



eCOMMONS

Loyola University Chicago
Loyola eCommons

Master's Theses

Theses and Dissertations

1965

A New Method of Scoring the How Supervise ? Test

Edward J. Hester
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses



Part of the [Psychology Commons](#)

Recommended Citation

Hester, Edward J., "A New Method of Scoring the How Supervise ? Test" (1965). *Master's Theses*. 1928.
https://ecommons.luc.edu/luc_theses/1928

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.
Copyright © 1965 Edward J. Hester

A NEW METHOD OF SCORING THE HOW SUPERVISE? TEST

by

Edward J. Hester

**A Thesis Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts**

January

1965

ACKNOWLEDGMENTS

The author is especially grateful to Mr. Earl C. Palmer, Director of Rehabilitation and Personnel Services, at Goodwill Industries of Chicago and Cook County, Inc. for permitting this study to be done at that agency.

The author likewise expresses his appreciation to Dr. Edmund P. Marx for his advice and guidance.

The author would also like to express his appreciation to Miss Sheila Pazio and Mr. John White for their able assistance.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF TABLES	iv
Chapter	
I INTRODUCTION	1
II REVIEW OF THE LITERATURE	3
III DEVELOPMENT OF THE SCALES	15
IV CROSS - VALIDATION OF THE SCALES	25
V SUMMARY	36
BIBLIOGRAPHY.	37
APPENDIX I	39
APPENDIX II	42
APPENDIX III	43
APPENDIX IV	45

TABLES

Table	Page
1. Age and Years of Education of the Subjects	15
2. Correlations for the Two Scales of the Leadership Opinion Questionnaire and Total and Parts of How Supervise? Test Form A	17
3. Intercorrelations of the Total and Three Parts of the How Supervise? Test Form Av.	18
4. Biserial Correlation of Each Item of the How Supervise? Test with Consideration (C) and Structure (S);	19
5. Special Scales Constructed for the How Supervise? Test to Predict Consideration and Structure	20
6. Correlation of the Special Scales of the How Supervise? with Consideration and Structure	21
7. The Standard Scores of Consideration and Structure When Estimated from the How Supervise?	22
8. Reliability, Means, and Standard Deviations of the How Supervise? Test	24
9. Reliability, Means, and Standard Deviations of the Leadership Opinion Questionnaire Scales . .	24
10. The Original and Cross-Validation Groups Compared for age and years of Education	25
11. Comparison of the Original and Cross-Validation Groups on the Basis of Sex	26
12. Comparison of the Original and Cross-Validation Groups on the Basis of Race	26
13. Correlation of the Special Scales of the How Supervise? with Consideration and Structure . . .	28
14. Correlations for the Two Scales of the Leadership Opinion Questionnaire and the Total and Three Parts of the How Supervise? Form A. . . .6.	29

Table	Page
15. The Standard Errors of Consideration When Estimated from the How Supervise?	29
16. Reliability, Means, and Standard Deviations of the How Supervise? Test	30
17. Reliability, Means, and Standard Deviations of the Leadership Opinion Questionnaire Scales . .	31
18. Intercorrelations of the Total and Three Parts of the How Supervise? Test Form A.	31
19. Comparison of the Original and Cross - Validation Groups on the Leadership Opinion Questionnaire.	33
20. Correlation of the Leadership Opinion Questionnaire and How Supervise? with age.	34
21. Comparison of the Leadership Opinion Questionnaire and How Supervise? Scores for Caucasians and Negroes	34

CHAPTER I

INTRODUCTION

The study of leadership has been one of the primary concerns in the fields of social and industrial psychology.

Apparently, one of the most rewarding such studies was that done at Ohio State University. In this analysis of leadership behavior two primary factors were discovered, entitled Consideration and Structure. These two dimensions of leadership behavior are defined by Fleishman (1960) as follows:

Consideration - Reflects the extent to which an individual is likely to have job relationships characterized by mutual trust, respect for subordinates' ideas, consideration of their feelings, and a certain warmth between supervisor and subordinates. A high score is indicative of a climate of good rapport and a two way communication. A low score indicates the supervisor is likely to be more impersonal in his relations with group members.

Structure - Reflects the extent to which an individual is likely to define and structure his own role and those of his subordinates towards goal attainment. A high score on this dimension characterizes individuals who play a more active role in directing group activities through planning, communicating information, scheduling, criticizing, trying out new ideas, etc.

The How Supervise? test of File (1945) is one of the most widely used tests of leadership opinion. It is composed of three parts entitled: Supervisory Practices, Company Policies, and Supervisor Opinions. It is possible that the How Supervise test is also measuring the two dimensions of leadership discovered by the Ohio State Leadership Studies. If these two factors are really universal dimensions of

leadership then they should be embedded in the How Supervise? test and might be predictable on the basis of the responses a subject makes to the test items.

This is the purpose of the current study, to measure the effect of Consideration and Structure upon the total score and the parts of the How Supervise? test Form A and, if possible, to construct special scales for the How Supervise? to measure Consideration and Structure.

CHAPTER II

REVIEW OF THE LITERATURE

When File (1945) developed the How Supervise? test, he made certain assumptions, the most important of which are as follows: (1) the ability to supervise is a general trait rather than specific to any particular job or company; (2) the lack of ability to deal with workers is the greatest single reason for supervisory failures and of management-worker friction; (3) the knowledge of supervisory functions can be measured by the responses to certain significant questions drawn from problems which frequently confront the supervisor.

File attempted to draw up questions which would be presented in problem form calling for operational responses. He also endeavored to select only questions which had face as well as statistical validity and only ones which are pertinent to industrial supervisors regardless of their department or company.

In testing the discriminating power of each item, it was felt that good supervisors as a group know the correct answers as well as do those who write books and articles on industrial supervision and men actually engaged in directing supervisory training programs, as a group. The statistical method employed was the critical ratio of the difference between the average

responses of the upper 20% and the lower 20% with respect to the total score on the test. It was found that the industrial experts as a group gave reliable answers to the problems presented in the test items. Two different groups of experts closely agreed on the answers ($r=.91$).

Apparently there have been only two studies published which are concerned with the development of a new scoring key for the How Supervise? test. One of these keys was developed by McCormick and Middaugh (1956) in order to increase the predictive efficiency of the test for supervisory personnel in a certain company. The study consisted of the analysis of item responses given in 1947 in relation to supervisory performances as rated in 1953. Ratings were made on three different rating factors: overall job performance, human relations ability; and ability to get the work done. However the inter-correlations between these ratings were of such a nature that only the first one, overall performance was used as a criterion. The split-half reliability of the How Supervise? used in this study was .21. When using the above merit rating as a criterion, the biserial correlation with the new scale was .27, while with the regular scoring system it was .05. The authors point out that the fact that the special scale differentiated while

the regular key did not probably is suggestive of the impact of differences of management philosophies.

The other study, an item analysis of the How Supervise? test using both internal and external criteria was done by Decker (1956). Over 200 college graduates who were members of the supervisory staff of a large manufacturing organization took form M of the How Supervise? test and were rated for supervisory performance. The results indicated no relation between the scores on the How Supervise? test and rated success in a supervisory position. An item analysis indicated that the items consistently measured some quality, possibly supervisory knowledge. Test records for the subjects were rescored on the basis of the 25 items which had significant coefficients of validity. The correlation between the total number right on these items and the criterion was found to be .35.

The greatest amount of research with the How Supervise? test seems to have been in the following areas:

1. The validity of the How Supervise? test in predicting supervisory success.
2. Whether or not the How Supervise? test is actually a measure of verbal intelligence or reading ability rather than of supervisory knowledge.

3. The improvement of scores on the How Supervise? test after various training courses.

The results of the validity studies of the How Supervise? test have been quite diversified. Carter (1952) used 48 foremen and assistant foremen in two metal fabricating plants as subjects. He found that part III of form A correlated .63 with ratings of supervisory ability made by fellow supervisors.

Holmes (1950) attempted to use the How Supervise? test to predict appraisals of job performance by management of two groups (A and B) of 100 and 50 supervisors of the office force of the State Farm Insurance Company. For group A the How Supervise? form A correlated .37 with the criterion. While with group B the How Supervise? form M correlated only .11 with the criterion. In defense of the varied results in this study, the author notes that the job of group A was only that of supervision while group B had specific duties to perform in addition to general supervision.

On the other hand, Decker (1957) Weitz and Nuckols (1953) found no significant relationship between the scores on the How Supervise? test and measures of supervisory success for a group of 55 supervisors in a medium-sized manufacturing plant and 78 district managers in a life insurance company. respectively. In the above mentioned article, Decker states that although the How Supervise? test may have some uses in

industry for such things as determining training needs among supervisors, it probably has little future as a selection device.

Sartain (1946) gave the How Supervise? test Experimental Edition, form A, to 40 members of supervision in the Texas Division of North American Aviation, Inc. Each man was rated by his superiors on two different rating forms, and the combination of the four ratings constituted the criterion of success. The How Supervise? was found to correlate $-.18$ with the criterion.

The fact that, in some studies, it was found that the How Supervise? test was not an adequate predictor of supervisory success raised the idea that it was actually not a test of supervisory ability, but rather of intelligence or reading ability.

Wickert (1952 b) investigated this claim using about 100 candidates for shop supervisory positions in a metal products company. He compared scores on the How Supervise? test with verbal intelligence and amount of education. The results seem to indicate that for those persons who have not graduated from high school, it measures intelligence rather than the knowledge of the principles of supervision. However, for relatively well-educated persons, the How Supervise? test scores have little relationship with intelligence test scores.

Similar results were obtained by Millard (1952). Using the Adaptability test as a measure of intelligence, he found a considerable correlation between intelligence and the How Supervise? test for factory supervisors and supervisors of newspaper carriers and dealers. For office supervisors, this correlation was considerable smaller and of less certain significance, while Holmes (1950) found correlations of .23 and .28 with another intelligence test (the Wonderlic Personnel Test).

Contradictory results were found by Sartain (1946) in the previously mentioned study of 40 supervisors in an aircraft plant. He found, in this study that the How Supervise? test was correlated $-.44$ with a measure of intelligence (the Tiffin and Lawshi Adaptability Test, form A).

Files (1945), Miller and Remmers (1950), and Millard (1952) using the years of education as an indirect measure of intelligence, found lower positive correlations with the How Supervise? test.

This same problem was studied from another aspect by Maloney (1952). He measured the readability of the How Supervise? test by the Flesh Formula. Accordingly, he found that the readability of the directions and items for the How Supervise? test is at the difficult level, with much of the material being at the high school graduate level of readability.

However, it should be stated that the Flesch Formula was not developed to be used on this type of material and, therefore, its validity in this instance might well be questioned.

In an attempt to learn whether or not a supervisory training course would have an appreciable effect on management's attitude toward supervisory practices, Miller and Remmers (1950) administered the How Supervise? test to a sample of 150 higher level managers. They found that it did not have any appreciable effect ($r .16$).

On the other hand, Wickert (1952 a) tested 227 college students with the How Supervise? test before and after taking various courses in psychology. The results showed that gains in mean scores for the four groups corresponded closely to the amount of human relations training believed to be included in each of the four courses. He also found that although the two forms of the test were roughly equivalent before training, form A following form B was considerably less sensitive in detecting the effects of human relations training than was form B following form A.

It might be that the discrepancy between the results of the above two studies may be attributable to the fact that the subjects used in the first study were probably much more firm in their attitudes than were the group used in the second study. Likewise, Mosel and Tracnaris (1959) in evaluating the results

of a six-week training program which produced small but significant improvement in supervisory attitudes as measured by the How Supervise? test, state that evaluation of training must involve the evaluation of the many other situational aspects in which training takes place, including organizational climate, attitudes and behavior of higher management.

LEADERSHIP OPINION QUESTIONNAIRE

Hemphill (1950), as part of the Ohio State Leadership Studies, developed the Leader Behavior Description Questionnaire. The questionnaire was composed of 150 items describing leadership behavior. The items were a priorily classified into nine groups: (1) integration, (2) communication, (3) production emphasis, (4) representation, (5) fraternization, (6) organization, (7) evaluation, (8) initiation, (9) domination. Studies showed that most of these scales were intercorrelated between .50 and .80.

Fleishman (1953 b) administered the Leader Behavior Description Questionnaire to 300 Air Force crew members who described their airplane commanders. A factor analysis of the results revealed that there were two major factors, termed "Consideration" and "Initiating Structure", and two minor factors which Fleishman tentatively labeled "Production Emphasis" and "Social Sensitivity".

On the basis of the above information, another questionnaire,

Supervisor Behavior Description Questionnaire, was developed. It was then applied to two groups of foremen in one of International Harvester Company's plants. After some revisions, the two dimensions of Consideration and Structure were found to be relatively independent of each other. The reliabilities of the two factors were .92 and .68, respectively, when applied to the second of the two groups at International Harvester. The intercorrelation of the two scales was $-.02$.

In the Supervisor Behavior Description Questionnaire the items were composed in such a way as to make it possible for an employee to rate his supervisor's behavior. The Leadership Opinion Questionnaire is essentially the same test but with the items reworded so that the subject expresses his own ideas of ideal supervisor behavior.

Bass (1956), in a validity study of the Leadership Opinion Questionnaire, found a correlation of .29 between the extent to which a supervisor believed he ought to be considerate of his employees and the extent to which he was rated a successful supervisor by his superiors two years later. No consistent relationship was found between favoring initiating structure and rated success.

In a second study, Bass (1958) administered this questionnaire to a group of 42 sales supervisors. The scores for the two scales were compared three years later with ratings

made by top management for these supervisors. Again, it was found that Consideration was significantly related to these ratings ($r .32$) while Structure was not correlated with the criteria ($r .05$).

Several researchers have attempted to determine the effects of differential degrees of Consideration and Structure in a supervisor's behavior upon his employees.

Oaklander and Fleishman (1963) studied nurses and non-medical supervisors in three hospitals. They found, as would be expected, that those supervisors who scored high on Consideration tended to have less intra-unit stress than those who scored low on this scale ($r -.31$). On the other hand, high Structure was related to low interdepartmental stress in voluntary hospitals ($r -.36$) but not in government hospitals. In the latter it was unrelated.

Fleishman and Harris (1962) employed the Supervisory Behavior Description Questionnaire to study the effect of Consideration and Structure in regards to employee grievances and turnover. Among the 57 production foremen in this plant it was found that Consideration and Structure were not pure factors but had an intercorrelation of $-.33$. Grievances were defined as the number presented in writing and placed in the company files. Turnover was measured by the number of workers who voluntarily left the company within an 11 month period. It

was found that there were significant relationships between the leadership behavior of foremen and the two measures of worker dissatisfaction. Grievances and turnover were lowest for groups headed by those foremen who were medium to high in Consideration together with low Structure. However, the most important relationship was with Consideration. High Consideration foremen had relatively low grievances and turnover regardless of the amount of structure in which he engaged. In other words, high Consideration foremen could increase Structure with very little increase in grievances and no increase in turnover.

In another study by Fleishman (1953 c), it was found that the higher people were in the plant hierarchy, the less consideration they felt the workers should get and the more structuring they felt should be initiated. However, in a subsequent study by Fleishman and Peters (1962) using another group of plant managers, he found contrary results. That is, the higher managerial levels tend to feel that they should initiate less structure in their relationship with their subordinates.

It seems that there is considerable permanence to the traits of Consideration and Structure. Fleishman (1953 a) administered the Leadership Opinion Questionnaire immediately before and after a leadership training course. He found that there was an average increase in Consideration scores during the course, while the Structure attitudes showed a general decrease for the foremen. The training, however, did not pro-

duce any kind of permanent change in either the attitudes or behavior of the trained foremen. Evaluation of the training back in the actual work situation showed that there were trends in the direction of more structuring and less consideration in those foremen who returned to the industrial environment.

Of course, in measuring the permanence of attitudes and traits by their susceptibility to change in a training program, there exists the major variable of the training program itself. Another approach would be to examine their relationship to the personality of the subject. This approach was taken by Fleishman and Peters (1962) and Hester and Daly (1964).

Fleishman and Peters administered the Leadership Opinion Questionnaire and the Survey of Interpersonal Values to 35 group department managers. The only relationship which he found significant beyond the .05 level was the inverse relationship of Structure and Independence ($r = -.39$).

Likewise, Structure was the only scales of the Leadership Opinion Questionnaire found by Hester and Daly to be related to some scales of the Minnesota Multiphasic Personality Inventory in a study of 50 supervisors and applicants for supervisory positions. Structure was determined to be negatively related at the .01 level with the following scales: Hypochondriasis (Hs), Hysteria (Hy), and Welch's Second Factor (R); while it was negatively correlated at the .05 level of confidence with the two scales: K-Correction (K) and Need for Affection (Hy_2).

CHAPTER III

DEVELOPMENT OF THE SCALES

SUBJECTS

The subjects in the development of this scale were 40 supervisors and applicants for supervisory positions at Goodwill Industries of Chicago and Cook County, Incorporated. The subjects, 32 males and 8 females, were divided into 22 Caucasians and 17 Negroes. Data on age and educational level is found in Table I.

TABLE I

AGE AND YEARS OF EDUCATION OF THE SUBJECTS

	Mean	σ
Age	31.77	10.11
Years of Education	12.30	2.68

PROCEDURE

The How Supervise? test Form A and the Leadership Opinion Questionnaire were administered together to the 40 subjects.

The total score and the scores for the three parts of the How Supervise were correlated with the two scales of the Leadership Opinion Questionnaire by means of the formula given by Lindquist (1942).

$$r_{xy} = \frac{\frac{\sum x y}{N} - M_x M_y}{\sigma_x \sigma_y}$$

Then the scores for the total and parts of the How Supervise? were intercorrelated, as well as the two scales of the Leadership Opinion Questionnaire, by means of the above formula.

In order to develop scales for the How Supervise? test (Form A) which will predict the scores on the two scales, Consideration and Structure, each item of the How Supervise? test was biserially correlated to the total scores on Consideration and Structure. The formula used is that given by McNemar (1962).

$$r_b = \frac{(M_z - M_y) p_2}{z \sigma_y}$$

In the above cases, the levels of significance were taken from Table 7 of the statistics manual by Crow, Davis, and Maxfield (1960).

RESULTS

The intercorrelations between the two scales of the Leadership Opinion Questionnaire (Consideration and Structure) and the three parts of the How Supervise? test Form A (I. Supervisory Practices, II. Company Practices, and III. Supervisory Opinion) are in Tables 2 and 3.

TABLE 2

CORRELATIONS FOR THE TWO SCALES OF THE LEADERSHIP
OPINION QUESTIONNAIRE AND TOTAL AND PARTS OF
HOW SUPERVISE? TEST FORM A

How Supervise?	Leadership Opinion Questionnaire	
	Consideration	Structure
Total	.461*	-.417 *
Part I	.326*	-.127
Part II	.246	.120
Part III	.323**	.050

* Significant at less than .01 level of confidence.

** Significant at the .05 level of confidence

TABLE 3
INTERCORRELATIONS OF THE TOTAL AND THREE PARTS OF
THE HOW SUPERVISE? TEST FORM A

	Total	Part I	Part II	Part III
Total502*	.666*	.783*
Part I	.502*148	.270
Part II	.666*	.148244
Part III	.783*	.270	.244

* Significant at less than .01 level of confidence.

From the above tables it seems that, at least for this group, the subscales of the How Supervise? test Form A are relatively independent factors. That is, the intercorrelations between the three parts are not significantly correlated with one another. The same thing applies to the two scales of the Leadership Opinion Questionnaire. The correlation between the Consideration and Structure scales is $-.115$; which is not significant at the .05 level of confidence.

However, the total score of the How Supervise? test is significantly correlated with both scales of the Leadership Opinion Questionnaire, i.e., $.461$ with Consideration and $-.417$ with Structure. Part I of the How Supervise test is significantly correlated with Consideration but not with Structure.

TABLE 4

BISERIAL CORRELATION OF EACH ITEM OF THE HOW SUPERVISE? TEST
WITH CONSIDERATION(C) AND STRUCTURE(S)

Item Number	C	S	Item Number	C	S
1	.344**	-.216	36	.016	.230
2	.340**	.093	37	.287	.192
3	.294	-.296	38	-.054	-.108
4	.122	-.087	39	.039	-.058
5	.000	.183	40	.330**	.017
6	-.048	.096	41	.123	-.098
7	.215	.063	42	.587*	-.372**
8	-.067	-.134	43	-.162	-.216
9	.206	-.276	44	.162	.162
10	-.090	.820*	45	.186	-.337**
11	.183	.275	46	.424*	-.135
12	.000	.000	47	.431*	.210
13	.784*	-.524*	48	.167	.560*
14	.267	-.134	49	-.116	.464*
15	.688*	-.129	50	.200	.043
16	.523*	.000	51	.252	-.115
17	.784*	-.262	52	.441*	-.148
18	-.154	.159	53	.405*	-.419*
19	.216	.196	54	.365**	-.092
20	.343**	.192	55	-.309**	.000
21	.708*	.075	56	.000	-.239
22	-.076	.083	57	.016	-.079
23	.365**	.000	58	-.406*	.243
24	-.186	-.186	59	.373**	-.094
25	.344**	.246	60	.256	.084
26	.054	-.216	61	.520*	-.109
27	.275	.092	62	.376**	.000
28	.000	.000	63	.034	.413*
29	.267	.402*	64	.149	.030
30	-.119	.239	65	.269	-.108
31	-.144	.576*	66	-.048	-.192
32	.132	-.370**	67	-.195	.130
33	.079	.248	68	.237	-.026
34	.347**	.097	69	.335**	.240
35	.000	.957*	70	.477*	-.239

* Significant at .01.

** Significant at .05.

Part II is not significantly correlated with either scale of the Leadership Opinion Questionnaire and Part III is significantly correlated with Consideration (.323) but not with Structure.

The biserial correlations of each item of the How Supervise? test form A with Consideration and Structure are found in Table 4.

On the basis of this data two scales were created to predict Consideration and two scales to predict Structure. Table 5 indicates the four scales, the code letters for them, what they are proported to predict, and the range of correlation that the scale items have with the criteria.

TABLE 5

Special Scales Constructed for the How Supervise?
Test To Predict Consideration and Structure.

Scale Code	Number of Items	Criteria	Range of Item Correlations with Criteria
C ₂₅	25	Consideration	> .300
C ₃₃	33	Consideration	> .250
S ₁₂	12	Structure	> .300
S ₁₆	16	Structure	> .250

Two scales were made to predict each criteria in order to determine if the longer scales, in spite of the fact that they contain items which do not correlate with the criteria at the .05 level of confidence, will be more predictive in view of greater reliability which they may have. The actual scales are given in Appendices I, II, III, and IV.

Of the 70 items in the How Supervise? test Form A, 43 items are used in the four scales developed here. Six of these items appear in both the Consideration scales and the Structure scales. Only one of these six items (#29) is positively correlated with both Consideration and Structure. The other five items are positively correlated with Consideration and negatively correlated with Structure.

When these scales were applied to this original group the correlations with Consideration and Structure were as follows in Table 6.

TABLE 6

CORRELATION OF THE SPECIAL SCALES OF THE HOW SUPERVISE?
WITH CONSIDERATION AND STRUCTURE

How Supervise? Scale	Consideration	Structure
C ₂₅	.695*	-.001
C ₃₃	.658*	-.003
S ₁₂	-.225	.582*
S ₁₆	-.264	.589*

* Significant at less than the .01 level of confidence.

According to the correlations as stated above, it makes little difference which scale, shorter or longer, is employed since the correlations are almost the same for the two Consideration scales, a difference of .037, and for the two Structure scales, a difference of .007.

The standard errors of the Leadership Opinion Questionnaire Scales, when estimated from the total How Supervise? test scores and the special scales of the latter test, are given in Table 7.

TABLE 7

THE STANDARD ERRORS OF CONSIDERATION AND STRUCTURE
WHEN ESTIMATED FROM THE HOW SUPERVISE?

How Supervise?	Consideration	Structure
Total	6.79	7.54
C ₂₅	5.51
C ₃₃	5.77
S ₁₂	6.20
S ₁₆	6.17

The best predictor of Consideration, according to the above table, would be scale C₂₅ with a standard error of 5.51. This means that a particular score for Consideration, as estimated by scale C₂₅, will be within 10.8 points on either side of that estimated score 95 percent of the time. Thus, using

the norms given by Fleishman (1960) for general supervisory personnel, if a subject's estimated score were 54, which is at the 50th percentile, it could be stated with reasonable confidence that the true score would not be more than 65, which is at the 91st percentile, nor less than 43, which is at the 7th percentile.

Since the standard error of Structure as estimated from scale S_{16} is larger than for the example above, the range will be even larger. If a person's estimated score is 48, which is at the 50th percentile in the same normative group, it could be stated with 95 percent accuracy that his true score is not greater than 60, which would be at the 95th percentile, nor less than 36, which is at the 3rd percentile. Therefore, it does not seem that these scales are accurate enough for prediction in individual cases.

The split-half reliabilities of the total score and special scales of the How Supervise? test Form A are listed in Table 8.

The same information is given for the Leadership Opinion Questionnaire's two scales, Consideration and Structure, in Table 9.

TABLE 8

RELIABILITY, MEANS, AND STANDARD DEVIATIONS OF THE
HOW SUPERVISE TEST?

How Supervise? Test Scales	Reliability	S-B Corrected Reliability	Mean	σ
Total	.282	.440	37.72	10.89
C ₂₅	.805	.892	17.00	3.65
C ₃₃	.653	.790	23.08	4.65
S ₁₂	.503	.669	6.95	2.10
S ₁₆	.527	.690	8.70	2.35
Part I	.139	.244	12.95	2.60
Part II	.534	.696	12.58	4.91
Part III	.774	.873	12.25	6.87

TABLE 9

RELIABILITY, MEANS, AND STANDARD DEVIATIONS OF THE
LEADERSHIP OPINION QUESTIONNAIRE SCALES

Scale	Reliability	S-B Corrected Reliability	Mean	σ
Consideration	.206	.342	52.38	7.66
Structure	.515	.680	56.10	7.63

CHAPTER IV

CROSS-VALIDATION OF THE SCALES

SUBJECTS

The 30 subjects for the cross-validation were also from Goodwill Industries of Chicago and Cook County, Incorporated. However, unlike the original group, they consisted entirely of applicants for supervisory positions.

It was at first believed that they would be comparable to the original group. However, when they were compared in terms of age, race, years of education, and sex, it was found that they were significantly different, by means of the t-test, at the .05 level of confidence, in regards to the first two characteristics. The comparison for age and years of education is given in Table 10, while that for sex is given in Table 11, and for race in Table 12.

TABLE 10

THE ORIGINAL AND CROSS-VALIDATION GROUPS COMPARED FOR
AGE AND YEARS OF EDUCATION

	Original		Cross-validation	
	Mean	σ	Mean	σ
Age	31.77	10.11	36.63	10.40**
Years of Education	12.30	2.68	12.86	1.79

** Significant at less than the .05 level of confidence.

TABLE 11

COMPARISON OF THE ORIGINAL AND CROSS-VALIDATION GROUPS
ON THE BASIS OF SEX

	Original	Cross-validation
Male	32	27
Female	8	3

N.B. Difference is not significant.

TABLE 12

COMPARISON OF THE ORIGINAL AND CROSS-VALIDATION GROUPS
ON THE BASIS OF RACE

	Original	Cross-validation
Caucasion	22	24
Negro	17	6

N.B. Difference is significant at the .05 level of confidence.

From the above tables it can be seen that the cross-validation group is significantly older and is composed of a significantly greater number of Caucasians than the original group.

PROCEDURE

After the How Supervise? test Form A and the Leadership Opinion Questionnaire had been administered to the 30 applicants for supervisory positions, each of the tests was scored according to the traditional method as well as with the special scales for the How Supervise? test developed in this study.

As in the preceding chapter, the total score, the scores for the three parts of the How Supervise, and the four special scales were correlated with the two scales of the Leadership Opinion Questionnaire. The correlations were also corrected for attenuation by means of the formula given by McNemar (1962).

$$r_{tt} = \frac{r_{xy}}{\sqrt{r_{xx}} \sqrt{r_{yy}}}$$

RESULTS

When the special scales for the How Supervise? test Form A were applied to the cross-validation group, the correlations with the Consideration and Structure scales of the Leadership Opinion Questionnaire were as follows in Table 13.

TABLE 13
CORRELATION OF THE SPECIAL SCALES OF THE HOW SUPERVISE?
WITH CONSIDERATION AND STRUCTURE

How Supervise? Scale	Consideration	Structure
C ₂₅	.300	.127
C ₃₃	.374**	.179
S ₁₂	-.059	-.089
S ₁₆	-.001	-.032

** Significant at less than the .05 level.

As it can be seen from the above table, only one of the scales, i.e., C₃₃ is significantly related to Consideration and none are related to Structure. In regard to the prediction of Consideration, the total score was much better in this than the scale C₃₃ as is evident from the information given in Table 14 and Table 15. Although the total score and the three parts of the How Supervise? test correlated with Consideration, none of them was significantly related to Structure.

TABLE 14

CORRELATIONS FOR THE TWO SCALES OF THE LEADERSHIP OPINION
QUESTIONNAIRE AND THE TOTAL AND THREE PARTS OF
THE HOW SUPERVISE? FORM A

How Supervise?	Leadership Opinion Questionnaire Consideration	Questionnaire Structure
Total	.447**	.139
Part I	.355**	.251
Part II	.371**	.085
Part III	.370**	.065

** Significant at less than the .05 level.

The standard errors of the Consideration scale of the Leadership Opinion Questionnaire, when predicted from the total score of the How Supervise? and the special scale C₃₃ are given in Table 14. The other standard errors are not presented, as was done with the Normative Group, because the scales in these cases were not significantly related for the cross-validation group.

TABLE 15

THE STANDARD ERRORS OF CONSIDERATION WHEN ESTIMATED
FROM THE HOW SUPERVISE?

How Supervise?	Consideration
C ₃₃	8.03
Total	7.74

The split-half reliability of the total score and special scales of the How Supervise? for the cross validation group is listed in Table 16. The reliabilities, as in the preceeding group, were computed by the odd-even method. The Spearman - Brown corrected reliabilities are also given as well as means and standard deviations.


TABLE 16
RELIABILITY, MEANS, AND STANDARD DEVIATIONS OF THE HOW
SUPERVISE? TEST

How Supervise? Test Scale	Reliability	S-B Corrected Reliability	Means	SD
Total	.655	.792	36.16	12.54
C ₂₅	.602	.752	16.86	3.44
C ₃₃	.457	.627	22.86	3.76
S ₁₂	.090	.165	6.90	1.65
S ₁₆	.244	.392	8.66	1.86
Part I	.174	.296	10.80	3.56
Part II	.644	.783	12.60	4.88
Part III	.451	.622	12.76	6.87

The same information is given for the Leadership Opinion Questionnaire's two scales, Consideration and Structure, in Table 17.

TABLE 17

RELIABILITY, MEANS, AND STANDARD DEVIATIONS OF THE
LEADERSHIP OPINION QUESTIONNAIRE SCALES

Scale	Reliability	S-B Corrected Reliability	Mean	
Consideration	.051	-.097	50.86	8.66
Structure	.479	.648	57.16	16.86

The intercorrelations reported below in table 18 indicate that, contrary to what was found with the original group, the parts of the How Supervise? for the cross-validation group are highly intercorrelated.

TABLE 18

INTERCORRELATIONS OF THE TOTAL AND THREE PARTS OF THE
HOW SUPERVISE? TEST FORM A

	Total	Part I	Part II	Part III
Total730*	.807*	.873*
Part I	.730*502*	.457*
Part II	.807*	.502*505*
Part III	.873*	.457*	.505*

* Significant at less than .01 level.

This significant intercorrelation suggests that a central factor is in effect here where it did not apply to the first group. On the other hand, the two scales of the Leadership Opinion Questionnaire are not significantly correlated (.182), so that at first glance they appear to remain independent factors in this group as they were in the former. However, in Table 16 it is seen that Consideration, for this group is completely unreliable. When this is taken into account and the correlation is corrected for attenuation, i.e., the unreliability of the scales, the correlation becomes .728 which is significant at less than the .01 level of confidence. In other words, if the two scales could be made perfectly reliable, then the theoretical correlation between the two would be .728. Therefore, the central factor is apparently affecting the scores on both the How Supervise? test and the Leadership Opinion Questionnaire for this group.

It will be recalled that Consideration and Structure for the Normative group had an intercorrelation which was not significant and thus they were considered to be independent factors. Even when this intercorrelation (-.115) was corrected for attenuation, it did not become significant (-.238).

In an attempt to further explain why the special scales of the How Supervise? test did not have predictive values in the cross-validation study, the means of the two groups were

compared for the two scales of the Leadership Opinion Questionnaire, the total score of the How Supervise? and the four special scales. The results given in Table 19 indicate that there are no significant differences according to the t-test.

TABLE 19

COMPARISON OF THE ORIGINAL AND CROSS-VALIDATION GROUPS ON THE LEADERSHIP OPINION QUESTIONNAIRE AND HOW SUPERVISE?

	Original		Cross Validation	
	Mean	σ	Mean	σ
Consideration	52.37	7.66	50.86	6.51
Structure	56.07	7.63	57.16	9.48
How Supervise?	37.72	10.88	36.16	12.53
C ₃₃	23.07	4.69	22.86	3.76
C ₂₅	17.00	3.68	16.86	3.44
S ₁₆	8.70	2.35	8.66	1.86
S ₁₂	6.95	2.10	6.90	1.65

N.B. None of the differences are significant.

Since it was found, as reported in Tables 10 and 12 that there were significant differences between the two groups in terms of age and race, these two were examined for any relationship that might exist between them and the subjects' scores on the Leadership Opinion Questionnaire and How Supervise?. It was found that neither age (see Table 20) nor race (see Table 21) was significantly related to the scores on the Leadership

LEWIS TOWERS
LOYOLA

Opinion Questionnaire. However, it was determined that Caucasians obtained significantly higher scores according to the t-test on the How Supervise? than did the Negroes.

TABLE 20

CORRELATION OF THE LEADERSHIP OPINION QUESTIONNAIRE AND
HOW SUPERVISE? WITH AGE

	Age
Consideration	.013
Structure	-.175
How Supervise?	-.091

N.B. None of the correlations are significant.

TABLE 21

COMPARISON OF THE LEADERSHIP OPINION QUESTIONNAIRE AND
HOW SUPERVISE? SCORES FOR CAUCASIANS AND NEGROES

	Caucasians		Negroes	
	Mean	σ	Mean	σ
Consideration	51.82	7.61	51.73	6.58
Structure	57.13	8.85	55.74	7.79
How Supervise	39.73	11.44	32.69**	9.74

** Difference is significant at less than the .05 level of confidence.

Thus, it seems that the total score of the How Supervise? test is moderately weighted with Consideration, with the effect of Structure still undetermined.

CHAPTER V

SUMMARY

The purpose of this thesis was to investigate the degree to which the How Supervise? test is affected by the two factors of Consideration and Structure as measured by the Leadership Opinion Questionnaire and to develop scales to predict these two factors.

The subjects for the scale development were 40 supervisors and applicants for supervisory positions at Goodwill Industries of Chicago and Cook County, Incorporated. The cross-validation group consisted entirely of applicants for supervisory positions. The second group was found to be significantly older and to be significantly different in terms of race.

It was determined that the How Supervise? is moderately weighted with Consideration but the effect of Structure is undetermined.

Special Scales were developed to predict Consideration but those for Structure were not predictive in the cross-validation.

The Caucasians in this study, it was learned, obtained significantly higher scores on the How Supervise? test than did Negroes. However, there were no significant racial differences on the Leadership Opinion Questionnaire.

BIBLIOGRAPHY

1. Bass, B.M. "Leadership Opinions as forecasts of supervisory success", J. Appl. Psychol. 1956, 40, 345-346
2. Bass, B.M. "Leadership Opinions as forecasts of supervisory success: a replication" Personnel Psychol. 1958, 11, 515-518.
3. Carter, G.C., "Measurement of supervisory ability", J. Appl. Psychol. 1952, 36, 393-395
4. Crow, E. L.; F.A. Davis & M.W. Maxfield, Statistic Manual New York: Dover Publications, Inc., 1960
5. Dicker, R.L. "An item analysis of 'How Supervise?' using both internal and external criteria", J. Appl. Psychol. 1956, 40, 406-411
6. Decker, R.L. "Across-validation study of a test of supervisory ability" Proc. W. Va. Acad. Sci. 1957, 20, 105-109.
7. File, Q. W. "The Measurement of supervisory quality in industry" J. Appl. Psychol. 1945, 29, 323-337
8. Fleishman, E.A., "Leadership climate, human relations training, and supervisory behavior", Personnel Psychol. 1953a, 6, 205-222.
9. Fleishman, E.A. "The description of supervisory behavior" J. Appl. Psychol. 1953b, 37, 1-6.
10. Fleishman, E.A. "The Measurement of leadership attitudes in industry" J. Appl. Psychol. 1953c, 37, 153-158.
11. Fleishman, E.A. Manual for the Leadership Opinion Questionnaire Chicago: Science Research Associates, Inc., 1960
12. Fleishman, E.A. & E.F. Harris "Patterns of leadership behavior related to employee grievances and turnover" Personnel Psychol. 1962, 15, 43-56.
13. Fleishman, E.A. & D.R. Peters, "Interpersonal values, leadership attitudes and managerial 'success'", Personnel Psychol. 1962, 15, 127-143
14. Hemphill, J.K. Leader Behavior Description, Personnel Research Board, Ohio State University, 1950

15. Hester, E.J. & R.M. Daly "Personality correlates of Consideration and Structure" Unpublished report 1964
16. Holmes, F.J. "Validity of tests for insurance office personnel", Personnel Psychol. 1950, 3, 57-69
17. Lindquist, E.F. A First Course in Statistics Cambridge, Mass: Riverside Press - 1942
18. Maloney, P.A. "Reading ease scores for Files 'How Supervise'" J. Appl. Psychol. 1952, 36, 225-227
19. McCormich, E.J. & R.W. Middaugh, "The development of a tailor-made scoring key for the 'How Supervise' Test" Personnel Psychol. 1956, 9, 27-37.
20. McNemar, Q. Psychological Statistics New York: John Wiley and Sons, Inc. 1962
21. Millard, K.A., "Is 'How Supervise?' an intelligence test?" J. Appl. Psychol. 1952, 36, 221-224
22. Miller, F.G. & H.H. Remmers, "Studies in industrial empathy: II Management's attitudes toward industrial supervision and their estimates of labor attitudes" Personnel Psychol. 1950, 3, 33-40
23. Mosel, J.N. & H.J. Trannaris, "Evaluating the supervisors programs" Engng. Industr. Psychol. 1959, 1, 18-23
24. Oaklander, H. & E.A. Fleishman "Patterns of leadership related to organizational stress in hospital settings" Unpublished report, 1963.
25. Sartain, A.Q. "Relation Between scores on certain standard tests and supervisory success in an aircraft factory" J. Appl. Psychol. 1946, 30, 328
26. Weitz, J. & R.C. Nuckols "A validation study of 'How Supervise?'" J. Appl. Psychol. 1953, 37, 7-8
27. Wickert, F.R., "How Supervise? scores before and after courses in Psychology" J. Appl. Psychol. 1952a, 36, 388-392
28. Wickert, F.R. "Relation between 'How Supervise?' intelligence and education for a group of supervisory candidates in industry" J. Appl. Psychol. 1952b, 36, 301-3

APPENDIX I

HOW SUPERVISE? ITEMS WHICH WERE FOUND TO BE RELATED AT BEYOND THE .05 LEVEL OF SIGNIFICANCE WITH THE CONSIDERATION SCALE OF THE LEADERSHIP OPINION QUESTIONNAIRE

Item No.	Related Response	Item
1.	Desirable	Asking your workers for suggestions before setting up an important project.
2.	Desirable	Transfer dissatisfied, but capable, workers to other jobs.
13.	Desirable	Giving a discharged worker a full explanation of your reasons for asking that he be fired.
15.	Undesirable	Urging employees to handle their own problems without seeking advice from anyone.
16.	Desirable	Telling poor workers when their work isn't measuring up to what it should be.
17.	Desirable	Dividing overtime as equally as possible among all workers.
20.	Desirable	Promoting employee recreation projects, such as athletic teams, hobby clubs, and social groups.
21.	Undesirable	Basing all promotions on how long the individual has worked for the company.
23.	Desirable	Making periodic surveys of the attitudes of employees toward company policies and management.
25.	Desirable	Asking employees to recommend individuals to be hired for new positions.

Item No.	Related Response	Item
34.	Desirable	Providing for special "exit interview" with all workers who have been fired.
40.	Undesirable	Giving supervisors longer vacations than those enjoyed by the average worker.
42.	Disagree	What the worker thinks is unimportant so long as he is doing his job well.
46.	Disagree	Sympathising with worker's difficulties only encourages unfounded protests against working conditions.
47.	Disagree	What the worker does during his "off hours" should be of no concern to his employer.
52.	Disagree	The usefulness of the product he is making is of little concern to the average employee.
53.	Disagree	The best way to make sure that rules will be obeyed is to put plenty of teeth in them.
54.	Disagree	Supervisors should be relieved of all responsibility for teaching new workers how to do their work.
55.	Disagree	Ability to handle workers is inborn, not learned.
58.	Agree or ?	No honest worker will go on strike against a company which provides its workers with a decent wage.
59.	Disagree	Supervisors are usually criticized more than they deserve.

ITEM NO	RELATED RESPONSE	ITEM
61.	Agree	If a worker goes over your head with a grievance, it is usually a sign of poor supervision on your part.
62.	Agree	A supervisor is a misfit unless he has the confidence and loyalty of his men.
69.	Disagree	The goals of management and labor are directly opposed and must always be in conflict with each other.
70.	Disagree	Rapid learners are usually quick forgetters.

APPENDIX II

HOW SUPERVISE? ITEMS WHICH WERE NOT FOUND TO BE SIGNIFICANTLY RELATED AT THE .05 LEVEL TO THE CONSIDERATION SCALE OF THE LEADERSHIP OPINION QUESTIONNAIRE BUT WERE CORRELATED BETWEEN .050 AND .299.

ITEM NO	RELATED RESPONSE	ITEM
3.	Undesirable	Impressing upon each worker that his job depends on how much work he turns out.
14.	Undesirable	Putting a loud individual in his place with a sarcastic remark.
27.	Desirable	Asking workers to comment about the way the company treats them.
29.	Desirable	Holding a supervisor responsible for the quality of the product produced in his department.
37.	Desirable	Requiring department heads to spend at least one week of the year visiting other up-to-date plants.
51.	Disagree	The only important requirement of a good supervisor is a complete understanding of the jobs he is to supervise.
60.	Disagree	The average supervisor can do nothing to reduce absenteeism.
65.	Disagree	You can tell when a person is lying by noting whether he looks you straight in the eye or not.

APPENDIX III

HOW SUPERVISE? ITEMS WHICH WERE FOUND TO BE RELATED BEYOND THE .05 LEVEL OF SIGNIFICANCE OR LESS WITH THE STRUCTURE SCALE OF THE LEADERSHIP OPINION QUESTIONNAIRE

ITEM NO	RELATED RESPONSE	ITEM
10.	Desirable	Talking over ways of cutting costs with your workers.
13.	Undesirable or "?"	Giving a discharged worker a full explanation of your reasons for asking that he be fired.
29.	Desirable	Holding a supervisor responsible for the quality of the product produced in his department.
31.	Desirable	Arranging monthly cost reports so as to give recognition to the department having the best record.
32.	Undesirable or "?"	Giving workers who turn in valuable suggestions a part of the money saved by putting their ideas into effect.
35.	Desirable	Giving supervisors special training on how to handle dismissal cases.
42.	Agree or "?"	What the worker thinks is unimportant so long as he is doing his job well.
45.	Agree or "?"	Praising workers for good work only leads to demands for more pay.
48.	Agree	The way an individual is treated by his fellow workers will probably determine whether he likes his job or not.
49.	Disagree	The kind of job which the individual does has little effect upon his social position.
52.	Agree or "?"	The best way to make sure that rules will be obeyed is to put plenty of teeth in them.

ITEM	RELATED	ITEM
NO	RESPONSE	
63	Agree	Knowing a great deal about an individual's home life is a great help in selecting the right person for a responsible job.

APPENDIX IV

HOW SUPERVISE? ITEMS WHICH WERE NOT FOUND TO BE SIGNIFICANTLY RELATED AT THE .05 LEVEL WITH THE STRUCTURE SCALE OF THE LEADERSHIP OPINION QUESTIONNAIRE BUT WERE CORRELATED BETWEEN .250 AND .299.

ITEM NO	RELATED RESPONSE	ITEM
3.	Desirable or "?"	Impressing upon each worker that his job depends on how much work he turns out.
9.	Desirable or "?"	Making an example of one worker to prevent further trouble with others.
111.	Undesirable	Prohibiting conversation between workers on routine jobs.
17.	Undesirable or "?"	Dividing overtime as equally as possible among all workers.

APPROVAL SHEET

The thesis submitted by Edward J. Hester has been read and approved by three members of the Department of Psychology.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated, and that the thesis is now given final approval with reference to content, form, and mechanical accuracy.

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

February 12, 1965
Date

Edmund P. Marx
Signature of Adviser