# MANIFEST STRUCTURE ANALYSIS OF SUPERVISORY TESTING

Thesis for the Degree of M. A.

MICHIGAN STATE COLLEGE

Clayton H. Rashleigh

1954

This is to certify that the

thesis entitled

MANIFEST STRUCTURE ANALYSIS OF SUPERVISORY TESTING

presented by

Clayton H. Rashleigh

has been accepted towards fulfillment
of the requirements for

M.A. degree in Psychology

G. M. Gilbert

Major professor

Date May 28, 1954

# MANIFEST STRUCTURE ANALYSIS OF SUPERVISORY TESTING

by

Clayton H. Rashleigh

A Thesis

Submitted to the School of Graduate Studies of Michigan State College of Agriculture and Applied Science in partial fulfullment of the requirements for the degree of

MASTER OF ARTS

Department of Psychology

Year 1954

## ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Dr. Frank M. du Mas for generously making available his knowledge, energy, and pioneering techniques of analysis.

He is also greatly indebted to Dr. Carl F. Frost for much concrete help, unfailing interest, and encouragement.

Grateful acknowledgement is also due to Dr. G. M. Gilbert for patience and cooperation in many ways.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

# INTRODUCTION

The purpose of the present study is to investigate a new technique of item validation and scale construction recently formulated by du Mas (4). The need for valid tests in industry and business, as well as in the armed services, has been emphasized by Super (14) and Lawshe (9).

Testing of Personnel for selection, placement, or promotion in industry has become an increasingly important development in psychology in recent years. Super (14) calls testing big business, stating that one million Americans took sixty million tests in one year. Lawshe (9) comments that recent war years clearly demonstrated the effectiveness of personnel tests both in industry and in the military services.

Lawshe (9) sees a need for tailor-made tests in all areas, and remarks on a growing tendency toward selective scoring of commercially available standard tests for a specific situation. The important question, he asks, is whether or not the test helps to identify the persons who are apt to be most successful on this particular job. In this connection, he states: "Whether it is a matter of test construction or the selection of significant items in commercially available tests, the problem of item validation is one and the same." (9, p. 17).

The usefulness of valid testing instruments has been illustrated by many studies. Wadsworth (17) has shown that test-selected employees proved satisfactory more often than non-test selected employees and

produced a smaller percent of problem employees.

Strong (12) showed that 56% of insurance salesmen scoring A on an interest test had individual sales totals of $150,000 while only 6% of salesmen scoring C on the test achieved that figure.

In the Army Aviation Testing Program, four percent of cadets with stanine score 9 were eliminated from primary flying school while seventy-seven percent of the cadets with stanine score 1 were eliminated.

File and Remmers (5) found that of forty-six men selected as supervisors in a company 80% scored above average on the How Supervise Test, while of fifteen men by-passed because judged lacking in ability only 15% scored above average on that test.

Wonderlic (18) found that among representatives of a personal-finance company, 86% of those employed a year or more made above a critical score on the Personnel Test, while only 35% of those who were dismissed or left the company made above that score.

However, many tests do not achieve adequate validity. Also, many tests fail to stand up under cross-validation. That is, the results achieved with the first sample are not verified in a comparable but independent sample, using the same criterion. Super (14) feels that external evidence of validity is the only adequate basis for judging a test, that is, verification against a criterion.

Validity immediately suggests the criterion problem. A question that Jenkins (8) asks is: "Validity for what?" The answer must be in terms of a good criterion. Clear and simple definitions of a good criterion are not plentiful. A criterion, in this context, might be

described as a measureable or quantifiable standard of behavior in a given situation, or a measure of worker performance. A criterion may be simple, such as the number of pieces per hour by men on a certain type of machine, or it may be composite and multivariate. For example, a criterion might consist of weighted combination of output, quality of work, ratings, and possibly other factors. Thus, criteria may be objective as records of production, or subjective as ratings of adaptibility. To attempt a consensus of expressed requirements, a good criterion should be: reliable, relevant, related to other criteria, suitable to the job analysis, available, acceptable to management, modifiable in terms of changes in the situation, and quantifiable.

Rush (10) in an elaborate factor analysis, concluded that the criterion of sales success is multidimensional rather than unitary, and hence the use of global measures of success or failure would seem undesireable, since this might obscure underlying relationships in validation studies. He also concluded that the development of effective selection devices may be facilitated by a knowledge of the component elements of job criteria.

The findings of Taylor, Schneider, and Symonds (15) seem to disagree with the above conclusions. Their factor analysis of 13 graphic rating scales of salesmen yielded only one clear factor. They concluded that basic salary constituted management's considered judgement of the value of the man to the organization, expressed on a dollar continuum. Using basic salary as their criterion, they found a cross validation coefficient of .47 for their form of forced-choice tetrads and rating

scales. The validity generalized to another group of salesmen in a different division of the company, the correlation being .52.

Super (14, p.48) uses the terms standardization and validation interchangeably, "because the standardization of a vocational test implies collecting data which make possible validation." It seems apparent that most of the steps dealing with selection of tests and test construction have as their goal a test which is valid for its specific use.

Super (14) says that the minimum correlation coefficient, or validity coefficient, for psychological tests has been generally set at .45 for individual tests; but lower validity coefficients may be combined usefully in test batteries. Validity coefficients are not likely to exceed .70, according to Super, because of the unreliability of criteria. As an example, he cites the unreliability between supervisors' ratings, that is, lack of agreement between raters. His argument appears logically sound. But one might ask the question, what if the criterion were more reliable?

In a highly competitive industrial situation, where incompetence cannot be tolerated, might not the supervisor's salary represent the carefully considered judgement of his value to the company, or even a relatively accurate extimate of his demonstrated value? Would it not seem logical to consider supervisors a selected group who reach and maintain their position through special effort for special reward? Some evidence supports the assumption that motivation is generally higher in higher socio-economic levels as suggested by Barnett, Handelsman, Stewart and Super (1).

Testing of supervisors offers special problems. Lawshe (9) comments

that supervisory jobs vary tremendously. Gibb (6) states that there is no one-leadership type of personality. Cleeton and Mason (2) agree that there is no general executive type. Super (14) points out that although a great deal of time and money is being spent on the application of psychological methods to the selection of executive personnel, little has been published on it in the psychological journals. He lists five current types of work in executive selection and evaluation: the development of custom-built batteries of tests such as the Cleeton-Mason Vocational Aptitude Examination; the validation of standard tests for this particular purpose, as in the University of Minnesota's College of Business Administration project; the development of single tests for executive interests or other traits, best illustrated by Strong's (11, 13) work with executives and public administrators; the clinical use of interviews and tests as commonly done by consulting psychologists and the use of clinically evaluated situation tests as developed by the British War Officer Selection Boards and carried further by the U. S. Office of Strategic Services.

In the field of executive selection, Thompson (16) found positive results with a battery of standard tests administered to 15 superior and 10 average executives of a firm of consulting management engineers. The tests included the Wonderlic Personnel Test, Michigan Vocabulary Profile Test, Cardall Test of Practical Judgement, Kuder Preference Record, Adams-Leply Personal Audit, Beckman Revision of the Allport A-S Reaction Study, Guilford-Martin Personnel Inventory, and Rood I-E Test. The criterion consisted of performance records (not described) and ratings by

partners (reliability not stated). Differences, at or above the 5% level, were found with the Wonderlic, Michigan Vocabulary Profile, Kuder, and the Adams-Leply. All of the reported differences favored the superior executives, except that on the Kuder Social Service Scale. The results describe the successful management engineer executive as superior to less successful partners in mental ability, interests, firmness, and stability, and inferior in interest in social service. No cross-validation study was reported, therefore these results must be considered highly tentative, especially with such a small sample.

Harrell (7) reported on 42 overseers, in three different cotton mills, rated satisfactory or unsatisfactory by their superiors. With a critical I.Q. of 100, on the Otis Self-Administering Test of Mental Ability, only 70% of the unsatisfactory, but 100% of the satisfactory achieved this I.Q. In view of the discussion of criteria, above, this study appears open to criticism.

Lawshe (9) comments that there is little evidence of successful validity studies in the executive brackets. He attributes this to the difficulty of setting up criterion groups at this level, and partially also to failure to develop adequate measuring instruments. Cleeton and Mason (2) point out that, since successful executives generally score relatively high on a wide variety of ability tests, they would seem to be well rounded personalities. Lawshe (9) suggests mental ability tests, temperament tests, interest tests, and, specifically, the Michigan Vocabulary Profile Test as most promising in this area.

Two important problems in testing generally seem to be: 1. choice of

test, or construction of a new one, and 2. validation of the test in a specific situation. For the development of a new vocational test, Super (14) suggests seven major steps: job analysis, selection of traits to test, selection of criteria of success, item construction, standardization, validation, and cross-validation. He points out that one or more of these steps may be slighted or omitted in special circumstances.

With reference to test construction, Super (14) stresses the importance of selecting a criterion early. He indicates that the criterion should be considered as soon as the characteristic to be tested has been isolated and selected on the basis of job analysis. This should also indicate the choice of the type of test to be constructed. Then should follow the problems of constructing apparatus and drawing or writing items, the first trial of the tentative form, further revision, collection of data on a larger group of subjects, analysis of the internal consistency, analysis of the scoring key, and another revision of the test.

These problems need not be gone into in detail here. However, it should be apparent that conventional test construction is a very complex and time-consuming task. All the problems mentioned above occur before data is collected for standardization and the establishing of norms. The test then must be validated in a specific situation, and cross-validated on another group not included in this first validation group, but using the same external criterion in both groups. This points up the complexity of test construction with conventional methods, and specifically the

crucial function of item analysis, or validation of the items which make up the test.

In conventional methods, test items are selected on the basis of values generated from theory or inference from properties of the stimulus items, according to du Mas (4). The finished instrument often does not have sufficient validity for effective use, resulting in a great loss of time and effort. Research which is so expensive and time-consuming, and which may turn out a total loss, is hard to justify to management.

There is a need for a procedure that will consider the situation as a whole, a total dynamic field, including the personality. Such a test should consider evaluation of biographical data, personality factors, and test performance in tests of achievement and skills. The special need for this type of evaluation in industry exists in the selection of supervisors, as has been pointed out above. Until now, a scale composed of items of such a heterogeneous nature has received little attention; but this study addresses itself precisely to this point.

Many techniques have been suggested for item analysis in test or scale construction, according to du Mas (4). In most of these, each item is related individually to a variable---often the total score for all items in the scale---then those items having high correlations with the variable and low correlation with each other are selected, and a weight is often assigned each item on the basis of the item-variable correlation. The correlational methods most often used are biserial, point biserial, tetrachoric and Phi.

Du Mas (4) holds, further, that these techniques are open to certain criticisms. The data seldom fit the assumptions upon which the various methods of item analysis are based. The methods most widely used in item analysis depend markedly upon agreement in difficulty of the items. The criteria of conventional item analysis for the retention of an item make it necessary to discard often perfect or near perfect scale items. Tests constructed by these methods practically always seriously violate scale concepts and/or criteria. Therefore he concludes that tests constructed from item analyses should not in general be regarded as scales, but rather as primitive, useful and probably necessary antecedents to more adequate instruments constructed by more rational methods.

Manifest Structure Analysis is a new method of scaling introduced and described in detail by du Mas (4) for the purpose of extracting an ordered set of categories from a domain. The set of categories then can be utilized as a measuring instrument and not merely as a set of predictors. He defines Manifest Structure Analysis as the analysis of an ordered structure which is operationally extracted from an apparently chaotic domain by reference to the manifest relations existing between a set of categories and a continuum of magnitudes, or criterion scale. It is different from all other methods of scaling in that values of the continuum are manifest and are not generated as an inference from the stimulus items. Because of this, objects, items, or events which exhibit no phenomenal similarity, relationship, or order may be scaled.

Practically, Manifest Structure Analysis utilizes an ordered criterion (e.g., income levels) as the ordinate, and categories (e.g., test

SEGMENTAL MODEL

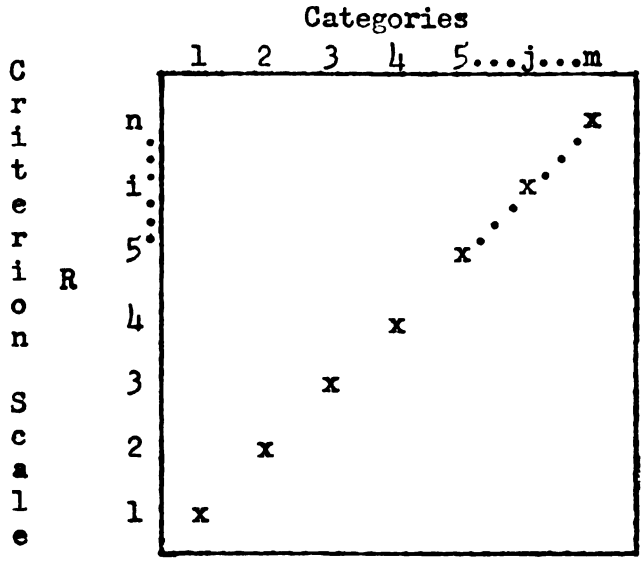


FIGURE I
Association surface for the
Segmental Model

(After du Mas, 4, p. 87)

scores) along the abcissa of a coordinate. The plotting of test scores automatically scales them to the criterion scale, yielding an ordered set of categories which indicate, or reflect, the magnitude, intensity, or degree of the criterion at any given point on the criterion scale. Scores which do not manifestly discriminate are dropped. The underlying notion is that categories may be differentially associated with some manifest variable in such a way as to form an ordered structure. (See Fig. 1, p. 9a). Thus, an ordered criterion of income may yield an ordered structure of scaled categories or items empirically obtained by Manifest Structure Analysis.

For the present study, data were made available by a client of an industrial psychological consultant of Michigan State College. The client, a manufacturing company of a highly specialized product, had accomplished a testing program of supervisors as part of a broader personnel evaluation program, under the direction of the consultant. The testing consisted of a battery of standard tests which had been found satisfactory in the previous testing and a personal information form. The testing was done in the offices of a professional psychologist who evaluated the results. The tests were administered and scored by a competent psychometrician. The testing session required about six hours for each subject.

## THE PROBLEM

The problem of this study was to select a set of ordered and weighted items or categories from biographical data forms and standard test information by means of which we could predict a subject's potential value to the company. The general hypothesis we wished to test was: There is a set of categories from biographical and test data which forms an empirical analogue of the Segmental Model (See Fig 1, p. 9a).

# PROCEDURE

## Subjects

Fifty-one supervisors or potential supervisors, all employees of the same manufacturing company, participated in a six hour testing program. Since criterion data were available on only forty-two of the subjects, by reason of transfer or leaving the company, it was necessary to discard the data of nine subjects. Of the remaining forty-two, ten subjects were held out to serve as a cross-validation group. They were selected on the basis of range along the criterion scale, only. That is, the second from the top, the second from the bottom, and then eight other subjects, were selected so as to be representative of the major criterion intervals. The number of subjects for the original sample was then 32. The number of subjects for the cross-validation group was 10.

## Criterion

The industrial psychological consultant, mentioned above, obtained from the company the following information for possible use as criteria: a numerical job classification which gave a numerical value to the level of supervisory responsibility; annual income (coded); hourly income (coded); merit ratings for 1953 and 1954; income change from 1950 to 1954; and job tenure. All names of supervisors were also represented by code numbers. Only the job classification and the coded hourly income appeared to reflect the various levels of supervisory function.

The coded hourly income, which excluded bonus and overtime pay, was selected as the best available criterion. This was operationally defined as a coded, dollars-per-hour, manifest scale of value of the supervisor's performance in this company.

## Apparatus

The du Mas Scaling Frame held 87 removeable slats placed together so as to present a flat surface slanting away from the vertical, at an angle convenient for placing thumbtacks into the slats while standing in front of it. Its outside dimensions were about four and one-half feet in height and five feet in width. One hundred holes for thumbtacks were drilled down the length of each slat at one-half inch intervals. Since the slats were one-half inch wide, and held firmly together, the holes formed straight lines top-to-bottom, across, and diagonally. The effect might be visualized as a rectangular surface made up of one-half inch squares or cells with a hole for a thumbtack in the center of each cell. There were 87 cells in each row across the frame, and 100 cells in each column. Each slat, or column, could be removed and shifted to another position where it would again fit this pattern of cells.

The criterion scale was attached to the left-hand margin of the scaling frame so that each individual, represented by a name code number and an hourly income code value, coincided with a horizontal row of cells. The criterion values were ordered from highest to lowest, with the highest at the top row of cells.

The slats, or columns, were numbered to represent categories, such as test score intervals. Thus, if an individual scored within this interval, the datum was entered into the appropriate cell where the individual's row and the category's column intersected.

## Basic Data

The battery of standard tests consisted of Bernreuter's Personality Inventory; the Social Intelligence Test prepared by Moss, Hunt, and Omwake; Wonderlic's Personnel Test; Bennett's Test of Mechanical Comprehension, Form AA; the Minnesota Clerical Test, by Andrew, Paterson, and Longstaff; How Supervise?, by File and Remmers; the Michigan Vocabulary Profile Test, by Greene; The Kuder Preference Record; the Washburne S-A Inventory (thaspic edition); the Study of Values, by Allport, Vernon and Lindzey.

The Guilford-Zimmerman Temperament Survey and the Thurstone Temperament schedule were used as alternate tests and therefore could not be utilized statistically. Records were not complete on the Wide Range Vocabulary Test, and therefore it could not be used.

The Personal History form was constructed by Harry C. Yudin, professional psychologist, in whose offices the testing was accomplished. The testing session was about six hours for each subject.

## Method of Analysis

If the data were numerical values, they were divided into three intervals, each representing roughly one-third of the subjects. Thus, if the scores for all the subjects on one test ranged from 60 to 90, and

if they appeared to be fairly evenly distributed, the three categories for this test became 0 to 70, 71 to 80, and 81 up. Each of these categories was numbered. If a subject's score placed him in the 71 to 80 category, a thumbtack was entered in the cell where his row and the category column intersected in the scaling frame. Biographical data, if not numerical, were divided into 'yes' and 'no' categories, on the assumption that a 'yes' response might be characteristic of top and bottom criterion ranges, and therefore not discriminate, while the 'no' response would, in this instance, discriminate the middle criterion range. The biographical categories were also given category numbers. Thus, Marital Status might become the following numbered categories: Married, Widower, Separated, Single, Divorced.

There was a total of 293 categories representing test scores, subtest scores, percentiles, numerical biographical data (e.g., age), and non-numerical biographical data (e.g., birthplace). Since the du Mas scaling frame held only 87 slats at a time, it was necessary to use extra slats, filling and evaluating 87 categories at a time. All the data were entered into their appropriate cells.

Certain categories were seen, when inspected individually, to discriminate some portion of the upper, lower, or middle range of the criterion scale. These categories were moved to the left side of the scaling frame by simply lifting the slat out of the frame and putting it back into the frame on the left side, after sliding the other slats to the right. Categories were rejected to the right side of the scaling frame if they were: multimodal, gappy, associated with a large part of the range, or

TABLE I

THE CATESCALE

| N | Categ. No. | Categories | Scale Value |
|---|---|---|---|
| 1. | 34 | Highest pay ever received. $151 per week - up | 4060.0 |
| 2. | 4 | Job at time of test. Supervision | 2888.0 |
| 3. | 94 | Organizations belong to. 1 | 2380.0 |
| 4. | 54 | Siblings. None | 2200.0 |
| 5. | 97 | Number of previous jobs. 0 - 1 | 2154.2 |
| 6. | 211 | Minn. Clerical Test, Numbers 212 - up | 2010.0 |
| 7. | 161 | Soc. Int. Test, Recog. Mental State. 0 - 15 | 1982.0 |
| 8. | 167 | Soc. Int. Test, Humor. 11 - 16 | 1980.0 |
| 9. | 151 | Washburn Social Adjust., Wishes, first 3. Job | 1957.5 |
| 10. | 214 | Minn Clerical, Numbers %ile. 86 - up | 1956.7 |
| 11. | 62 | Number of books or other important last month.0-1 | 1932.7 |
| 12. | 221 | Minn. Clerical, Names score. 0 - 110 | 1916.6 |
| 13. | 181 | Personnel Test. %ile 0 - 34 | 1841.0 |
| | | * | |
| 14. | 11 | Education 8th or below. | 1833.0 |
| 15. | 199 | Bennett Mech. Comprehension, Foremen %ile 67-up | 1811.7 |
| 16. | 273 | Kuder Pref. Record, Clerical 0 - 40 | 1802.7 |
| 17. | 185 | Personnel Test. I. Q. below 115 | 1799.6 |
| 18. | 100 | Bernreuter, Emotional Stability 0 - 34 | 1791.6 |
| 19. | 158 | Social Intelig. Test, Judgement. 0 - 20 | 1780.0 |
| 20. | 226a | How Supervise?, Shop Practice Sc. 0 - 12 | 1764.6 |
| 21. | 229 | How Supervise?, Company Policy. 0 - 12 | 1764.3 |
| 22. | 73 | Weight 186 - 200 | 1764.0 |
| 23. | 241 | Mich. Vocabulary Prof., Human Rel. 0 - 14 | 1761.1 |
| 24. | 244 | Mich. Vocabulary Prof., Commercial 0 - 12 | 1759.0 |
| 25. | 106 | Bernreuter, Extrovert. 0 - 34 | 1753.7 |
| 26. | 170 | Social Intelligence Test, Total 0 - 90 | 1751.4 |
| | | * | |
| 27. | 87 | Psychologist's evaluation (rated by author) low-1 | 1742.5 |
| 28. | 220 | Minn. Clerical, Names. Wrong. 5 - up | 1738.9 |
| 29. | 84 | Psych. eval. (rated on 7 point scale) average- 4 | 1735.0 |
| 30. | 172 | Soc. Intel. Test., College %ile. 0 - 34 | 1723.3 |
| 31. | 178 | Personnel Test, Number right. 0 - 12 | 1700.0 |
| 32. | 31 | Highest pay ever received. 0 - 99 per wk | 1682.6 |
| 33. | 57 | Siblings. Two | 1665.7 |
| 34. | 166 | Social Intelligence Test, Humor. 0 - 10 | 1660.9 |
| 35. | 153 | Washburn S-A Inv., one of first 3 wishes Social. | 1627.5 |
| 36. | 5 | Job at time of test. Non-supervisory | 1622.2 |
| 37. | 120 | Washburn S-A Inv. Score: "t" 6 - 10 | 1618.6 |
| 38. | 41 | Number of children 1 | 1600.0 |
| 39. | 286 | Scale of Values, Social score 31 - 35 | 1562.8 |

* Catescale is divided into upper, middle, and lower thirds.

Fig. 2

# Segmental Catescale
## Original Sample

| N | Name Code Number | Criterion R | 15628 286 | 16009 41 | 16186 120 | 16222 5 | 16275 153 | 16609 166 | 16657 57 | 16826 31 | 17000 178 | 17233 172 | 17350 84 | 17389 220 | 17425 87 | 17514 170 | 17537 106 | 17590 244 | 17611 241 | 17640 73 | 17643 229 | 17646 226 | 17800 158 | 17916 100 | 17996 185 | 18027 273 | 18117 199 | 18330 11 | 18410 181 | 19164 221 | 19327 62 | 19567 214 | 19575 151 | 19800 167 | 19826 161 | 20100 211 | 21542 97 | 22009 54 | 23800 94 | 28880 4 | 40600 34 | Catescale Number | Predicted Scores | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10 | 6670 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 3 | 4.7 | 4.0 |
| 2 | 16 | 3440 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 | 8.7 | 0.6 |
| 3 | 6 | 3330 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 | 1.4 | 6.6 |
| 4 | 32 | 2770 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 9.6 | 5.9 |
| 5 | 1 | 2660 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 | 1.3 | 1.9 |
| 6 | 5 | 2440 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 9.9 | 7.0 |
| 7 | 25 | 2070 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 2 | 1.1 | 8.6 |
| 8 | 39 | 2070 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.5 | 8.7 |
| 9 | 43 | 1990 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.0 | 5.1 |
| 10 | 38 | 1960 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 9.3 | 4.1 |
| 11 | 37 | 1850 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 9.3 | 9.3 |
| 12 | 27 | 1850 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.0 | 2.0 |
| 13 | 33 | 1790 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.9 | 0.4 |
| 14 | 8 | 1790 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.3 | 2.5 |
| 15 | 22 | 1750 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.7 | 9.1 |
| 16 | 51 | 1740 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.8 | 3.4 |
| 17 | 17 | 1740 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.7 | 3.9 |
| 18 | 41 | 1690 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.9 | 7.4 |
| 19 | 47 | 1680 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.6 | 6.8 |
| 20 | 29 | 1660 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.1 | 9.6 |
| 21 | 9 | 1600 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.2 | 1.0 |
| 22 | 7 | 1590 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.2 | 4.0 |
| 23 | 46 | 1590 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.7 | 5.9 |
| 24 | 36 | 1570 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.2 | 4.7 |
| 25 | 20 | 1560 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.6 | 1.0 |
| 26 | 12 | 1540 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.2 | 5.1 |
| 27 | 49 | 1510 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.0 | 2.7 |
| 28 | 21 | 1490 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 6.9 | 8.6 |
| 29 | 40 | 1480 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 0 | | |
| 30 | 13 | 1470 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.3 | 1.4 |
| 31 | 50 | 1470 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.5 | 4.3 |
| 32 | 31 | 1460 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.5 | 2.3 |

r = .95

## Cross-Validation Sample

| N | Name Code | R | 15628 286 | 16009 41 | 16186 120 | 16222 5 | 16275 153 | 16609 166 | 16657 57 | 16826 31 | 17000 178 | 17233 172 | 17350 84 | 17389 220 | 17425 87 | 17514 170 | 17537 106 | 17590 244 | 17611 241 | 17640 73 | 17643 229 | 17646 226 | 17800 158 | 17916 100 | 17996 185 | 18027 273 | 18117 199 | 18330 11 | 18410 181 | 19164 221 | 19327 62 | 19567 214 | 19575 151 | 19800 167 | 19826 161 | 20100 211 | 21542 97 | 22009 54 | 23800 94 | 28880 4 | 40600 34 | Catescale Numbers | Predicted Scores | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11 | 6000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 9.6 | 9.2 |
| 2 | 44 | 3330 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 9.5 | 8.0 |
| 3 | 24 | 2440 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.0 | 2.4 |
| 4 | 48 | 1970 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.7 | 1.0 |
| 5 | 34 | 1810 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.7 | 7.5 |
| 6 | 4 | 1740 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.2 | 4.2 |
| 7 | 26 | 1680 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.9 | 1.9 |
| 8 | 30 | 1580 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.0 | 2.8 |
| 9 | 3 | 1490 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 8.0 | 8.5 |
| 10 | 42 | 1470 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 7.1 | 2.3 |

r = .80

tescale
mple

| 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | Cate-scale | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17916 | 17995 | 18027 | 18117 | 18330 | 18410 | 19146 | 19327 | 19567 | 19575 | 19840 | 19826 | 20100 | 21543 | 22009 | 23800 | 23880 | 40469 | Cabescale Numbers | Predicted Scores | S |
| 100 | 185 | 273 | 199 | 11 | 181 | 221 | 62 | 214 | 151 | 167 | 161 | 211 | 97 | 54 | 94 | 4 | 34 | | | |
| | | | | | | | | | | | | | | | | x | x | 3 | 47 | 4.0 |
| | | | | | | | | | | | | | x | | x | x | x | 2 | 87 | 0.6 |
| | x | | | | | x | | | | | x | | | | | x | | 2 | 14 | 6.6 |
| | x | x | | | x | x | x | | x | x | | | | | | x | | 1 | 96 | 5.9 |
| | | x | | | | | | x | | x | | x | x | | | x | | 2 | 13 | 1.9 |
| x | | | x | | | | x | x | x | | | x | | x | | x | | 1 | 99 | 7.0 |
| x | | | x | x | | x | x | x | | x | | x | | | | x | x | 2 | 11 | 8.6 |
| | x | | x | x | | | x | x | | x | | x | x | | | | | 1 | 85 | 8.7 |
| | x | x | x | x | x | x | | x | | x | | x | | | | | | 1 | 80 | 5.1 |
| x | x | | x | x | | | | x | | x | | x | x | x | x | x | | 1 | 93 | 4.1 |
| x | | | | x | x | x | x | | x | | | | | | | | | 1 | 93 | 9.3 |
| x | x | x | x | | | x | | | x | | | | | | | | | 1 | 80 | 2.0 |
| | x | | | x | x | | x | x | x | | | x | | | | | | 1 | 79 | 0.4 |
| x | x | | x | | x | x | | | | | | | x | | | x | | 1 | 83 | 2.5 |
| | x | x | | | | x | | | x | x | | | | | | x | | 1 | 87 | 9.1 |
| | x | x | | | x | x | | | x | | | | | | x | | | 1 | 78 | 3.4 |
| | x | | | x | x | x | x | | | | x | | | | | | | 1 | 77 | 3.9 |
| x | x | x | | | x | | | x | | x | x | x | | | | | | 1 | 79 | 7.4 |
| x | x | x | | x | x | | | | | | | | | | | | | 1 | 76 | 6.8 |
| | x | x | | | | x | x | | | x | | | | | | | | 1 | 81 | 9.6 |
| x | x | | | | | x | x | | | | | | | | | | | 1 | 72 | 1.0 |
| | x | x | x | x | | | | | | x | x | | x | | | | | 2 | 82 | 4.0 |
| x | x | x | | x | x | x | x | | | | | | x | | | | | 1 | 77 | 5.9 |
| | x | | | | | | | | | x | | | | | | | | 1 | 72 | 4.7 |
| | x | x | | | | x | | | | | x | | | | | | | 1 | 76 | 1.0 |
| | x | x | x | | | | | | | | | | | | | | | 1 | 72 | 5.1 |
| x | x | x | x | | | | | | | | | | | | | | | 1 | 70 | 2.7 |
| | x | | | | | | | | | | | | | | | | | 1 | 69 | 8.6 |
| | | | | | | | | | | | | | | | | | | 0 | | |
| x | x | | | | x | | | | | | | | | | | | | 1 | 73 | 1.4 |
| | | | x | | | | | x | x | | | | | | | | | 1 | 75 | 4.3 |
| | x | x | x | | | x | | x | | | | x | | | | | | 1 | 75 | 2.3 |
| | | | | | | | | | | | | | | | r=.95 | | | | | |

mple

| 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | Cabescale Values. | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17916 | 17995 | 18027 | 18117 | 18330 | 18410 | 19146 | 19327 | 19567 | 19575 | 19840 | 19826 | 20100 | 21543 | 22009 | 23800 | 23880 | 40469 | | | |
| 100 | 185 | 273 | 199 | 11 | 181 | 221 | 62 | 214 | 151 | 167 | 161 | 211 | 97 | 54 | 94 | 4 | 34 | Cabescale Numbers | S | |
| | | x | | | | | x | x | x | | | x | | | | x | | 1 | 96 | 9.2 |
| | x | | | | | x | | | | x | | | | | | x | | 1 | 95 | 8.0 |
| x | x | | | | | | | x | x | x | | x | | | | | | 1 | 80 | 2.4 |
| | x | x | | | | x | | | | x | | | | | | x | | 1 | 87 | 1.0 |
| x | x | | | x | x | x | x | | x | x | x | | | | | | | 1 | 77 | 7.5 |
| | | | | | | x | | | | | | | x | | | | | 1 | 72 | 4.2 |
| | x | | | | | x | | | x | | x | | | | | | | 1 | 79 | 1.9 |
| | x | | | | | | x | | | | | | | | | | | 1 | 70 | 2.8 |
| | | x | x | | · | x | | | x | x | | | | | | | | 1 | 80 | 8.5 |
| x | | | | | | x | | | | | | | | | | | | 1 | 71 | 2.3 |
| | | | | | | | | | | | | | | | r=.80 | | | | | |

not sufficiently associated with individuals in the sample, in accordance with the du Mas (4) specifications.

Thirty-nine categories were thus selected. (See Table I, p. 16). The scale value for each category was calculated by means of the du Mas (4) formula:

$$V = \frac{\text{Sum } R}{N} ,$$

where:

V: category scale value
Sum R: criterion score (income) of all individuals associated with a a particular category
N: number of R values associated with the category.

The selected categories were then ordered with regard to the magnitude of their scale values, in accordance with the du Mas (4) instructions. (See Fig. 2, p. 17) These scale values constituted a catescale. (See Table I)

A score was then calculated for each individual, by means of the du Mas (4) formula:

$$S = \frac{\text{Sum } V}{n} ,$$

where:

S: score from the catescale.
Sum V: sum of catescale values of all categories with which an individual is associated.
n: number of categories with which an individual is associated.

This operation attempted to predict the criterion score from the catescale extracted from the data.

The product moment correlation was calculated between the criterion distribution R and the predicted score distribution S.

One supervisor did not appear in any category, and was therefore not scored. The proportion D of the sample for whom a score is determinate was therefore calculated, by the du Mas formula:

$$D = \frac{N_S}{N_R},$$

where $N_R$ : the number of individuals in the sample that have an R value; and $N_S$ : the number of individuals in the sample for which a score is determinate.

Cross-validation was done by applying the scale values, or weights, of the catescale constructed from the first sample, to the cross-validation sample of 10 subjects. A product moment correlation was calculated between the distribution of criterion values in this sample and the predicted score distribution.

## RESULTS AND DISCUSSION

A Catescale (categories possessing scale values) consisting of 39 categories, was operationally extracted from a total of 293 categories of biographical and standard test data. (See Table I, p. 16). The Catescale values were used to predict the original criterion scale values of the original sample of 32 supervisors. The validity coefficient was $r = .95$ .

The same Catescale values were used to predict the criterion scale values of a cross-validation group of 10 supervisors. The correlation between the predicted criterion values and the original criterion values of the cross-validation group resulted in a coefficient of $r = .80$ .

The general hypothesis: There is a set of categories from biographical and standard test data which forms an empirical analogue of the Segmental Model (Fig. 1, p. 9a). This was clearly supported.

Reference to Figure 2, page 17, will reveal the Criterion, R, values in the third column from the left. Also, in the extreme right hand column, the Predicted Scores, S, will be seen. It will also be noted that each individual is represented by a name code number in the column next to his criterion value column, and that the N of the original sample is 32.

The Criterion, R, column is the original criterion scale, ordered from highest to lowest, of the original sample. The Predicted Scores, S, column contains the predicted criterion scores, or values, calculated from the Catescale Values in the row along the top of Figure 2. A predicted score is the mean of the catescale values represented by x's in the individual's row.

Considering the Original Sample, upper part of Figure 2, the validity coefficient has been expressed as the correlation between the distribution of the Criterion, R, values and the distribution of predicted criterion, S, values. As stated above, the validity coefficient was: r = .95 .

Considering the cross-validation sample, the lower part of Figure 2, the same column headings will be seen. Also, it will be noted that the Catescale values are identical with those in the original sample above. The N of this group is 10. The name code numbers reveal that these supervisors are not included in the original sample. The Catescale found for the first sample was applied to the cross-validation sample, and the predicted criterion values, S, calculated in the same way, by taking the mean of the catescale values represented by x's in the rows of the individuals. Correlation between the criterion, R, distribution of the cross-validation group and its distribution of predicted criterion scores, S, yielded: r = .80 .

The Catescale of 39 categories, with category numbers and scale values, is presented separately in Table I, page 16. The category numbers and scale values may also be identified in the second and third rows of Figure 2,(p.17). In Table I, (p.16), the content of the categories of the Catescale are shown. It may be recalled that the categories were made, and numbered individually, by dividing numerical data (e.g., scores on a test) into three intervals and biographical data into appropriate categories (e.g., married, single). The 39 categories of Table I, with their scale values, represent the Catescale operationally extracted from 293 categories as described under Method of Analysis beginnign on page 14 of this thesis. It may be noted in Table I, page 16, that the Catescale is divided into thirds.

Apparently, it should be possible to read a general description of the upper third of the discriminating supervisory qualities of our catescale. However, all of the supervisors in the upper third of our sample are not uniformly associated with the upper third of the categories. How these categories interact, what effect they may have on each other when associated together, or what combinations produce what results, requires a type of analysis or speculation beyond the scope of this study.

The categories themselves are of interest. Some were quite surprising and unexpected. Some might have been expected in the light of conventional theory. However, these categories were operationally and objectively extracted from the chaotic domain of all the data in conformance with the principles of manifest structure analysis, as presented by du Mas (4). The fact remains that these categories, or combinations of categories, appear to discriminate with high validation and cross-validation the various levels of the criterion scale.

Several modifications of procedure suggest themselves which might yield a special type of catescale for a specific purpose. For example, an intensive study might be made of a specific area, or of one test, or of one kind of test, for which du Mas (4) had described an intensive model. In this study, items were selected for highest possible discrimination. Item selection could be somewhat more liberal in specific areas, or in all areas, so as to include a more complete description from the catescale, even though this would increase the variance and therefore decrease the validity somewhat. Dichotomizing all data, instead of dividing them into three intervals was another possibility in dealing with categories.

It would be possible to make several interesting a posteriori interpretations of these categories. This suggests itself as a potentially rich source of new ideas and psychological insights. This however, was not the purpose of this study. The speed with which new catescales can be selected, and weights calculated, seems to offer possibilities for successive approximations to a best ordered structure, as du Mas (4) suggests. Also it would also appear quite possible to substitute different criteria in the same set of data.

Since this was the first empirical study of supervisors with Manifest Structure Analysis, the apparent utility of this method has by no means been fully explored. This is especially true in view of the scarcity of published evaluative results, and the even greater scarcity of positive findings, in the field of supervision, as pointed out in the introduction of this paper. Further research and greater experience with this method should reveal the most fruitful areas of application and the specific uses in which results would be most definitive.

## SUMMARY AND CONCLUSIONS

The problem of this study was to select a set of ordered and weighted categories or items from biographical forms and standard test information, by means of which a subject's potential value to a company might be predicted. A group of 32 supervisors or potential supervisors participated in a six hour testing program which included a personal information form and a battery of standard tests. A criterion scale of coded hourly income was obtained and defined as a coded, dollars-per-hour, manifest scale of value of the supervisor's performance in this company.

A Catescale (literally, scale-weighted categories) was extracted from the biographical and test data arranged on a criterion continuum of coded hourly income. (See Fig. 2, p. 17). The validity of the Catescale in predicting the original criterion values was: $r = .95$ (See Table I, p.16). A cross-validation group of 10 supervisors, selected only for range along the criterion continuum, was scored with the catescale values found on the original sample. Cross-validation correlated: $r = .80$ (See Fig. 2, p.17). Both validity and cross-validation coefficients were well beyond the one percent level of confidence.

The general hypothesis was: There is a set of categories from biographical and standard test data which forms an empirical analogue of the Segmental Model. The hypothesis was clearly supported. The Segmental Model (Fig.1, p. 9a) is one of several models presented by du Mas (4) for scale construction in Manifest Structure Analysis. It is a mathematical

model representing a perfect correlation. It is represented as a scatter diagram with a criterion scale at the ordinate, the left side, and categories along the abcissa, at the top. The criterion scale is ordered with the highest value at the top. The model would represent a Pearson r of 1.

The important difference between conventional item analysis and manifest structure analysis is that categories (columns) are selected to meet the assumptions of Pearson r computation, by inspection of each column (category with individuals plotted) separately, ordering the columns in terms of mean value, and computing Pearson r between the original criterion distribution and a predicted distribution. (See Fig. 2, p. 17). Category values, or catescale weights, are the means of the criterion scale values represented in the columns. Predicted criterion values are the means the rows in terms of the column values. Data are represented in the scatter by x's. Justification for the procedure is cross-validation.

Practical conclusions: A new technique has been developed which in one case has extracted a catescale of 39 categories from 293 categories of biographical and test data, with extremely high validation and cross-validation. Practical advantages of this technique are: simplicity, speed, and low cost. Highly trained personnel would not be required in a situation where scoring tables and graphs could be set up. Flexibility and extremely wide applicability, "wherever rating scales or tests are applicable" were indicated in agreement with du Mas (3, p. 117). Analysis of case histories, interview forms, and other multivariate data-gathering instruments of psychosocial phenomena appears to be uniquely within the scope of this technique.

BIBLIOGRAPHY


1. Barnett, G. J., Handelsman, I., Stewart, L. H., and Super, D. E. The occupational level scale as a measure of drive. Psych. Monographs, 1952, 342.

2. Cleeton, G. U. and Mason, C. W. Executive ability. The Antioch Press, Yellow Springs, Ohio. 1946.

3. du Mas, F. M. Continua, catemensions, catescales. J. Clin. Psychol., 7, 1951. 112-117.

4. du Mas, F. M. Manifest structure analysis. Unpublished treatise.

5. File, Q. W. and Remmers, H. H. Studies in supervisory evaluation. J. Appl. Psychol. 30, 1946, 421-425.

6. Gibb, C. A. The principles and traits of leadership. J. Abn. and Soc. Psychol. 1947, 272.

7. Harrell, W. Testing cotton mill supervisors. J. Appl. Psychol. 1940, 24, 31-35.

8. Jenkins, J. G. Validity for what? J. Consult. Psychol. 1946, 10, 93-98

9. Lawsche, C. H., Jr. Principles of personnel testing. First Edition. McGraw Hill, Inc. N.Y. 1948.

10. Rush, C. H., Jr. A factorial study of sales criteria. Pers. Psychol. 6, 1953.

11. Strong, E. K., Jr. Interests of senior and junior public administrators. J. Appl. Psychol. 1946, 30, 55-71.

12. Strong, E. K., Jr. Vocational interests of men and women. Stanford Univ. Press, Palo Alto. 1943.

13. Strong, E. K., Jr. Vocational guidance of executives. J. Appl. Psychol., 1927, 11, 331-347

14. Super, D. E. Appraising vocational fitness by means of psychological tests. Harper Bros., N.Y., 1949.

15. Taylor, E. K., Schneider, D. E., and Symonds, N. A. A short forced-choice evaluation form for salesmen. Pers. Psychol., 6, 1954.

BIBLIOGRAPHY

16. Thompson, C. E. Selecting executives by psychological tests. Educ. Psychol. Measmt. 1947, 7, 773-778.

17. Wadsworth, G. W. Tests prove their worth in a utility. Pers. J., 1935, 14, 183-187.

18. Wonderlic, E. F., and Hovland, C. J. The personnel test; a restandardized abridgement of the Otis S-A test for business and industrial use. J. Appl. Psychol., 1939, 23, 685-702.

ROOM USE ONLY