. If they are similar with respect to these dimensions, it does not seem unreasonable to suspect that they also agree a great deal about what they expect in their personnel.

A fourth possibility is that with occupations a kind of sub-culture develops which may in turn initially have depended upon the first three factors but which grew to the point where it exerts an influence in its own right. The military is perhaps the best example of this, but one might conceive of an engineering or accounting sub-culture on the professional level. Likewise, it seems probable that there may be different sub-cultures extending downward to the skilled trade level. It seems doubtful, however, that any distinct sub-cultures could be expected on the unskilled and semi-skilled level.

These four possibilities then seem to support the reasonableness of the hypothesis on an a priori basis. The task of the study will be to determine whether or not the hypothesis can be confirmed empirically.

The relevancy of the hypotheses to the criterion problem may be summarized in the following way. It is recognized that supervisory judgments often represent the only foundation, in a practical sense, on which a criterion can be constructed.
Moreover, even where other measures are available, these other measures cannot be allowed to completely determine personnel decisions.

**OTHER POSSIBLE VARIABLES**

**Industry as a Variable**

The idiosyncratic expectations of a particular industry may have an effect upon supervisor perception and judgment of worker performance. Some industries, for example the oil and chemical, have reputations for being exceptionally "progressive" in their handling of workers while others, e.g. coal and textiles, do not enjoy such a reputation. It is possible, that systematic differences would be found in supervisor values as a result of this variable alone. This study, however, will not consider or attempt to isolate this effect.

**Company as a Variable**

The effect of company may likewise be important in the formation of supervisor judgments about performance. It is well known that most companies have some policies which may be peculiar to them. This may in turn effect supervisors in various ways.

**The Individual Supervisor as a Variable**

One of the most thoroughly confirmed laws of psychology has been the phenomenon of the individual differences. There is no reason to suspect that this phenomenon will not operate in
supervisor perception and judgment of work performance. Other factors must be taken into consideration. Since supervisors are, in a sense, a measuring instrument for the psychologist, it seems that improvement of this measuring instrument can come about only through a study of its nature - how it operates and to a degree - why it operates the way it does. Once some of this information becomes available, the measuring instrument may be evaluated as to its adequacy, how it might become more effective, etc. A number of variables have been suggested in a preliminary way as influencing the operation.

Two of these variables have been selected for empirical study. Two hypotheses have been formulated in order to test the effect of the experimental variables.

Hypothesis #1 - There should be greater consistency of performance value systems for members of a supervisory group than for members of a non-supervisory group (Supervision Hypothesis).

Hypothesis #2 - In addition to the effect postulated by Hypothesis #1, among members of a supervisory group there should be greater consistency of performance value systems for supervisors who are classified as occupationally homogeneous as compared to supervisors who are classified as heterogenous along this same dimension. (Occupation Hypothesis).
PHASE II - CONGRUENCE AND THE PERCEPTION OF WORK PERFORMANCE

Statement of the Hypothesis

Chapter 11 discusses a study which demonstrated a relationship between congruence of values of instructor and student, and academic achievement. The second phase of this study bears a resemblance to Bills' investigation in its methodological approach. However, the behavior involves work achievement rather than academic achievement, the values investigated are quite different, and the theoretical rationale utilized to explain the congruence phenomenon is based on different considerations. The hypothesis of the second phase of the study is as follows:

Hypothesis #3 - There will be a tendency for the variable, congruence of worker-supervisor performance value systems, to be positively associated with supervisory judgment of level of performance of the worker. (Congruence Hypothesis).

The Congruence Hypothesis and Performance Prediction

Previous discussion has made clear that there are many variables which may be and have been related to worker performance. An objection might be raised concerning the economy of adding another variable, congruence of values, to the plethora of

---

112 R. Bills, op. cit.
variables already investigated. This objection is quite legitimate, but may be answered if it is remembered that the main purpose of the study is not related to prediction as such. Rather, the purpose involves the empirical testing of a hypothesis which is derived from a theory of value. The first phase of the study investigates the effect of certain variables upon the performance value systems of supervisors. It is concerned with performance behavior in the abstract. The second phase of the study postulates a relationship between the similarity of value systems and the judgment of effectiveness of actual job behavior. It is very possible, however, that variables involving aptitudes and personality might predict future job success more efficiently than the variable, congruence of values. As a matter of fact, congruence is really not a predictor variable at all since it must depend upon the individual worker and supervisor interacting with one another for a period of time.

It is believed, however, that the congruence variable will be useful in two major respects. First, it should give more information about the processes which affect a final judgment of individual worker performance by the supervisor. Second, it may be of value in illuminating the area of supervisor - worker interaction especially as this interaction reflects the communication process.
In many respects this two-way breakdown underscores the two alternate explanations which may be conceived to underlie the experimental hypothesis. The efficacy of these explanations depends upon where the major locus of the phenomenon is placed. These two loci represent respectively the supervisory judgmental processes or the actual worker behavior. If the major emphasis is placed upon the judgmental processes of the supervisor, the manner in which changes have occurred in worker behavior as a result of the communication process is largely irrelevant. Since supervisor judgments are in fact the major concern of this study, it is upon the theoretical ramifications of this aspect that the experimental hypothesis rests. Nevertheless, there is a possible alternate explanation for the hypothesis which may be derived from learning. In this case, however, the major locus is in the behavior of the worker and not in the judging of that behavior by the supervisor. This learning theory paradigm will be briefly discussed at the conclusion of this chapter.

DERIVATION OF THE CONGRUENCE HYPOTHESIS.

Values and Supervisor Perception of Actual Work Behavior

One common factor in virtually every job situation is the necessity for a judgment as to how well an individual is performing the duties required by the job. This judgmental process may not
be formalized, but someone must evaluate a worker at one time or another. This evaluation may be conceived as representing a comparison between the worker's behavior and some hypothetical ideal standard. The various possible factors influential in building up this ideal standard within a supervisor has been discussed in Phase 1 of the study. The ideal standard of an individual supervisor may be conceived to manifest itself through the medium of his performance value system. When a supervisor observes worker behavior he values this behavior along a high - low continuum. It is believed that the value system of the supervisor must influence in a major fashion, the placement of behavior along this continuum by the supervisor. Thus, if a worker repeatedly exhibits behavior which is positively valued by a particular supervisor, this supervisor will positively value that worker's performance. The same process should hold true for a worker who exhibits behavior which is negatively valued by a supervisor, except that in this case the worker's performance would be judged as poor. Therefore, in a study of a group of workers and their supervisor, an investigator may properly assume that the workers have exhibited behavior which was positively or negatively valued by the supervisor. On a theoretical level, a supervisor's ordering of his workers on a continuum from best to worst should perfectly parallel the degree to which the workers
behavior was positive or negative with respect to the supervisory value system. Naturally, this argument is derived from the ideal case.

Worker Value System and Worker Behavior

It is a basic assumption of this study that the worker possesses a performance value system just as does the supervisor. There are many possible factors which influence the value systems development - not the least of which is the role played by the supervisor. This development, however, is not relevant at present. The essential point is that the worker does in fact, have a performance value system. Now if it is assumed that the worker will tend to behave in a manner consistent with his own value system, it is possible to directly derive the congruence hypothesis.

Congruence of Performance Values and the Perception of Performance Behavior.

Congruence of performance value systems is defined as the degree to which worker values approximate those of a supervisor and vice-versa. The necessary and sufficient conditions for the congruence hypothesis to hold true are as follows:

1. A supervisor must judge the appropriateness of work behavior through the medium of his own performance value systems.
2. Workers tend to behave in a manner consistent with their own performance value systems.

3. Workers who hold value systems which are relatively similar to that of the supervisor should tend to behave in a manner which is positively valued by the supervisor and vice-versa.

4. From the foregoing reasoning, it is only necessary that the supervisor value overall work performance in a manner which parallels the values placed on the discrete behavior exhibited by the worker.

AN ADDENDUM TO THE DERIVATION OF THE CONGRUENCE HYPOTHESIS.

It has been mentioned previously that there is an alternate rationale which might be invoked to derive the congruence hypothesis. This rationale, however, is not antithetical but rather may be considered complementary to the previous derivation. In the previous derivation, it was hypothesized that worker behavior was determined by the worker's performance value system. However, the valuing of this behavior by the supervisor was the result of the influence of the supervisory value system. It was stated that a supervisor's ordering of his workers on a continuum of best to worst could be predicted from an order constructed from the congruence of value systems.

It may be noticed that there are no stated antecedent conditions underlying the congruence dimension, i.e. there are no variables
postulated to account for the clearly implied distribution of the workers along this dimension. In most empirical studies of work performance, it is generally the case that a selected predictor variable has no necessary antecedent conditions associated with it. Thus, aptitudes, intelligence, personality, interests, etc., are all presumed to be relatively stable worker characteristics. These characteristics are then correlated with the performance behavior of the worker. The congruence dimension, however, involves not only worker characteristics but supervisor as well. A congruence score for an individual worker has no meaning except as related to the value system of the supervisor. Now, there are undoubtedly a number of variables which determine the nature of the performance value systems of the individual worker on the one hand and the supervisor on the other. It seems highly probable, however, that one of the important variables involve the interaction process between worker and supervisor. The value systems of both worker and supervisor may be partially regarded as a product of this interaction process. Interaction, in turn, implies communication. In the work setting this communication is intimately connected with the learning of job requirements by the worker. These job requirements, however, are interpreted by the supervisor. Thus, in a very important sense, the worker
must learn to some degree the performance value system of the supervisor. Nevertheless, workers learn the supervisory value system to differential degrees.

**Communication of Value - Expectations**

While the final overall judgment of the supervisor is utilized by the psychologist to validate his instruments, it must be remembered that supervisor and worker are interacting on a day to day basis. Furthermore, the worker's behavior is approximating or not, as the case may be, the supervisory value expectations. Good supervisory technique demands that the worker be apprised of the fact that his behavior is meeting expectations or not. The supervisor must therefore tell the worker in what respect his behavior is deficient. It is also considered good supervisory practice to let the worker know when his behavior is appropriate, although this is not nearly as widespread among supervisors as is the pointing out of deficiencies. In any case, communication takes place between supervisor and worker with the former doing the talking and the latter listening.

**Incorporation-Modification of Worker Value System**

The worker then reacts to the communication either positively thereby changing or modifying the behavior or not. The worker also gains the information that this particular behavior is either appropriate or inappropriate in the supervisors
opinion. He also may be able to perceive degrees of appropriateness depending upon the supervisor's reaction.

It should be apparent that continued interactions between worker and supervisor will lead over the course of time to a workers becoming cognizant of the various types of behavior which will be reinforced either positively or negatively. While the worker originally comes to the job situation with a ready-made work value system which derives from a number of factors including culture, family, previous work experience, etc., it also seems clear that the present job must have an effect upon this value system to a greater or lesser extent. The present supervisor must also play a crucial role in modifying or reinforcing this value system. A set of values with respect to the job may therefore be theorized in the worker as well as the supervisor. In the case of the supervisor it is values with respect to worker behavior. For workers, however, the values center around what is expected of himself with respect to his own job.

Knowledge of Results and Learning

It has frequently been demonstrated in the laboratory that knowledge of how one is performing frequently facilitates the learning process. These findings were first demonstrated by
Thorndike in the learning of lists of foreign words.\textsuperscript{113} It has also been clearly corroborated under certain conditions in the learning of motor tasks. Trowbridge and Cason using two groups of blindfolded subjects found that learning to draw three-inch lines was significantly increased when subjects were told how close they were coming to the criterion line. A comparable group who were not given this information did not improve their performance.\textsuperscript{114} Underwood reported a study involving the tracking and ranging of targets in which two groups of subjects were given differential amounts of feedback about their success in these tasks. It was found that the two groups differed significantly on the ranging but not on the tracking task.\textsuperscript{115} These results suggested that in order for the effect to be shown, the task must be relatively complex and

\textsuperscript{113} E.L. Thorndike, "An Experimental Study of Rewards", Teachers College Contra. Education (1933) No. 580


the subject not be in a position to easily check his progress.

An axiom of modern supervision demands that workers be adequately informed about their job requirements. It has been recognized that people cannot perform their jobs in the most satisfactory manner unless they are aware of what is expected of them. The responsibility for communicating this information rests with the supervisor. The worker, however, assumes some responsibility because he must learn the substance of the supervisory communication. There are a great many problems connected with this communication process between worker and supervisor. While the various factors which facilitate or inhibit communication are not under investigation in this study, it should be recognized that this question has important implications for the entire area of supervisor-worker relations. It will be recalled that in the 'knowledge of results' studies, the information communicated to the subject was relatively simple, i.e. his line was too long, his ranging was off when the light was not lit, etc. A theoretical model may, however, be constructed of the interaction of worker and supervisor using the 'knowledge of results' study as a paradigm. In this case the worker is the subject and the performance criterion represents the results - or success vs. failure. It is the supervisor who sets the criterion as well as
communicates information to the worker on how he is meeting this criterion.

It would follow then that workers who were given information about the criterion should learn faster and therefore perform their jobs better than workers who do not enjoy this advantage.

Congruence as Related to 'Knowledge of Results'

The communication between worker and supervisor is analogous to the 'knowledge of results' paradigm in that the worker is told which behavior is appropriate and which not. This does not necessarily have to be accomplished directly through the supervisor - his surrogate, perhaps a more experienced worker, would suffice. The crucial factor is that the worker gets feedback concerning his own behavior. Congruence represents the degree of accord between worker and supervisor about the appropriateness of the behavior. Congruence then can be taken as an indication of the degree of effectiveness of communication between worker and supervisor on the one hand. It can in a slightly different sense, be used as an indication of the awareness of the worker of the appropriateness of the behavior for his job as viewed by the person who is judging his behavior - his supervisor. If in the 'knowledge of results' the subject was asked how it was that he was able to improve his performance, he would undoubtedly reply that he was
told his line was in error by such and such and so he corrected it. He was able to profit by his mistakes in other words.

The methodological approach of the present study is quite similar to the previous example. If the investigator asks a subject why he was able to improve his performance he would in effect be asking the subject if he were aware of one of the experimental conditions of the 'knowledge of results' study. In the present study, however, the worker is asked what behavior is appropriate for his job. If the subject's performance value system is markedly different from the supervisor, there is an implication of a breakdown of communication between worker and supervisor. This breakdown may be referred to a number of factors which include deficiency of supervisory communication, failure for some reason by the worker to learn the content of the supervisory communication, and possibly an interaction effect between the first two variables. The crucial point for this study, however, is the fact that the workers performance value system which presumably guides his behavior is different from the supervisory performance value system. The worker does not seem to be aware of what behavior is appropriate and what is not, as judged by the supervisor. An example will perhaps make this point clearer. Suppose a worker positively values successful interpersonal interaction and negatively values
productivity. His supervisor, on the other hand, values these activities in the reverse order. It may be predicted, therefore, that the workers behavior on the job if it reflects his performance value system will be negatively valued by the supervisor. Nevertheless, the worker does not know what the requirements of the job are - again as viewed by the supervisor. It may be concluded, therefore, that in some way the communication process between worker and supervisor has gone awry.

SUMMARY

There are two phases to the study. The first investigates the role played by the variable of reference group in determining supervisory performance value systems. Two hypotheses were derived to test the general hypothesis.

Hypothesis # 1 - There will be a tendency for performance value systems of supervisors to be more consistent than the performance value systems of non-supervisors, (Supervision Hypothesis).

Hypothesis # 2 - In addition to the previous effect, there will be a tendency for supervisors who are classified as occupationally homogeneous to hold more consistent performance value systems than supervisors who are classified as occupationally heterogeneous, (Occupation Hypothesis).

The second phase of the study investigates the supposed relationship between congruence of worker - supervisor value systems and supervisor perception of individual worker effectiveness.
Hypothesis # 3 - There will be a tendency for the variable of congruence of worker - supervisor performance value systems to be positively associated with supervisor perception of the worth of individual worker performance.
CHAPTER IV

METHOD

INSTRUMENTS EMPLOYED

Performance Assessment Scale

Selection of the Items

A special rating scale was designed to measure the effects under study. (See Appendix A). It is based on the forced-choice rationale although differing in format. The scale consists of fifteen groups of six statements each. Five of the statements may be considered favorable and one unfavorable. The five statements in each group (pentad) were carefully selected to have relatively equal social desirability connotations (equal preference values). The unfavorable statement in each group was inserted only as a control device and was not used in the compilation of the data or in the analysis of the results. The necessity for these control statements will soon become clear. A sample group of statements (#5 from the scale) follows:

( ) Undertakes work on his own initiative.
( ) Has shown ability to improve his work.
( ) Has shown he can take on increased responsibility.
( ) Carries out assigned tasks well.
( ) Does not come up with alibis or excuses for his mistakes.
( ) Evidence for excessive use of alcohol.
The first five statements are considered positive in that almost all judges could be expected to view the behavior or qualities as desirable for an employee to possess. The last statement is considered unfavorable in that almost all judges view the behavior as quite undesirable.

The items were selected with the following criteria in mind:

1. They must be unambiguous in meaning.

2. They must refer as much as possible to behavior or traits for which the rater has had a reasonable opportunity to observe. As a corollary, it was felt imperative to avoid concepts or inferences about behavior which exceeded the average rater's sophistication. Statements which assumed the need for complicated psychological inferences were therefore eliminated.

3. The items must refer to behavior or qualities which a priori have some clear relationship to the work setting.

4. The items finally selected must lend themselves to a breakdown on some psychologically meaningful dimension.

5. The items must have no definite a priori pull in a positive or negative direction. The use of superlatives was therefore carefully avoided.

6. Items which seemed to imply some overall degree of success were not to be employed. Since the study was interested in the elements making up job success, statements indicating overall success could only confuse the issue.
Construction of Dimensions.

It was felt that while the main hypotheses of the study were not particularly dependent upon psychological dimensions built into the scale ancillary suggestions for future research could possibly be derived if the item could be classified in some fashion which was meaningful from a psychological standpoint. Accordingly, it was decided to include six dimensions which appeared to be particularly pertinent to the work setting. A description of these dimensions will be found in Appendix B.

Construction of the Scale

(a) Rationale for the method of rank order. It will be recalled that the items were selected with a view to controlling, as much as possible, social desirability connotations. If the items were to be judged individually, for example on a Thurstone type of scale, the scale values would probably all cluster near the high end. This would hamper discrimination. Further, with all of the items being positive, there might be an adverse affect upon the judges from a psychological standpoint. This could come about if the judge decided that since all of the qualities or behavior would be desirable for an employee to possess, there is not much point in trying to make discriminations. Through the use of rank order format, one forces, in a sense, the judge to make the maximum use of his
discriminative powers. Although the paired-companion method would yield equally good results, the procedure was rejected on practical grounds.

(b) Rationale for the Use of the Pentad Form

The pentad form was decided upon for a number of reasons. First, since it was desirable from a practical point of view to avoid paired-comparisons, it was also necessary to take maximum advantage of the rank order technique. There is a pragmatic upper limit to the rank order technique, however. From an a priori consideration and from a study of the literature, 116 it was decided that the use of the pentad form would provide maximum discriminatory power while avoiding building up resentment on the part of the judges towards the task.

(c) Placing the Items in Pentad Form

A group of items were selected which seemed to have no overwhelming pull in a positive direction. These items were then classified according to dimension by two judges. A total of 171 items remained upon which the judges agreed as to dimension. Each of these dimensions was assigned a number. The items representing the dimensions were then numbered from 1 to n for each dimension.

The pentads were then constructed by randomizing the serial appearance of the dimension in the various pentads and then from a table of random numbers selecting the item to represent the dimension. There were qualifications, however, to the randomizing procedure. First each dimension had to appear at least once in each pentad whether the statement representing it was positive or negative. Since each group contained one negative statement, this indicates in effect that one dimension was negative for each group. Secondly, it had been decided not to repeat statements in the scale. This, in turn, signified that once a statement had been selected, it was no longer in the sample of items.

Twenty-four groups of items were constructed. Inspection of the scale, however, revealed that two of the groups were so constituted as to almost insure that the items could be ordered in a pre-determined way. These two groups were therefore discarded.

**Pre-testing of the Scale**

A scale was then constructed consisting of the remaining twenty-two groups. A decision was then made to pre-test the scale on a limited number of subjects in order to eliminate groups for which agreement seemed to be too high or too low and to clarify or drop out items which were ambiguous. A sample of ten supervisors of four organizations were administered the scale. Occupations
represented were two welding foremen, two machinist foremen, three electronic engineers, one psychologist, one supervisor of car wash operations and one supervisor of clerk typists. In addition to the standard instructions, the judges were requested to circle items which they did not understand and also to make any comments about the scale itself. Of the twenty-two groups of pentads four coefficients of concordance ranged from .55 to .78, three from .40 to .53, five from .30 to .39, four from .20 to .29, three from .13 to .19 and three below .09. It was felt that the four groups with coefficients above .55 were too high for the purpose of the study while the three groups below .09 were too low. In the latter instance it was assumed that discrimination was so difficult that the reliability of the groups of statements were open to question or the groups were tapping idiosyncratic values of the supervisors. In the case of the four high groups, it was also believed that two sets of possibilities were operative. Either the groups lent themselves, despite precautions, to easy ordering on a social desirability dimension or the groups were particularly suited to the tapping of the general supervisor dimension. In either case, however, it did not seem likely that enough room would be left for the effect of occupation. Accordingly these seven groups were discarded and a new scale constructed consisting of the fifteen pentads.
in the middle range. Items in these groups which the supervisors had noted as ambiguous were revised. There were, however, only a few questionable items among these fifteen groups.

**Population**

Companies represented. It was decided to get as wide a range of supervisors included in the sample as possible which was consistent with later grouping of the supervisors into meaningful homogeneous occupational groups. A total of 28 of the 37 industrial, commercial, and hospital organizations contacted agreed to cooperate in the study. For a breakdown of these organizations, see Appendix C. Of these companies, nineteen sent completed returns, and are included in the sample. See Appendix C.

**Administration**

Arrangements were made through the personnel departments of the cooperating companies to have the subjects complete the scale. These supervisors, wherever possible were brought together in a group when they independently filled out the scale. The only information about the purpose of the study was in the written instructions on the form itself. In some few cases it was not possible to bring the supervisors together in groups and therefore the fact that the scale was to be independently completed was stressed by the personnel officer to the individual supervisors. Analysis of these few cases reveal no particular difference in the amount of consonance obtained.
While there was no time limit for the completion of the scale, it was observed in most cases to take on the average about twenty minutes. Virtually all supervisors completed the scale in thirty minutes.

The completion of the scale by the various supervisors was for the most part voluntary. The fact that the scale was being filled out for a research project which had no connection with the individual's organization was stressed. On the third page of the scale information was requested from the supervisor which included name, age, number of employees supervised, education completed, and the type of job supervised. Of this the only absolutely essential data was the type of job supervised. The respondents were requested to be as specific as possible about this. Anonymity in completion of the scale could, therefore, be afforded if the situation so warranted. It was felt that the additional information, however, might be of some value in case of ambiguity in results.

Occupations Represented

The following occupations have had at least one return and are therefore represented in the sample.

Skilled Trades - Industrial

1. Electric Arc Welders
2. Heli-Arc Welders
3. Heavy Equipment Machinists
4. Light Equipment Machinists
5. Tool and Die Makers
6. Mechanical Inspectors
7. Steel Fabricators
8. Oxy-acetylene Burners
9. Maintenance Men
10. Lab. Technicians
11. Automobile Mechanics
12. Printers

Semi-skilled trades - Industrial

1. Fire Extinguisher Assemblers
2. Rigmen
3. Blade Wrappers
4. Machine Operators (Razor Blade Manufacturers)
5. Lathe Operators
6. Production Honers
7. Press Operators
8. Miscellaneous Electronic Assemblers
9. Platers
10. Solderers
11. Coil Winders
12. Transformer Assemblers
13. Capacitor Assemblers
14. Electrical Assemblers
15. Confectionary Assemblers
16. Stock Room Shippers

Unskilled Trades

1. Janitors
2. Packers
3. Helpers
4. Car Washers
5. Filling Station Attendants

Clerical Workers

1. Stenographers
2. Clerk-typists
3. File Clerks
4. Billing Clerks
5. Statistical Control Clerks
6. Customer Contact Clerks
7. Bookkeepers
8. Shipping Clerks
9. Interviewers
10. Office Machine Operators
11. Instructors
12. Credit Authorizers
13. Clerk Checkers
14. Statistical Control Clerks

Semi-Professional and Managerial

1. Draftsman
2. Electronic Assembly Foreman
3. Salesman

Professional

1. Engineers - Analysts
2. Accountants
3. Psychologists

Miscellaneous Occupations

1. Chauffeurs
2. Bartenders
3. Waitresses
4. Unclassified Factory
5. Unclassified Clerical

The Makeup of the Homogeneous Groups

Breakdown is by occupational grouping. In order to qualify as a group for this study, however, each classification must have a minimum of four. The groups which met these requirements are the following:

1. Welders - Electric Arc and Heliarc
2. Machinists - includes tool and die makers - light and heavy equipment machinists
3. Lab Technicians

**Semi-skilled Trades**

2. Electronic Assemblers
3. Stock Room Shippers

**Clerical Workers**

1. Clerk Typists
2. Statistical Control Clerks
3. Customer Contact Clerks
4. Office Machine Operators
5. Clerk Checkers

**Semi-Professional - Managerial**

1. Electronic Assembly Foremen
2. Draftsmen

**Professional**

1. Engineers
2. Accountants

The above mentioned groups were arrived at by utilizing the following criteria:

1. For many of the homogeneous groups there was no question as to how they should be combined. This was especially true for those occupations which have a special type of identity. This identity can be justified mainly on the basis of special education and training as well as the specialized functions of the occupations. Homogeneous groups which a priori fit this category are the engineers, accountants, welders, draftsmen, machinists, and lab technicians.
2. In the case of the semi-skilled and unskilled industrial occupations, there was something of a problem insofar as determining the degree of homogeneity. It was decided to combine on the bases of two criteria – the degree of training required as outlined in the Dictionary of Occupational Titles and the type of activity involved. 117

3. The clerical occupations presented special problems in view of the fact that for virtually all of the occupations there were some common basic skills, viz. typing and filing. Specialization in the clerical area can be assumed to come about through the specialization and elaboration of various skills or of other types of abilities in some cases. Thus, the typists are pooled because of specialization in this one skill. The office machine operators are pooled using the same rationale. The customer contact clerks are grouped because the one function of meeting and dealing with the public sets them apart from the other clerical occupations. All groups in the clerical occupations are similarly constructed.

4. In some cases an individual supervisor had authority over different occupational groups. Thus a machine shop foreman might have included in his work force a clerk, a stock handler, etc. When this occurred the following was the basis for eventual classification.

(a) The main purpose of the department was decided and the auxiliary functions noted. A machine shop is set up primarily for the purpose of machining products and the major portion of the work force are machinists. Thus in this case the return would be grouped under machinist occupation.

(b) Where the main purpose of the department was obscure or mixed, the return was classified according to the occupation which was highest in terms of skill and training required. This occurred in only a few cases.

STATISTICS UTILIZED

(a) Coefficient of Concordance

The methodology of the study involves the ranking of statements by a number of different judges. This represents ordinal measurement. Only a non-parametric mathematical model is therefore appropriate. The Coefficient of Concordance, or \( \text{W} \) 118, 119 developed by Kendall allows the investigator to compare

118 M.G. Kendall, Rank Correlation Methods, (London: Griffith, 1948)

the amount of agreement of k sets of judges ranking n sets of objects. The statistic can range in value from 0.00 to 1.00 with 1.00 representing complete agreement among the judges and 0.00 indicating complete disagreement. In the case of this study the judges are supervisors and workers and the objects to be ranked are statements which represent the behavior of workers on the job.

Values of the Coefficient of Concordance were computed for all of the various groups of judges in the study for all fifteen pentads. The Coefficient therefore, represents the amount of agreement within these groups for each pentad.

Random Blocks Design

The main purpose of this phase of the study is to compare the amount of agreement between groups of judges combined on the basis of homogeneity of reference group. Comparisons, however, are to be made not for the individual pentad, but over all fifteen. Each group has a value of W for each pentad. It is known that a function of W is distributed approximately as chi square.\textsuperscript{120}

\[ \text{chi square} = K \times (N-1) \times W \]

\textsuperscript{120} Ibid.
This function of $W$, however, since it is not normally distributed cannot be properly used with a parametric statistical model which assumes normality. Nevertheless, a transformation of chi square which is chi does approximate normality. Accordingly all values of $W$ were transformed to the corresponding value of chi. Since the above-mentioned formula would in our sample groups result in different values of chi for the same value of $W$ because of the differing size of $k$, it was necessary to equalize all groups so that the same value of $W$ would result in the same value of chi. This could be properly done because with the statistical model chosen for the group comparisons, the size of $k$ is irrelevant. Therefore, a convention was decided upon in which the value of $W$ was multiplied by 100 and the square root extracted. The Random Blocks analysis of variance technique was selected as the appropriate statistical model. This technique is described by Edwards. Each homogeneous group has fifteen values of chi as does the heterogeneous group. Since the size of the groups make no difference for the analysis each group may be considered to represent one

\begin{footnote}{A.L. Edwards, Experimental Design in Psychological Research, (Rinehart & Co., New York) pp. 284-297}\end{footnote}
individual. The different pentads represent, for statistical purposes, different experimental conditions. Since the same subjects are involved among groups variance is correlated. As already indicated, each individual subject actually represents a group of subjects. The within groups variance is therefore uncorrelated. The interaction term, subject by condition, is the error term which will be used to test the effect. The mean square of the within groups variance will be divided by the mean square of the subject by condition term to obtain a value of F. In this study F has one (two groups of subjects) and fourteen (fifteen experimental conditions) degrees of freedom. A value of F which exceeded that value of F at the .05 level of significance, was taken to indicate that the hypothesis of no difference from homogeneous to heterogeneous group should be rejected. Values of F which do not attain this level of significance indicated that the hypothesis of no difference should be accepted.

Combining Probabilities to Test the Hypotheses

Since there is only one comparison involved in the test of Hypothesis #1 (Supervision Hypothesis) the probability of the obtained F may be directly computed.

Hypothesis #2, however, states the variable of occupation will have a significant effect upon the consistency of value systems. In order to test this hypothesis, fifteen groups of occupationally
homogeneous supervisors are being compared with the one group of occupationallly heterogeneous supervisors. Hypothesis #2, therefore, does not stand or fall upon any one of these comparisons, but rather upon all of them. It is necessary then to test the general hypothesis by combining the obtained probabilities of the fifteen independent comparisons into an overall probability figure. The most parsimonious method for obtaining this overall probability figure is by computing the probability which could be expected on a chance basis of the number of groups which exceeded a specified level of significance. For Hypothesis #2 of this study, this significance level has been set at .01

Plotting the Values of the Dimensions

It has previously been mentioned that the study was set up in such a way that the individual items could be related to psychological dimensions. This was ancillary, however, to the main purpose of the study. The methodology involved in obtaining the data for the main purpose of the study precluded powerful statistic analysis of the effect of the dimensions. Nevertheless, the relative emphasis by the various homogeneous groups on each of these dimensions has been tabled in Appendix D. Comparisons of the groups may be made but not statistically evaluated.
PHASE II. CONGRUENCE AND THE PERCEPTION OF WORK PERFORMANCE

Instruments Employed

There were two instruments employed in the second phase of the study, the Wonderlic Personnel Test and a modification of the Performance Assessment Scale. The Wonderlic was used to assess the effect of the variable of intelligence upon the congruence dimension and also upon the supervisory ranking of work performance. The PAS, which was described earlier in this chapter was used to measure worker value systems. The instructions of this, the employees', form of the PAS are somewhat different from those contained in the supervisory form. The instructions are both more explicit as to the rank order method and ask the subject to rank the statements from a slightly different standpoint.

Wonderlic Personnel Test

Since intelligence was a variable which might prove very significant both in terms of the criterion itself and also for the congruence scores, the Wonderlic Personnel Test was administered to all subjects. This test is a modification of the Otis Self-Administering Test of Mental Ability and is described in the
literature. 122, 123 Besides being a group test, the Wonderlic had
the crucial advantage of a short administration time. Published
reliability and validity figures are adequate. 124

SUBJECTS

Subjects were recruited through the personnel divisions of
three organizations, 18 engine lathe operators from a manufacturer
of electronic equipment, 12 shoe salesmen from a specialty shoe
store, and 30 janitors working in a neuropsychiatric hospital. The
janitors, however, were divided into two groups since there were
actually two supervisors involved whose men functioned independ-
ently. The subjects were drawn from the middle and lower skill
levels not from design, but from pragmatic considerations.
Generalizations from the results to higher skill levels therefore,
must be made with considerable caution.

122 O.K., Buros, Third Mental Measurements Yearbook

123 Wonderlic Personnel Test Manual, Psychological
Corp. New York.

124 Ibid.
ADMINISTRATION

All subjects were gathered together in their respective groups and the general nature of the experiment was described in very vague terms. It was explained that the exact purpose of the study could not be revealed because such might affect the results. The subjects were assured that the individual results would be held in strict confidence. Management, therefore, would be given no information about the findings for any individual. After the experiment the subjects would be given any information they might wish concerning the purpose and the ideas which were being tested. Virtually all those approached agreed to cooperate in the project.

The subjects were administered the PAS and the Wonderlic in small groups of from five to nine each. Testing for each of the main groups was completed within one working day. Average time per subject to complete the two tasks was thirty-five to forty minutes. The instructions for the PAS were read aloud to each group of subjects. Almost all subjects seemed to understand the procedure. In some few cases it was necessary to explain the rank order method in more detail. Standard instructions were given to the subjects for the Wonderlic test. The order of the tests administration was PAS followed by the Wonderlic. While random-
izing the order might have been desirable from a theoretical point of view, it was felt imperative that the PAS be completed in as non-threatening as atmosphere as possible. Administration of an intelligence test prior to the PAS might very well have raised defensiveness which might in turn affect results.

SCORING

The Wonderlic test was scored in the standard fashion. 125

The following procedure was followed in the scoring of the PAS:

1. Each individual form was examined. A form was considered unreliable and therefore discarded from the study if in two or more groups of statements the control negative statement was given a rank less than 5. If the instructions were understood and the subject actually complying with them, the negative statement should have been placed sixth or at best fifth for each group of statements. Accordingly four returns were discarded from the group of lathe operators, six from the janitors, and two from the sales group.

2. The remaining returns were then compared against that of the supervisor. Deviations of the worker rankings from that of the supervisor for each pentad were squared and totaled. This gave a deviation for the individual pentad. The final score was tabulated by adding together the deviation scores for all fifteen pentads. High congruence is therefore indicated by low total deviation scores and vice versa.

3. All of the total deviation scores were then placed in a distribution for statistical analysis.

---

125 Ibid.
CRITERION OF WORK PERFORMANCE

Since all of the groups of subjects were small (10 to 14) the rank order method of determining the quality of work performance was employed. Supervisors for all groups were instructed in the method. The supervisor placed his best worker number one, his second best number two through n, his worst worker. The aim of the study was to order the workers on a relative basis within each group. Absolute differences among workers between groups were not considered.

STATISTICS

Kendall's Tau

In order to test the main effect under study correlations were computed between the rank order of the congruence measure and the rank order on the criterion measure using Kendall's statistic tau. This statistic is described in the literature. 126, 127 The statistic can take on values from -1.00 to -1.00. It should be noted that tau gives values generally smaller than that of the other commonly used non-parametric statistic, Spearman's rho, unless the correlations are perfect in either a positive or negative

126. M.G. Kendall, op. cit.

127 S. Siegel, op. cit.
direction. Both statistics, however, use the same amount of information from the data, and both are equally significant even though the value of tau is smaller than that of rho.\textsuperscript{128} The sampling distribution of tau is virtually normal given n's of the magnitude used in this study. In order to test the effect of intelligence upon both the congruence measure and the criterion measure, tau for all groups was computed.

To investigate the effect of the congruence measure with intelligence held constant, partial tau was also computed for all groups. Unfortunately the sampling distribution of partial tau has not yet been established and therefore, tests of significance cannot be made.\textsuperscript{129} Combining Probabilities

In view of the fact that all of the samples are relatively small and consequently extremely high values of the correlation would have to be obtained in order to reach significance, it was decided to test the overall hypothesis by computing the exact probability of the results for each group. When this information

\begin{itemize}
\item \textsuperscript{128} Ibid.
\item \textsuperscript{129} Ibid.
\end{itemize}
is available, the probabilities may be combined using the associated values of chi square. In this manner, the chi squares may be added and the total chi square tested for significance using the appropriate degrees of freedom.
CHAPTER V

RESULTS

PHASE #1 - HOMOGENEITY OF GROUP AND PERFORMANCE VALUE SYSTEMS

Hypothesis #1 - Supervision as a Determiner of Consistency of Value Systems

A total of forty-nine employees representing the occupations of shoe salesmen (n=10), janitor (n=14), industrial therapist (n=11) and engine lathe operator (n=14) were utilized. Coefficients of concordance were computed for the employees and supervisors over all fifteen pentads. The coefficients for the two groups are presented in Table 1.

TABLE 1
PENTAD COEFFICIENTS OF CONCORDANCE OF HETEROGENEOUS EMPLOYEES AND HETEROGENEOUS SUPERVISORS

<table>
<thead>
<tr>
<th>Pentad</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>Pentad mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suprs.</td>
<td>151.24.23.38.27.21.20.26.25.18.28.42.36.13.45.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The values in Table 1 were then transformed as described in Chapter IV and a comparison made between the two groups by means of the random blocks design. It was hypothesized that if there was an effect that could be traced to supervision itself, then the heterogeneous supervisors should agree significantly more than the group of relatively heterogeneous employees.
Table II presents the statistical analysis of the data of Table I transformed as described in Chapter IV.

**TABLE II**

**HYPOTHESIS NO. 1**

**STATISTICAL ANALYSIS OF RELATIVE AGREEMENT OF EMPLOYEES VERSUS SUPERVISORS**

<table>
<thead>
<tr>
<th></th>
<th>Supervisors Effect</th>
<th>Pentad Effect</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>mean square</td>
<td>F</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>43.0</td>
<td>3.51*</td>
</tr>
</tbody>
</table>

*Significant beyond .05 level F 4.60
** Significant beyond .01 level F 8.86

It may be observed from Table II that the supervisors agree significantly more among themselves on the scale than do the employees. The coefficients of concordance presented in Table I for the employees may be conceived to result from a number of different factors. The first of these factors undoubtedly revolve around the social desirability of the item. A second factor may be the possibility of common expectations of employees regardless of job. In any case, it seems that the social desirability of the items of the scale could not account for an appreciably greater degree of agreement than that found and tabulated for the heterogeneous employees. If it were possible to eliminate some of the other factors operative in employees but not in non-employees, the agreement
resulting from social desirability might be even less. It should also be noted, however, that the group of heterogeneous employees is far less heterogeneous on the variable of occupation than the supervisory group. The resulting supervisory effect is probably therefore somewhat attenuated. Regardless, it seems reasonable to suppose that the greater agreement found among the supervisor's must result from factors associated with supervision itself. If would follow then that the PAS scale is a relatively potent instrument for the investigation of the supervisory effect.

Hypothesis #2 - Occupation as a Determiner of Consistency of Value Systems.

The confirmation of Hypothesis #1 has demonstrated that the variable of supervision which acts to make a group of judges more homogeneous also is associated with greater consistency of performance value systems. Hypothesis #2 states that there should be significantly greater agreement among supervisors who are drawn from an occupationally homogeneous reference than among those drawn from heterogeneous occupational groups. It will, of course, be noted that the heterogeneous supervisors of Hypothesis #2 were considered the more homogeneous in Hypothesis #1. According to Hypothesis #2, however, the occupationally homogeneous supervisors will be more consistent.
Coefficients of Concordance for the fifteen groups of homogeneous judges, therefore, should be higher than those for the heterogeneous judges. Table 111 presents this data.

**Table 111**

COEFFICIENTS OF CONCORDANCE OF FIFTEEN GROUPS OF HOMOGENEOUS SUPERVISORS AND ONE GROUP OF HETEROGENEOUS SUPERVISORS

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machinists</td>
<td>5</td>
<td>.50</td>
<td>.36</td>
<td>.68</td>
<td>.42</td>
<td>.54</td>
<td>.52</td>
<td>.68</td>
<td>.55</td>
<td>.70</td>
<td>.33</td>
<td>.33</td>
<td>.55</td>
<td>.38</td>
<td>.71</td>
<td>.64</td>
<td>.53</td>
</tr>
<tr>
<td>Welders</td>
<td>9</td>
<td>.54</td>
<td>.73</td>
<td>.79</td>
<td>.41</td>
<td>.61</td>
<td>.68</td>
<td>.60</td>
<td>.44</td>
<td>.17</td>
<td>.40</td>
<td>.37</td>
<td>.56</td>
<td>.14</td>
<td>.61</td>
<td>.44</td>
<td>.50</td>
</tr>
<tr>
<td>Laboratory</td>
<td>4</td>
<td>.51</td>
<td>.69</td>
<td>.78</td>
<td>.46</td>
<td>.51</td>
<td>.44</td>
<td>.78</td>
<td>.43</td>
<td>.42</td>
<td>.66</td>
<td>.71</td>
<td>.59</td>
<td>.48</td>
<td>.78</td>
<td>.34</td>
<td>.57</td>
</tr>
<tr>
<td>Technicians</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock Room</td>
<td>4</td>
<td>.38</td>
<td>.73</td>
<td>.54</td>
<td>.44</td>
<td>.40</td>
<td>.29</td>
<td>.49</td>
<td>.34</td>
<td>.24</td>
<td>.49</td>
<td>.75</td>
<td>.35</td>
<td>.59</td>
<td>.78</td>
<td>.34</td>
<td>.47</td>
</tr>
<tr>
<td>Shippers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Electric</td>
<td>20</td>
<td>.30</td>
<td>.35</td>
<td>.29</td>
<td>.31</td>
<td>.41</td>
<td>.33</td>
<td>.54</td>
<td>.10</td>
<td>.30</td>
<td>.14</td>
<td>.81</td>
<td>.09</td>
<td>.57</td>
<td>.21</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Operators</td>
<td>13</td>
<td>.17</td>
<td>.28</td>
<td>.33</td>
<td>.26</td>
<td>.40</td>
<td>.12</td>
<td>.40</td>
<td>.33</td>
<td>.09</td>
<td>.49</td>
<td>.55</td>
<td>.52</td>
<td>.13</td>
<td>.54</td>
<td>.10</td>
<td>.31</td>
</tr>
<tr>
<td>Electronic Operators</td>
<td>7</td>
<td>.35</td>
<td>.60</td>
<td>.71</td>
<td>.37</td>
<td>.59</td>
<td>.26</td>
<td>.74</td>
<td>.28</td>
<td>.23</td>
<td>.35</td>
<td>.56</td>
<td>.54</td>
<td>.62</td>
<td>.70</td>
<td>.44</td>
<td>.49</td>
</tr>
<tr>
<td>Electronic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foremen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clerk</td>
<td>14</td>
<td>.25</td>
<td>.04</td>
<td>.21</td>
<td>.27</td>
<td>.13</td>
<td>.12</td>
<td>.20</td>
<td>.38</td>
<td>.32</td>
<td>.41</td>
<td>.62</td>
<td>.49</td>
<td>.10</td>
<td>.26</td>
<td>.23</td>
<td>.27</td>
</tr>
<tr>
<td>Typists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer Cont.</td>
<td>4</td>
<td>.69</td>
<td>.28</td>
<td>.57</td>
<td>.50</td>
<td>.43</td>
<td>.35</td>
<td>.41</td>
<td>.69</td>
<td>.61</td>
<td>.39</td>
<td>.92</td>
<td>.81</td>
<td>.21</td>
<td>.35</td>
<td>.28</td>
<td>.50</td>
</tr>
<tr>
<td>Clerks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical</td>
<td>4</td>
<td>.70</td>
<td>.38</td>
<td>.64</td>
<td>.18</td>
<td>.31</td>
<td>.62</td>
<td>.51</td>
<td>.52</td>
<td>.69</td>
<td>.60</td>
<td>.62</td>
<td>.89</td>
<td>.74</td>
<td>.78</td>
<td>.73</td>
<td>.60</td>
</tr>
<tr>
<td>Clerks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Mach.</td>
<td>6</td>
<td>.43</td>
<td>.35</td>
<td>.56</td>
<td>.26</td>
<td>.13</td>
<td>.26</td>
<td>.26</td>
<td>.06</td>
<td>.52</td>
<td>.64</td>
<td>.30</td>
<td>.42</td>
<td>.54</td>
<td>.55</td>
<td>.44</td>
<td>.19</td>
</tr>
<tr>
<td>Operators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clerk</td>
<td>4</td>
<td>.68</td>
<td>.33</td>
<td>.58</td>
<td>.51</td>
<td>.78</td>
<td>.36</td>
<td>.33</td>
<td>.28</td>
<td>.29</td>
<td>.25</td>
<td>.65</td>
<td>.25</td>
<td>.50</td>
<td>.31</td>
<td>.33</td>
<td>.43</td>
</tr>
<tr>
<td>Checkers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draftsmen</td>
<td>5</td>
<td>.39</td>
<td>.64</td>
<td>.55</td>
<td>.49</td>
<td>.49</td>
<td>.27</td>
<td>.22</td>
<td>.21</td>
<td>.16</td>
<td>.64</td>
<td>.78</td>
<td>.59</td>
<td>.55</td>
<td>.28</td>
<td>.38</td>
<td>.44</td>
</tr>
<tr>
<td>Engineers</td>
<td>8</td>
<td>.39</td>
<td>.34</td>
<td>.64</td>
<td>.65</td>
<td>.32</td>
<td>.48</td>
<td>.28</td>
<td>.39</td>
<td>.49</td>
<td>.73</td>
<td>.69</td>
<td>.58</td>
<td>.60</td>
<td>.24</td>
<td>.31</td>
<td>.48</td>
</tr>
<tr>
<td>Accountants</td>
<td>4</td>
<td>.86</td>
<td>.80</td>
<td>.83</td>
<td>.60</td>
<td>.53</td>
<td>.76</td>
<td>.71</td>
<td>.87</td>
<td>.70</td>
<td>.98</td>
<td>.65</td>
<td>.56</td>
<td>.85</td>
<td>.39</td>
<td>.47</td>
<td>.71</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>151</td>
<td>.24</td>
<td>.23</td>
<td>.38</td>
<td>.27</td>
<td>.21</td>
<td>.20</td>
<td>.26</td>
<td>.25</td>
<td>.18</td>
<td>.28</td>
<td>.42</td>
<td>.36</td>
<td>.13</td>
<td>.45</td>
<td>.11</td>
<td>.27</td>
</tr>
</tbody>
</table>
It may be observed by inspection of Table 111 that most of the homogeneous groups present consistently higher coefficients than does the heterogeneous group of supervisors. The question has arisen as to whether or not the differences between the homogeneous and heterogeneous groups might be explained on the basis of the size of the N's within the homogeneous groups, i.e. with decreasing values of the size of N the coefficients increase. In order to rule out this possibility a series of random samples were drawn from the data with differing sizes of N ranging from 4 to 20. The magnitudes of the obtained coefficients were then correlated to sample size by means of Spearman's statistic, rho. A small insignificant negative relationship was found. It was therefore concluded that sample size in itself could not account for the magnitude of the coefficients for the homogeneous groups.
TABLE IV
HYPOTHESIS #2

STATISTICAL COMPARISONS OF EFFECT OF HOMOGENEITY OF OCCUPATION BETWEEN FIFTEEN GROUPS OF HOMOGENEOUS OCCUPATIONS AND ONE GROUP OF HETEROGENEOUS OCCUPATIONS

STATISTICAL ANALYSIS OF PENTAD EFFECT

<table>
<thead>
<tr>
<th>Occupation</th>
<th>N</th>
<th>Occupation</th>
<th>df</th>
<th>mean square</th>
<th>F</th>
<th>df</th>
<th>mean square</th>
<th>F</th>
<th>df</th>
<th>mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machinists</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>34.6</td>
<td>49.6***</td>
<td>14</td>
<td>10.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Welders</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>259.2</td>
<td>26.4***</td>
<td>14</td>
<td>21.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Lab. Tech.</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>447.4</td>
<td>176.1***</td>
<td>14</td>
<td>17.4</td>
<td>6.8**</td>
</tr>
<tr>
<td>Stock Room</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>224.8</td>
<td>43.1***</td>
<td>14</td>
<td>19.1</td>
<td>3.6*</td>
</tr>
<tr>
<td>Shippers</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>224.8</td>
<td>43.1***</td>
<td>14</td>
<td>19.1</td>
<td>3.6*</td>
</tr>
<tr>
<td>Non-Elect. Operators</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.2</td>
<td>3.7</td>
<td>14</td>
<td>1.8</td>
<td>3.2*</td>
</tr>
<tr>
<td>Electronic Operators</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>6.0</td>
<td>1.5</td>
<td>14</td>
<td>30.3</td>
<td>7.3**</td>
</tr>
<tr>
<td>Electronic Foremen</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>246.9</td>
<td>29.1***</td>
<td>14</td>
<td>17.8</td>
<td>2.1</td>
</tr>
<tr>
<td>Clerk</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.8</td>
<td>0.1</td>
<td>14</td>
<td>24.9</td>
<td>2.7*</td>
</tr>
<tr>
<td>Typists</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.8</td>
<td>0.1</td>
<td>14</td>
<td>24.9</td>
<td>2.7*</td>
</tr>
<tr>
<td>Customer</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>259.6</td>
<td>34.3***</td>
<td>14</td>
<td>23.8</td>
<td>3.1*</td>
</tr>
<tr>
<td>Cont. Clks</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>259.6</td>
<td>34.3***</td>
<td>14</td>
<td>23.8</td>
<td>3.1*</td>
</tr>
<tr>
<td>Statistical Clerks</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>484.7</td>
<td>37.1***</td>
<td>14</td>
<td>15.4</td>
<td>1.2</td>
</tr>
<tr>
<td>Office Mach. Operators</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>62.6</td>
<td>4.4</td>
<td>14</td>
<td>20.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Clerk</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>139.9</td>
<td>11.4**</td>
<td>14</td>
<td>13.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Checkers</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>139.9</td>
<td>11.4**</td>
<td>14</td>
<td>13.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Draftsmen</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>153.1</td>
<td>13.0**</td>
<td>14</td>
<td>19.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Engineers</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>173.9</td>
<td>22.6***</td>
<td>14</td>
<td>18.3</td>
<td>2.4</td>
</tr>
<tr>
<td>Accountants</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>805.5</td>
<td>272.1***</td>
<td>14</td>
<td>17.7</td>
<td>6.0**</td>
</tr>
</tbody>
</table>

* Significant beyond .05 level F > 4.6 for 1 and 14 df
** Significant beyond .01 level F > 8.7 for 1 and 14 df
*** Significant beyond .001 level F > 17.1 for 1 and 14 df

* Significant beyond .05 level F > 2.5 for 14 and 14 df
** Significant beyond .01 level F > 3.7 for 14 and 14 df
From Table IV it may be observed that of the fifteen occupationally homogeneous groups, eleven differ significantly from the heterogeneous group in the amount of agreement on the PAS scale. Of these eleven groups which are significant beyond the .01 level, nine are significant far beyond the .001 level. The probability of obtaining by chance as many as eleven groups out of fifteen differing at the .01 level from the reference group is far beyond the .001 level of significance. While this is clearly a highly significant finding and indicates that the overall null hypothesis may be rejected with great confidence, it does not explain the variability of the individual homogeneous groups with respect to the sub-hypotheses. These sub-hypotheses concern outcomes in the individual homogeneous groups. The null hypothesis for four of these sub-groups could not be rejected. This indicates that the overall hypothesis must be restricted somewhat. While it is possible to argue that these four groups were not as homogeneous in terms of occupation as the other eleven, it is also equally clear that they are more homogeneous along this dimension than the heterogeneous group. The failure to confirm for these four homogeneous groups therefore, must be construed to weaken the generality of the overall hypothesis.
However, it should be noted that none of these groups can be said to have a marked occupational identity. Furthermore, all these groups fall in either the semi-skilled or clerical occupations, and the two clerical groups do not represent very high skill levels. It is possible that the overall hypothesis will have to be reformulated when an investigator studies occupations below the higher skill levels. Unfortunately, there was not enough data to analyze unskilled occupations. When such data becomes available, it will be possible to state whether agreement is a function of skill level as well as occupation itself.

Tables II and IV also present F values which are specific to the pentad effect. This effect refers to differential responding by the subjects with occupational group held constant over the fifteen pentads. These F values are for the most part non-significant. Significant values can be traced for the most part to defects in the scale itself.

**Dimensional Analysis of the PAS by Occupational Group**

The PAS was constructed with an additional purpose in mind. While this purpose was of secondary importance to the main hypotheses of the study, it was believed that a charting of the various occupational groups relative emphasis along the six dimensions might prove of general interest. This information is summarized in Appendix D.
Reliability of the PAS

In order to test the stability of the results obtained, the PAS was administered to a group of ten supervisors three to six months after the initial administration. Table V presents the average Spearman rank order correlation over the fifteen pentads. These data were obtained by computing correlations between successive administrations of the scale for each subject and then averaging the correlations over all ten subjects.

**TABLE V**

**SPEARMAN RANK ORDER CORRELATIONS BETWEEN SUCCESSIVE ADMINISTRATIONS OF THE PAS N=10 - PENTAD RELIABILITY OF THE PAS**

<table>
<thead>
<tr>
<th>Pentad</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.74</td>
<td>.57</td>
<td>.52</td>
<td>.61</td>
<td>.44</td>
<td>.67</td>
<td>.82</td>
<td>.41</td>
<td>.44</td>
<td>.74</td>
<td>.78</td>
<td>.59</td>
<td>.61</td>
<td>.59</td>
<td>.51</td>
</tr>
</tbody>
</table>

From the data of Table V, it may be observed that pentad reliability of the PAS varied from .41 to .82. An average correlation over all fifteen pentads was calculated as .61. These reliability figures are somewhat lower than the reliability generally considered desirable for a scale. It should be pointed out, however, that the index of reliability employed in the study, Spearman's rho, was very crude. Minor reversals in the ranking of five statements
markedly lowered the correlations obtained. Another possible factor affecting the reliability of the scale was the methodology of the study. With so much effort devoted to the minimizing of the general cultural effect, familiar cues were at a minimum. In a sense then, the procedure worked against the obtaining of very high reliability. It seems clear, however, that the scale can and should be revised in an effort to improve its reliability.
CHAPTER V

RESULTS - PHASE II

CONGRUENCE AND THE PERCEPTION OF WORK PERFORMANCE

The hypothesis under test in phase II of this study concerns an association postulated between the congruence of worker and supervisor value systems on the one hand, and supervisor ranking of the worker along the performance dimension on the other. The test hypothesis states that there should be a significant tendency in a positive direction for workers who agree with their supervisor to be ranked higher than workers who disagree. This tendency or lack of same will be demonstrated in the form of correlations.

Table VI presents data found for the four independent groups which have been described in Chapter IV.

TABLE VI

TAU CORRELATIONS BETWEEN THE VARIABLE OF WORKER - SUPERVISOR CONGRUENCE AND SUPERVISOR RANKING OF WORKER

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Tau</th>
<th>Z</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janitors (Group 1)</td>
<td>14</td>
<td>.20</td>
<td>0.99</td>
<td>.1611</td>
</tr>
<tr>
<td>Janitors (Group 11)</td>
<td>10</td>
<td>.51</td>
<td>2.06</td>
<td>.0197</td>
</tr>
<tr>
<td>Inside Shoe Salesmen</td>
<td>10</td>
<td>.56</td>
<td>2.27</td>
<td>.0116</td>
</tr>
<tr>
<td>Engine Lathe Operators</td>
<td>14</td>
<td>.28</td>
<td>1.38</td>
<td>.0838</td>
</tr>
</tbody>
</table>
From Table VI it may be observed that all of the correlations are in the predicted direction. Two of the four are significant beyond the .05 level of significance and one correlation approaches this level. Since the groups are independent of each other and the exact probabilities are known, a chi square test of the overall hypothesis is appropriate. Table VII was constructed by converting each probability by means of the following formula:

\[ \text{Chi square} = -2 \log_e p \]

### TABLE VII

**COMBINING OF CHI SQUARE TO OBTAIN OVERALL PROBABILITY**

<table>
<thead>
<tr>
<th></th>
<th>Exact Probability</th>
<th>Chi Square</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janitors (Group 1)</td>
<td>.1611</td>
<td>3.654</td>
<td>2</td>
</tr>
<tr>
<td>Janitors (Group II)</td>
<td>.0197</td>
<td>8.032</td>
<td>2</td>
</tr>
<tr>
<td>Inside Shoe Salesmen</td>
<td>.0116</td>
<td>9.016</td>
<td>2</td>
</tr>
<tr>
<td>Engine Lathe Operators</td>
<td>.0838</td>
<td>4.964</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>25.666</strong></td>
<td><strong>8</strong></td>
</tr>
</tbody>
</table>

* Significant beyond .05 level \( \chi^2 \geq 15.51 \)
** Significant beyond .01 level \( \chi^2 \geq 20.09 \)
*** Significant beyond .001 level \( \chi^2 \geq 26.12 \)

From Table VII it may be observed that the significance for the overall hypothesis is beyond the .01 level and approaches significance at the .001 level. This finding indicates that the
overall null hypothesis may be rejected with confidence and is taken as a confirmation of the test hypothesis that worker-supervisor congruence as measured by the PAS is significantly related to supervisor perceptions of the individual worker's performance. Although all four groups exhibit this relationship, they do so to differential degrees. There are no clues as to the reasons for this variability from the data. The fact, however, that the two groups of janitors (which are very comparable) show such a wide disparity may constitute preliminary evidence that occupation in itself is not of crucial importance in accounting for the magnitude of the congruence effect. A great deal more research, of course, will be needed before this question can be definitely settled.

INTELLIGENCE AND SUPERVISOR RANKING OF INDIVIDUAL WORKER PERFORMANCE

It is well known that intelligence tests of workers tend to correlate with supervisors' rankings of performance. Intelligence test data were gathered for the groups under study not to investigate the effect of intelligence on performance rankings as such, but rather as a means of evaluating its effect upon the variable of congruence. Table V11, however, presents the correlations between intelligence and performance rankings as a matter of general interest. Since there were no preliminary predictions as to
direction of the correlations, a two tailed test was employed to figure the exact probabilities.

**TABLE VIII**

TAU CORRELATIONS BETWEEN THE VARIABLE OF WORKER INTELLIGENCE AND SUPERVISOR RANKING OF WORKER ASSOCIATED VALUES OF CHI SQUARE

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Tau</th>
<th>z</th>
<th>Probability</th>
<th>df</th>
<th>Chi Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janitors (Group 1)</td>
<td>14</td>
<td>.33</td>
<td>1.64</td>
<td>.10</td>
<td>2</td>
<td>9.606</td>
</tr>
<tr>
<td>Janitors (Group 11)</td>
<td>10</td>
<td>.13</td>
<td>.448</td>
<td>.654</td>
<td>2</td>
<td>.848</td>
</tr>
<tr>
<td>Inside Shoe Salesman</td>
<td>10</td>
<td>.19</td>
<td>.658</td>
<td>.512</td>
<td>2</td>
<td>1.338</td>
</tr>
<tr>
<td>Engine Lathe Operators</td>
<td>14</td>
<td>.15</td>
<td>.745</td>
<td>.454</td>
<td>2</td>
<td>1.580</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>8.362</td>
</tr>
</tbody>
</table>

*Significant beyond .05 level chi square ≥15.51

From Table VIII it is obvious that the obtained values of chi square in no case reach significance nor even approach significance although all the correlations are positive. This is the result primarily of two factors. A two tailed test was employed and therefore all probabilities are double what they would be using a one tailed test. Secondly, the correlations reported for these four groups although somewhat lower than those reported in the literature are still roughly comparable. The small number of subjects in each group, however, naturally raise the probability associated with each group. A correlation with an N of 200 of the same magnitude
would undoubtedly be significant. There is nothing in this data which would discredit a hypothesis that intelligence is significantly related to work performance as perceived by supervisors.

Worker - Supervisor Congruence and Worker Intelligence

Of far more importance for this study is the question of the relationship of the variable of congruence to that of intelligence. If these two variables are very significantly related, a natural conclusion would be that congruence score is only another method of measuring a subject's intelligence. This would, of course, nullify to a large extent the practical and theoretical significance of the previously reported findings. Table IX presents an analysis of the relationships found in the four groups between congruence scores and intelligence.

TABLE IX

TAU CORRELATIONS BETWEEN THE VARIABLE OF WORKER - SUPERVISOR CONGRUENCE AND WORKER INTELLIGENCE - ASSOCIATED VALUES OF CHI SQUARE

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Tau</th>
<th>z</th>
<th>Exact Probability</th>
<th>df</th>
<th>Chi Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janitors (Group 1)</td>
<td>14</td>
<td>.12</td>
<td>.599</td>
<td>.548</td>
<td>2</td>
<td>1.206</td>
</tr>
<tr>
<td>Janitors (Group 11)</td>
<td>10</td>
<td>.09</td>
<td>.312</td>
<td>.750</td>
<td>2</td>
<td>0.576</td>
</tr>
<tr>
<td>Inside shoe salesmen</td>
<td>10</td>
<td>.18</td>
<td>.622</td>
<td>.534</td>
<td>2</td>
<td>1.254</td>
</tr>
<tr>
<td>Engine Lathe Operators</td>
<td>14</td>
<td>.13</td>
<td>.647</td>
<td>.518</td>
<td>2</td>
<td>1.316</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.352</td>
</tr>
</tbody>
</table>

*Significant beyond .05 level chi square ≥ 15.51
From Table IX it may be observed that although all of the correlations are in a positive direction none are significant nor do the associated values of chi square total to any level of conventional significance. It may therefore be concluded that intelligence in itself has little relationship to congruence scores, - at least in the groups under study.

**Congruence and Supervisor Ranking of Worker Performance with Intelligence Effect Partialled Out**

Since both intelligence and the variable of congruence correlated positively with supervisor rank of the individual worker and in turn congruence and intelligence correlated positively for all groups, it would be of interest to estimate the relationship of congruence alone to supervisor rank with the effects of intelligence partialled out. Table X presents the results of such a statistical analysis.

**TABLE X**

**RELATIONSHIP OF WORKER - SUPERVISOR CONGRUENCE TO SUPERVISOR RANK OF WORKER - INTELLIGENCE PARTIALLED OUT**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Partial Tau</th>
<th>Tau-Congruence to Rank (From VI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janitors (Group 1)</td>
<td>14</td>
<td>+.18</td>
<td>+.20</td>
</tr>
<tr>
<td>Janitors, (Group 11)</td>
<td>10</td>
<td>+.50</td>
<td>+.51</td>
</tr>
<tr>
<td>Inside shoe salesmen</td>
<td>10</td>
<td>+.54</td>
<td>+.56</td>
</tr>
<tr>
<td>Engine Lathe Operators</td>
<td>14</td>
<td>+.27</td>
<td>+.28</td>
</tr>
</tbody>
</table>
From Table X it may be seen that the effects of partialling out intelligence are quite negligible upon the original correlations. There is, at present, no means of determining the significance of the statistic employed, partial tau. It seems clear, however, that intelligence as such has little bearing upon the results reported in Table VI, i.e., the relationship of worker-supervisor congruence to supervisor ranking of workers.
CHAPTER VI
CONCLUSIONS AND DISCUSSION

PHASE I - REFERENCE GROUP AND PERFORMANCE VALUE SYSTEMS

Conclusions

1. Supervision is crucially related to the content of work performance value systems. Supervisory experience tends to make judges more consistent among themselves with respect to assigning value to work behavior.

2. The variable of occupation is also significantly related to the consistency of judges' performance value system. Within the groups of supervisory subjects, homogeneity of occupation was found to be associated with increased consistency in performance value systems.

3. The confirmation of these two hypotheses indirectly tends to confirm the overall hypothesis which postulates a relationship between the degree of homogeneity of reference group and consistency of performance value systems.

4. There is variability in the degree to which individual groups of homogeneous supervisors agree upon the adequacy or appropriateness of work behavior. Of the fifteen occupationally homogeneous groups in the study, eleven differed from the heterogeneous group to a very high degree - significant beyond the .01 level. Four of the homogeneous groups, however, did not reach significance although two of these four approached significance at the .05 level. It can be concluded, therefore, in view of this lack of complete confirmation of the sub-hypotheses, that generalizations cannot be safely made to individual sub-groups other than those for which the findings were positive. In this connection,
however, it should be pointed out that the number of subjects in most of the groups was quite small and further research in needed. Investigations of the sub-groups where the findings were inconclusive should, of course, be made with even more caution.

**Discussion - Implications for Future Research**

It is believed that the corroboration of the reference group hypothesis has significant implications for the criterion problem in industrial psychology. This belief is based upon a number of factors.

First, the study has demonstrated the usefulness of viewing performance behavior within a general framework of values. This general orientation should prevent us from becoming involved in a never-ending search for some ultimate standard against which work behavior is to be evaluated. In the final analysis, the desirability of behavior of one human being must be referred to the value system of another. Nevertheless, value systems are not absolutely idiosyncratic. Rather, values tend to be shared. One of the major problems of this study was in determining which variables seem to be important for the sharing of these values. As a first approximation, it was decided to investigate the effect of homogeneity of reference group upon this phenomenon. The experimentally tested variables were almost inevitable given the basic framework of the study and the nature of the behavior under
investigation. It is apparent, however, that the variables of supervision and occupation do not exhaust all the possibilities inherent in the homogeneity of reference group hypothesis. Some of these possibilities for future research include the effects of industry, company, department, and socio-economic group upon performance value systems.

A second inference may be drawn from the finding of relatively unique performance value systems in the various occupations under study. It seems clear that the worker behavior is judged on different bases depending upon the occupation in which they are engaged. The hypothetical "good performance versus poor performance" is thus seen to be a very complex affair inasmuch as the same work behavior will be judged differentially appropriate or desirable in two different occupations. For example, an individual worker with good interpersonal skills will find this talent differentially valued in an occupation such as machinist as against an occupation such as electronic assembler. While for some occupations, this type of differentiation might be made on the basis of a job description, it seems equally clear that this deduction could not be made in the above-mentioned example just by a study of the job descriptions.
From the foregoing, a third inference may be drawn. This concerns rating scale methodology. It would seem to follow that if there are inter-occupational differences in the manner in which judges value performance behavior, the rating scales must be constructed to take this variable into account. The omnibus type of rating scale which gives equal weight to a wide variety of performance behavior will illustrate this point. In certain of the clerical occupations, the assignment of equal weight to behavior suggesting promotional potential as opposed to effective interpersonal behavior, is unjustified. In this instance, harmonious interpersonal behavior is much more positively valued than behavior which suggests possibilities for upward mobility.

Forced-choice theorists have recognized this problem and have suggested that a new scale be constructed for every new job under study. In like manner, critical incidents may not be generalized to jobs which differ markedly from the investigated job. It is the opinion of the writer that the seeming lack of generality of the foregoing rating scale techniques lies in a failure to appreciate fully the crucial importance of the degree of homogeneity of judges of work behavior in different job situations.

There is, of course, more to the criterion problem than merely investigating the factors which make judges performance
value systems more consistent. There are a great many problems of even greater interest to psychologists. These might include the effects upon worker behavior of various types of supervisor value systems, the effect of deviancy of supervisory value systems upon supervisory behavior, effecting changes in supervisory value systems through psychological means, etc. It is believed, however, that the adoption of the concept of value on the theoretical level brings these problems more sharply into focus.

In summary, the first phase of this study has attempted to demonstrate that the reference group of the judge has a crucial significance upon the whole criterion problem. The criterion problem may be understood to a greater extent if it is recognized that the problem intrinsically involves subjective judgments. The ordering of these subjective judgments can be most effectively made if care is taken in the definition and delineation of the significant reference group from which the judges are drawn.
CHAPTER VI

PHASE II - CONGRUENCE AND THE PERCEPTION OF WORK PERFORMANCE

Conclusions

1. The overall hypothesis of worker-supervisor congruence being positively associated with supervisor ranking of individual worker performance is clearly confirmed with a significance beyond the .01 level and approaching the .001 level obtained.

2. There is variability in the magnitude of the effect of the congruence variable among the four groups of independent subjects. No conclusive inferences could be drawn from the data as to the reasons for this variability. However, it was suggested that occupation as such does not seem to play too great a role in this variability.

3. Intelligence does not seem to be of crucial importance in determining worker-supervisor congruence, although consistent low correlations in a positive direction were found between the two variables.

Discussion - Implications for Future Research

The congruence variable was investigated as a means of furthering the extension of a theory of value in the area of work performance. While the first phase of the study postulates a relationship the degree of homogeneity of certain groups and the consistency of performance value systems, the second phase predicts a relationship between worker-supervisor performance value systems and the way supervisors value actual work behavior.
The derivation of the congruence hypothesis as described in Chapter IV, does not depend upon variables presumed influential in the building of performance value systems. Unlike the first phase, therefore, the etiology of value systems is not the subject of inquiry. It was theorized that if workers tended to behave in accordance with their own value systems and if a supervisor judged this behavior from the framework of his value systems, then the congruence phenomenon should occur. There are, however, further implications of the congruence hypothesis.

In the second phase, values are presumed to affect actual behavior and the perception of actual behavior while the first phase is primarily concerned with behavior in the abstract, i.e. with performance value systems. The primary purpose of Phase I was to apply the concept of value to the criterion problem. The congruence hypothesis, however, transcends the criterion problem. This conclusion should be apparent if it is remembered that a criterion refers specifically to the measurement of work behavior. The congruence hypothesis involves not only the valuing of work behavior by the supervisor, but a relationship between value systems and the actual behavior of workers. The second phase of the study, therefore, may be viewed as an investigation of work performance in general and not merely in
its perception. The 'knowledge of results' paradigm was introduced more as a suggestion for future research rather than as an explanatory model. In this respect, it seems clear that a crucial aspect of work performance involves the communication process between supervisor and worker. The etiological factors in the building of value systems are more complicated than the reference group hypothesis would imply. The reference group hypothesis does not adequately encompass these etiological factors. Performance values must be transmitted from supervisor to worker and vice-versa. Mutual communication of values imply learning by either worker or supervisor or both.

A natural consequence of such a learning process should be more congruence between worker and supervisor. It has been demonstrated, however, that there is considerable variability along the congruence dimension in a group of workers. This variability would seem to imply differential learning by the workers. Of great significance, however, is the fact that unlike most learning, congruence is not appreciably dependent upon the variable of intelligence. What variables, therefore, underlie the congruence dimension? It is believed that promising clues may lie in non-intellectual factors in the worker, in the supervisor, and in their dynamic interaction. In any case, it would seem that
low congruence may be taken as an indication of a breakdown in communication between worker and supervisor. If the congruence dimension is regarded as a measure of the effectiveness of worker - supervisor communication, a great many research avenues are opened upon problems in the area of worker performance. These problems involve not only the prediction of performance, but also questions relating to the process of worker - supervisor interaction, variables involving the facilitation and retardation of the learning of job requirements, etc. Another major area where the congruence dimension would seem to have special relevance is that of worker morale.

Another promising avenue of research may lie in the investigation of congruence as related to group behavior. Not only is there variability along the congruence dimension among individuals, but also among groups. Congruence has been shown to be positively associated with individual worker performance. An important extension would be the demonstration that congruence is also positively associated with group performance.

From the foregoing, it is concluded that the congruence dimension may have important implications in our understanding of the behavior of people in the work setting.
SUMMARY

A study is described which advocates that the area of work performance be viewed within a general value framework. It was divided into two phases. The first phase demonstrated a positive relationship between the degree of homogeneity of reference group and the consistency of performance value systems. Two sub-hypotheses were used to test this relationship. It was found that the performance value systems of supervisors were significantly more consistent than the performance value systems of non-supervisors. It was demonstrated that occupationally homogeneous supervisors were significantly more consistent in performance value systems than supervisors drawn from heterogeneous occupations.

The second phase of the study postulated a positive relationship between the congruence of worker - supervisor value systems and supervisor rank of individual worker effectiveness. Four independent groups were used to test the hypothesis. All of the groups exhibited the predicted effect with two of the four reaching statistical significance.

The implications of these findings for the area of worker performance were discussed.
APPENDIX A

PERFORMANCE ASSESSMENT SCALE

Instructions

Following are a series of statements which describe behavior of people on the job. They form groups of six statements each.

The purpose of the scale is to find out the relative merits towards overall job success of the behavior outlined in each statement. It is possible that many of the statements in each group would be considered desirable for an employee to possess and perhaps some which are irrelevant or undesirable. As a result, it may sometimes be difficult to decide between them. It is necessary for the study, however, that a ranking in order of importance for job success be made of the statements within each group of six.

No doubt you can think of various types of jobs in which the order would be different, but for this scale, it is important to consider only the job types which you directly supervise.

Read the statements for each group carefully. Then choose the statement which in your opinion is most essential or most desirable in an employee working for you. This statement is to be numbered "1" in the box to the immediate left. Then choose the statement which is not quite as essential or desirable as "1", but which is more important than the remaining statements and number this "2". Following the same procedure rank the remaining statements "3", "4", "5", and finally "6" - the statement which is least essential or perhaps even undesirable.

Do not spend too much time making your judgments or considering all the possibilities which might make one statement better than another. Consider only the overall importance of the statement to effective job performance.

1. ( ) Seems to "catch on" to new skills easily.
   ( ) Tolerant of fellow workers faults or peculiarities.
   ( ) Cannot make decisions easily.
   ( ) Asks for work when he has completed an assigned task.
   ( ) Keeps his tools or equipment in good repair.
   ( ) Is expert in his own line of work.
2. ( ) Always comes to work on time.
   ( ) Capable of taking over when the occasion demands.
   ( ) Gets along well with fellow workers.
   ( ) Does not seem too interested in his job.
   ( ) Makes few mistakes.
   ( ) Continually improves his knowledge of his job.

3. ( ) Shows little potential for increased responsibility.
   ( ) Good attendance record.
   ( ) Friendly manner.
   ( ) Seems to have a good deal of aptitude for his job.
   ( ) Production is better than average.
   ( ) Works well without being driven or urged.

4. ( ) Understands easily what he is to do without need for much explanation.
   ( ) No management problem.
   ( ) Has demonstrated a capacity for leadership.
   ( ) Seems willing to learn.
   ( ) Bungles routine tasks.
   ( ) Is orderly in his work habits.

5. ( ) Undertakes work on his own initiative.
   ( ) Has shown ability to improve his work.
   ( ) Has shown he can take on increased responsibility.
   ( ) Carries out assigned tasks well.
   ( ) Does not come up with alibis or excuses for his mistakes.
   ( ) Evidence for excessive use of alcohol.

6. ( ) More skillful than the average in his job.
   ( ) Has been a troublemaker.
   ( ) Appearance is neat and clean.
   ( ) Keeps busy all day without direction.
   ( ) Capable of making quick decisions.
   ( ) Never makes the same mistake twice.

7. ( ) Well liked by his fellow workers.
   ( ) Seems to do the best that he can.
   ( ) Rarely comes up with new ideas.
   ( ) Brings to the job a high degree of skill.
   ( ) Has had extensive experience in his line of work.
   ( ) Never quits ahead of time.
8. () Falls behind in his work.
   () Encourages others to do their best.
   () Maintains a steady, sustained pace in his work.
   () Knows the requirements of his job thoroughly.
   () Not easily discouraged.
   () Has a good many worth-while ideas.

9. () Willing to assume the more unpleasant tasks connected
   with the job.
   () Completes his tasks in a reasonable length of time.
   () Attention wanders from his work.
   () Helps new people get adjusted to the job.
   () Well informed about new advances in his field.
   () Works well under pressure.

10. () Requires few instructions.
    () Has many complaints.
    () Willing to work overtime to get the work out.
    () Is a perfectionist in his work.
    () Likes to take on responsibilities.
    () Has better than average ability to plan and lay out
        his work.

11. () Seems to like his job.
    () Never injured on the job.
    () Necessary to explain to him almost every part of the job.
    () Other workers have difficulty keeping up with him.
    () Can work with just about anybody.
    () Looks for new and improved methods for doing the work.

12. () Cooperates well with supervisor
    () Understands the fine points about his trade or profession.
    () Has never damaged company equipment.
    () Evades the difficult jobs.
    () Work is accurately performed.
    () Tries to learn the more complicated skills of higher
        paying jobs.

13. () Handles routine jobs well.
    () Devotes time and effort to master the job.
    () Does not learn new skills easily.
    () Always does what he is told.
    () Can depend upon him to get the work out on time.
    () Inspires confidence in those around him.
14. ( ) Work is slovenly or inaccurately performed.
   ( ) Works after hours or comes in at night when it is desirable.
   ( ) Handles customers or clients well.
   ( ) Knows what to do next without being told.
   ( ) Anticipates supervisor's wants and needs.
   ( ) His abilities fit in well with the job requirements.

15. ( ) No undesirable personal habits.
   ( ) Gives up easily on the difficult jobs.
   ( ) Helps others with their work whenever possible.
   ( ) Is in sympathy with the long-range goals of the organization.
   ( ) Has made new contributions to his field.
   ( ) Always makes his production quota.

The following information is requested in order to facilitate the analysis of our results and also to make them more meaningful. Would you kindly supply the following.

_________________________  ___________________
(Name)  (Age)  (No. employees supervised)

Education completed ____________________________

Types of jobs you supervise (welders, stenos, clerks, etc.) ________

_________________________________________________________________________

Approximate length of time in your present field________________________

Length of time as a supervisor _______ Length of time-present

Position __________________________

How would you say your people compare with an ideal group of workers? Please circle the appropriate point on the scale.

Very Poor  Poor  Fair  Average  Good  Very Good  Excellent
APPENDIX B

DIMENSIONS OF THE PERFORMANCE ASSESSMENT SCALE

1. Motivation. The importance of adequate motivation for successful performance is well recognized and accepted not only among psychologists but also by personnel people, top management and first-line supervisors as well. There is much evidence that supervisors are ready to tolerate such things as inexperience as long as the worker seems "willing to learn". Items are included in this category which refer to interest in the job, sticking to the job in the face of unpleasant circumstances, working to potential capacity, and willingness or eagerness to learn and perform the job.

The motivation category relates only to the individual's particular job and does not refer to his potential upward mobility. This latter aspect of work performance while certainly dependent to some extent upon motivational factors is also clearly related to other traits and qualities such as ability to handle people, intelligence, etc.

2. Work Habits. This category refers to an individual's "style of work" so to speak. It includes such things as attendance and tardiness record, traits such as neatness, steadiness, and perfectionism in the work, personal habits which have no direct
interpersonal reference on the job such as addictions, and care taken to protect himself and company property. It seems that this is an omnibus category and probably is more complex than any other with the possible exception of #6 Promotional Potential. The category also seems to stress behavior and traits of a more routine nature and could consequently be expected to be more important in lower level types of jobs.

3. **Interpersonal Harmony.** It is also well recognized that individuals who have difficulty getting along with other people are likely to cause many problems in social settings. Since the work situation is social in nature, there is no reason to believe that this will not be the case there also. It should also be realized, however, that the extent of the importance of this category probably will vary from situation to situation. Nevertheless similar occupations could be expected to be relatively uniform on its importance. Items in this category refer specifically to the individual's interpersonal relations both with co-workers and supervisor.

4. **Job Knowledge and Aptitude.** There are clearly two dimensions operative in this category. The reasons for the amalgamation of these two relatively discrete dimensions were partly pragmatic. It is difficult to construct many items for them which are not virtually synonymous. The task might therefore raise undue resistance from
the raters. The most important reason, however, was theoretical. It was felt that while both of these dimensions were highly related to the fifth category - skill, there were probably major differences as well which could be obscured if all three were thrown together. It is well known that there are classes of individuals who have a great deal of theoretical knowledge about the job, but who seem unable to put this knowledge to practical use (as the skill category implies). Further it was believed that different kinds of jobs may differentially value the knowledge dimension. For certain high level jobs (such as diagnostic physician), knowledge is of crucial importance while skill may be secondary.

Aptitude likewise was felt to be probably different from skill in vital respects. For certain types of jobs only minimal skill may be required (e.g. entry jobs) and most of the training needed is provided directly on the job. In this case ability to "catch on" (aptitude for the job) is a crucial factor. Conversely, in other occupations (e.g. the skilled trades) ability to do the job has presumably been developed to the point where aptitude for the work can be assumed. Aptitude and knowledge while dissimilar in many respects, may be viewed as different steps of the learning process. Items are included under aptitude where this is mentioned by name or seems to be clearly implied. Knowledge items include those
where it is explicitly stated or behavior is described which implicitly assumes a high degree of knowledge.

5. **Skill** This dimension is one which has been found in many studies to account for a general common factor in job requirements. Items which represent this dimension on the scale utilize the word 'skill', or natural concomitants of skill such as production record.

6. **Promotional Potential** It is commonplace in the industrial setting to look for employees who seem to have the potential for increased responsibility. While it is usually understood that persons considered eligible for promotion possess qualities and abilities tapped by the other dimensions of the scale, it is also equally clear that upward mobility is dependent upon a number of other qualities and abilities. This dimension was designed to measure the relative importance over various types of jobs of these relatively unique qualities and abilities. Items representing what is commonly assumed to be behavior or traits representing promotional potential include such things as leadership, bearing and interest in increased responsibilities, the use of the decision-making process, interest in the company and improving its products or services, successful dealings with customers and the like. This dimension, like dimension 11 (Work Habits) is complex. It also sometimes blends almost imperceptibly into other dimensions. A reasonably high motivation, for example, is almost
a necessary condition for advancement as is getting along with other people. The postulation of a unique set of traits and abilities is represented by Promotional Potential.
APPENDIX C

COMPANIES AGREEING TO COOPERATE

2. Steel Fabricating Companies
   1. Steel Wire Mfg. Company
   2. Confectionary Companies
   1. Metallurgical Company
   1. Insurance Company
   4. Electronic Equipment Companies
   2. Chemical Manufacturers
   1. Machine Tool Company
   2. Department Stores
   2. Car Wash Companies
   1. Can Company
   1. Milk Company
   2. Printing Companies
   1. Neuropsychiatric Hospital
   2. Restaurants
   1. Aircraft Company
   2. Machine Shop Companies
   2. Gasoline Stations
   1. Trucking Company

COMPANIES INCLUDED IN THE SAMPLE

2. Steel Fabricating Companies
   1. Confectionary Company
   1. Insurance Company
   4. Electronic Equipment Companies
   1. Machine Tool Company
   2. Department Stores
   1. Car Wash Company
   1. Neuropsychiatric Hospital
   1. Restaurant
   1. Machine Shop Company
   1. Aircraft Company
   1. Gasoline Station
## APPENDIX D

### RANK BY OCCUPATIONAL GROUPS OF SIX DIMENSIONS OF PAS*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Machinists</td>
<td>5</td>
<td>3.25</td>
<td>3.62</td>
<td>3.92</td>
<td>2.51</td>
<td>2.11</td>
<td>2.62</td>
</tr>
<tr>
<td>Welders</td>
<td>9</td>
<td>3.26</td>
<td>3.63</td>
<td>3.86</td>
<td>2.27</td>
<td>2.17</td>
<td>2.70</td>
</tr>
<tr>
<td>Lab. Technicians</td>
<td>4</td>
<td>3.36</td>
<td>3.98</td>
<td>3.73</td>
<td>2.27</td>
<td>2.17</td>
<td>2.60</td>
</tr>
<tr>
<td>Shippers</td>
<td>4</td>
<td>3.00</td>
<td>3.81</td>
<td>3.27</td>
<td>2.65</td>
<td>2.73</td>
<td>2.56</td>
</tr>
<tr>
<td>Non-Elect. Operators</td>
<td>20</td>
<td>3.16</td>
<td>3.68</td>
<td>3.39</td>
<td>2.52</td>
<td>2.67</td>
<td>2.64</td>
</tr>
<tr>
<td>Elect. Ops.</td>
<td>13</td>
<td>3.21</td>
<td>3.68</td>
<td>3.16</td>
<td>2.62</td>
<td>2.70</td>
<td>2.71</td>
</tr>
<tr>
<td>Elect. Foremen</td>
<td>7</td>
<td>3.24</td>
<td>3.60</td>
<td>3.37</td>
<td>2.52</td>
<td>2.85</td>
<td>2.50</td>
</tr>
<tr>
<td>Clerk Typists</td>
<td>14</td>
<td>2.89</td>
<td>3.57</td>
<td>3.04</td>
<td>2.73</td>
<td>2.70</td>
<td>3.02</td>
</tr>
<tr>
<td>Cust. Cont. Clerks</td>
<td>4</td>
<td>2.40</td>
<td>3.52</td>
<td>3.14</td>
<td>2.90</td>
<td>2.92</td>
<td>3.08</td>
</tr>
<tr>
<td>Off. Mach. Operators</td>
<td>6</td>
<td>2.85</td>
<td>3.18</td>
<td>3.22</td>
<td>2.73</td>
<td>2.92</td>
<td>3.35</td>
</tr>
<tr>
<td>Clerk Checkers</td>
<td>4</td>
<td>2.64</td>
<td>3.69</td>
<td>3.00</td>
<td>2.46</td>
<td>3.03</td>
<td>3.14</td>
</tr>
<tr>
<td>Draftsmen</td>
<td>5</td>
<td>3.13</td>
<td>3.73</td>
<td>3.48</td>
<td>2.26</td>
<td>2.68</td>
<td>2.80</td>
</tr>
<tr>
<td>Engineers</td>
<td>8</td>
<td>2.89</td>
<td>4.15</td>
<td>3.57</td>
<td>2.38</td>
<td>2.75</td>
<td>2.33</td>
</tr>
<tr>
<td>Accountants</td>
<td>4</td>
<td>2.83</td>
<td>4.21</td>
<td>3.92</td>
<td>1.96</td>
<td>2.23</td>
<td>2.79</td>
</tr>
<tr>
<td>Heterogeneous</td>
<td>151</td>
<td>3.06</td>
<td>3.69</td>
<td>3.34</td>
<td>2.52</td>
<td>2.67</td>
<td>2.77</td>
</tr>
</tbody>
</table>

* Large values indicate the dimension was considered relatively unimportant. Small values indicate the dimension was considered relatively important. The average of five ranks would, of course, be '3.00'.
APPENDIX E

PERFORMANCE ASSESSMENT SCALE

FORM E

Instructions

On the following pages you will find statements which describe many ways in which people behave or act on their jobs. These statements are put together in groups. Each group contains six statements.

The purpose of this research is to find out how the people who are actually doing the job feel about the importance of these statements.

In looking them over, you will see statements which may seem absolutely essential for a person to possess in order to do a good job if he were doing your type of work. Others may not appear as essential and may perhaps even be undesirable.

What you are asked to do is to read the statements in each group carefully. Then make a ranking of the statements in each group in order of desirability for your work.

For example, Suppose you were asked to be a judge or critic of six movies. You then saw these movies. You first were asked to pick the one you thought was the best. Then you were asked to pick the one that was next best. You continued to choose the rest of the movies in order of merit. If you had a list of movies and marked on the list the one which you thought was best with a 1, and the one which was next best with a #2, and so on down the line to the #6 movie, you would have a setup very close to the one here. Of course #6 would be the sixth best movie which would naturally make it the worst of the group.

The rules to be followed here are exactly the same, but instead of movies you will be judging statements.
After reading each group of statements, pick the one which in your opinion, would be most desirable or essential in a person doing your type of work. In the box to the left of the statement put #1. Then choose the statement which is next most essential or desirable and mark this #2. Then pick the statements which are 3rd, 4th, 5th and 6th in importance in order for a person to do a good job. #6 will be the statement which seems to you to be the least essential or perhaps even undesirable for a person to possess.

There are fifteen groups of statements. Follow the same rules for each group.

Do not spend too much time making up your mind about the statements or considering all the possibilities which might make one statement more important than another. Consider only the overall importance of each statement for doing a good job in your line of work.

Scale items same as Appendix A.
CONFIDENTIAL

Name ___________________ Age _______ Marital Status ______

No. Dependents _______ Job Title ______

How long (approximately) have you worked in your present job? ________

How long have you worked in your present Dept.? ________________

Length of time in your field?

Have you worked in related fields? ____ (Yes or No)

If so, how long? ________________

How promising do your chances look for advancement in your present job? Circle one.

None Very Poor Poor Fair Average Good Very Good Excellent

How does your present supervisor measure up to your ideal boss?

(a) In terms of ability? Circle one.

Very Poor Poor Average Good Excellent

(b) In terms of the way he treats his people? Circle one.

Poor Fair Average Good Excellent

(c) In terms of how well you like him as a person? Circle one.

Poor Fair Average Good Excellent

How does your present job measure up to what you would really like to do if you had the chance? Circle one.

Very Poor Poor Fair Average Good Very Good Excellent

We would like to thank you for cooperating in this research project. We hope that you have found the various tasks interesting.
BIBLIOGRAPHY


BIBLIOGRAPHY

Duffy, E., A Critical Review of Investigation Employing the Allport Vernon Study of Values on Other Tests of Evaluative Attitude, Psychological Bulletin V37, 1940.


BIBLIOGRAPHY


BIBLIOGRAPHY


BIBLIOGRAPHY


Sisson, E. D. Forced-Choice The New Army Rating Personnel Psychology, 1948, VI.


Thorndike, E. L. An Experimental Study of Rewards, Teachers College Contra. Education, 1953.


Thurstone, L. L. Ability, Motivation and Speed, Psychometrika V2, 1937.

Tiffin, J. and R. T. Phelan, The Use of the Kuder Preference Record to Predict Turnover in an Industrial Plant, Personnel Psychology V6, 1953.

Toops, H. A. The Criterion, Education and Psychological Measurement V4, 1944.
BIBLIOGRAPHY


Uhrbrock, R.S. Standardization of 724 Rating Scale Statements, Personnel Psychology V3, 1950.


PERCEPTION OF WORK PERFORMANCE BY WORKERS AND SUPERVISORS
(Library of Congress No. Mic 59)

James Taylor Mahoney, Ph.D.
Boston University Graduate School, 1959

Major Professor: Professor Henry Weinberg

Background and Purpose

Work performance is an area of major interest for psychologists in their study of human behavior. The present study will investigate two important problems in this area. Phase I is concerned with criterion development. It is postulated that the bases which supervisors utilize in judging work performance are dependent upon the occupational group of which they are members as well as membership in the supervisory group, i.e. that membership in these groups to some extent determines the direction which supervisory expectations take with respect to work performance. It is hypothesized that the consistency of performance expectations of supervisory judges is a function of the homogeneity of the groups from which these judges are drawn. Derived from this general hypothesis are two specific hypotheses to be tested:

Hypothesis #1 - The performance expectations of supervisors will be significantly more consistent than of non-supervisors.
Hypothesis #2 - The performance expectations of supervisors within homogeneous occupational groups will be significantly more consistent than among supervisors who are occupationally heterogeneous.
The second general question, considered in Phase II, concerns the congruence of worker-supervisor performance expectations. A positive relationship is postulated between the degree of congruence of the expectations of worker and supervisor and the evaluation by the supervisor of the individual worker's performance.

Hypothesis #3 - The ranking by a supervisor of worker performance is a function of the degree of congruence of worker-supervisor performance expectations.

**Instrument Employed**

A special instrument, the Performance Assessment Scale or PAS, was constructed to measure the effects under study. The scale is designed to minimize social desirability factors. The scale consists of fifteen groups of statements describing work behavior. The subject is asked to rank the five statements in such a way that they reflect order of importance for effective job performance. In this way, a measure of his performance expectations is obtained.

**Statistics**

For Phase I, the Coefficient of Concordance was utilized to measure consistency of expectations within groups of subjects. The values for this statistic were then transformed and placed in a Random Blocks Design.

For Phase II, a congruence score for each worker was obtained by squaring the deviations of his rankings of the PAS statements from the supervisor rankings. These scores
were then correlated to the ranking by the supervisor of the workers' performance by means of the statistic tau.

Results

Hypothesis #1 - 151 supervisors from widely disparate occupations were found to agree significantly more in their rankings of the PAS statements than 49 non-supervisors. An F of 8.51 was obtained. This is significant beyond the .05 level.

Hypothesis #2 - Of fifteen groups of occupationally homogeneous supervisors, eleven agreed in their rankings to a significantly greater extent than did the group of 151 supervisors of heterogeneous occupational background. Values of F for these eleven groups ranged from 11.33 to 252.13. Under the conditions of this study an F greater than 8.86 is significant beyond the .01 level.

Hypothesis #3 - The variable, congruence of expectations, was found to be related beyond the .05 level to supervisor rankings of workers in two out of the four groups tested.

It is interesting to note that none of the individual groups used to test the above hypotheses showed results which were contrary to prediction.

Conclusions

1. Membership group appears to be significantly related to the degree of consistency among judges in their evaluation of work performance.

2. The evaluation by a supervisor of the overall performance
of an individual worker is positively related to the degree of agreement between their performance expectations.

3. It would seem that cognizance of performance expectations is necessary in the understanding and refinement of criteria of worker performance.

4. Performance expectations may represent a promising new area of research upon the communication process between worker and supervisor.
James Taylor Mahoney, Jr.
30 Walnut Knoll,
Canton, Mass.
CA-6-2879

Age: 33
Married with one child, age 6

Born: New York, N.Y.

Parents Names: James T. and Eleanor Mahoney of Arlington, Mass.

Degrees: A.B. - Harvard College, 6/50
A.M. - Boston University, 6/57

World War II Veteran - U.S. Navy 1944 to 1946

Associate Member: Sigma Xi

Two years experience as a Personnel Director

Two professional publications

Currently employed at Veterans Hospital, Brockton, Mass.