1954

The Thurstone Temperament Schedule as an Instrument of Supervisory Evaluation

Francis Xavier Paone

Loyola University Chicago

Recommended Citation
http://ecommons.luc.edu/luc_theses/1200

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.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1954 Francis Xavier Paone
THE THURSTON TEMPERAMENT SCHEDULE AS AN INSTRUMENT OF SUPERVISORY EVALUATION

by
FRANCIS XAVIER PAONE

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

June 1954
LIFE

Francis Xavier Paone was born in New York City, December 19, 1919.

He was graduated from St. Francis Xavier High School, New York City, New York, June, 1938, and from Fordham University, Bronx, New York, June, 1942, with the degree of Bachelor of Arts.

From 1942 to 1946 the author served in the Armed Forces as an Air Force Communication Officer. Upon discharge he obtained a position as an electrical engineer with a large midwestern communications equipment corporation, which position he holds at the present time. He began his graduate studies in Psychology at Loyola University in February of 1951. The outbreak of the Korean War caused him to be recalled to active duty with the Armed Forces for fifteen months. Upon his return he resumed his graduate studies at this same University. A portion of his graduate work was completed at the University of New Mexico while in the Armed Forces in 1951-1952.
PREFACE

The name "West Company," used in this thesis is fictitious. The nature of the information obtained in the compilation of the data made it necessary for the author to adopt this procedure. The company at which the research was conducted was policy-bound against allowing the results of this research to be published in any form. There was no objection, however, to the use of the data in this thesis, so long as the anonymity of the company be safeguarded. This the author has attempted to do.

Thanks are due Doctor H. V. Jones of the University of Chicago for his aid in obtaining the Psychometric Laboratory Reports of Doctor L. L. Thurstone upon which the Thurstone Temperament Schedule is based.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A. Statement of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>B. An Hypothesis</td>
<td>7</td>
</tr>
<tr>
<td>II. REVIEW OF THE RELATED LITERATURE - PART I</td>
<td>9</td>
</tr>
<tr>
<td>A. The Merit Rating, Early Plans</td>
<td>9</td>
</tr>
<tr>
<td>B. Some Early Rating Studies at West</td>
<td>10</td>
</tr>
<tr>
<td>C. The Introduction of the Over-All Rating</td>
<td>11</td>
</tr>
<tr>
<td>D. Recent Developments in Rating at the West</td>
<td>12</td>
</tr>
<tr>
<td>E. Special Research on the Reliability of Ratings</td>
<td>15</td>
</tr>
<tr>
<td>F. The Validity of the Merit Rating</td>
<td>16</td>
</tr>
<tr>
<td>REVIEW OF THE RELATED LITERATURE - PART II</td>
<td>17</td>
</tr>
<tr>
<td>A. The Dimensions of Temperament, The Trait Theory</td>
<td>17</td>
</tr>
<tr>
<td>B. Guilford's Research on the Dimensions of Temperament</td>
<td>18</td>
</tr>
<tr>
<td>C. The Contribution of Constance Lovell</td>
<td>24</td>
</tr>
<tr>
<td>D. Thurstone's Research on Dimensions of Temperament</td>
<td>27</td>
</tr>
<tr>
<td>REVIEW OF THE RELATED LITERATURE - PART III</td>
<td>31</td>
</tr>
<tr>
<td>A. The Thurstone Temperament Schedule, Its Purpose</td>
<td>31</td>
</tr>
<tr>
<td>B. A Description of the Areas Covered</td>
<td>31</td>
</tr>
<tr>
<td>C. Format of the Thurstone Temperament Schedule</td>
<td>33</td>
</tr>
<tr>
<td>D. Administration of the Test</td>
<td>34</td>
</tr>
<tr>
<td>E. Plotting the Individual Profile</td>
<td>35</td>
</tr>
<tr>
<td>F. Interpretation of a Profile</td>
<td>35</td>
</tr>
<tr>
<td>G. Norms for Men and Women, Boys and Girls</td>
<td>38</td>
</tr>
<tr>
<td>H. The Reliabilities for the Seven Areas</td>
<td>39</td>
</tr>
<tr>
<td>I. Validity Studies</td>
<td>41</td>
</tr>
<tr>
<td>III. THE RESEARCH AND ITS RESULTS</td>
<td>46</td>
</tr>
<tr>
<td>A. The Test Situation</td>
<td>46</td>
</tr>
<tr>
<td>B. The Sample</td>
<td>47</td>
</tr>
<tr>
<td>C. The Merit Rating</td>
<td>48</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>D. Treatment of the Data</td>
<td>49</td>
</tr>
<tr>
<td>E. An Average Supervisor's Profile</td>
<td>50</td>
</tr>
<tr>
<td>IV. CONCLUSIONS</td>
<td>53</td>
</tr>
<tr>
<td>A. Summary</td>
<td>53</td>
</tr>
<tr>
<td>B. Interpretation of the Test Data</td>
<td>54</td>
</tr>
<tr>
<td>C. The Direction of Future Research</td>
<td>58</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>60</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>63</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Chapter III</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. RELIABILITY COEFFICIENTS FOR THE SEVEN AREAS</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>II. TEST-RETEST RELIABILITY COEFFICIENTS, MEANS AND STANDARD DEVIATIONS FOR EIGHTY-ONE MALE EXECUTIVES</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>III. MEANS, STANDARD DEVIATIONS, PEARSON CORRELATIONS AND SIGNIFICANCE BETWEEN MERIT RATINGS AND PRIMARY TRAITS</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTERQUARTILE RANGES OF THE SEVEN TEMPERAMENT TRAITS PLOTTED ON THE THURSTONE TEMPERAMENT SCHEDULE PROFILE GRAPH TO PORTRAY THE COMPARISON BETWEEN THE TEST GROUP AND THE STANDARDIZATION POPULATION</td>
<td>51</td>
</tr>
</tbody>
</table>

vii
CHAPTER I

STATEMENT OF THE PROBLEM

In modern industry there is little room for erroneous thinking on the part of management in its decisions. The criterion by which every activity is measured is the yard-stick of utilitarianism and economical efficiency. Nevertheless, the man is not disregarded. In the architecture of an industry he has become the mortar which is used to bind together the bricks of cohesive organization. Where the mortar is poorly utilized or poorly selected the bricks become loose and the organization falls apart. For this reason in recent years it has become apparent to management in the industrial world that the selection of the man is the crucible in which the success of the venture will be tested. In his thinking about his help the typical executive in the past was more or less guided by the question, "How much will we save if we did without him?" Today, the typical personnel manager asks himself, "How will we be helped with him?" This change of emphasis from the negative to the positive is evident in the developmental trends of modern industry.

It is only natural for one going a step further in the
direction of this kind of thinking to ask himself how it is possible to know if he is most advantageously utilizing his employee to the optimal betterment of his business. The answer to this question is not a simple one and is not one found in the report of the executive committee on current expenditures. When a management board comes to this realization it has begun to solve the problem for only then does it admit that the question has too many facets for a simple solution. This is the realm of the industrial psychologist and the solution of this problem with his help is not nearly as formidable.

Psychological attempts at measuring the worth of an individual in respect to his job have been more or less successful depending largely on the extent to which management has been willing to accept counsel in this connection. The ramifications of this question are not within the scope of this thesis for we are interested in something more than the routine selection of the right worker for the right job. However, it might be said in passing that unless the thinking of management is healthy in regard to this basic process there is small hope of achieving the optimal effectiveness toward which one might strive.

Going beyond the normal selection processes which are characteristic of modern industrial procedures we might ask ourselves if the goal of effective employment practice is reached with the intelligent use of selective screening. If this were
true there would be no loss of efficiency in industries in which sound and reasonably effective screening programs have been installed. One would be unschooled indeed in the problems of modern business if he believed this to be the case. Most industries in these times have come to realize the importance of adequate screening procedures in the selection of employees. This does not seem, however, to be an adequate solution to the problem. Many millions of dollars are lost each year because of employees, who although apparently suited to the job, prove to be detrimental to the good of the organization as a whole. Millions more are lost because the problem employee is not controlled, is not aided, and indeed is not recognized as such by his superiors. Here is, then, the nucleus of our problem.

Statement of the Problem

Between the board of directors of any corporation and the people who actually manufacture the product there exists a great gap which must be filled by the middle man known as the supervisor. This person is the important link in the chain of communication between the president of the board and the punch press operator or truck driver upon whom the existence of the company depends. Much is expected of this person and much is usually received. Here is the person who interprets the needs of the employees and passes his feelings along to the people who
are too far removed from the scene to appreciate the problems of the working man. Here is the person who must demand the best from his group and must exact from the worker the finest of creations. Upon this person, also, the responsibility falls when production slackens and a scapegoat is needed to explain away a failure. This is the man upon whom the efficiency and productivity of the worker finally depend.

It is not at all difficult to understand why the conception of what constitutes good supervisory material is important to management. It would be extremely desirable if we could find the qualities which go to make up good supervision in a simple psychological experiment in which all the intervening variables were held constant and supervisory ability were the only unknown. If such a thing were possible human nature would have long since been reduced to a numerical value on an arithmetical continuum and psychological studies could be concluded with the completion of good courses in physics and geometry. Unfortunately the problem cannot be so easily solved. The interrelationships of human instincts, drives, needs and personality forces are far too complex to lend themselves to ready measurement.

This is not intended to mean, however, that it would be impossible to employ psychological instrumentation in an effort to discover some of the idiosyncracies of human behavior
relevant to the supervising situation. This thesis shall attempt to do just this. Although it would be perhaps impossible to delineate the perfect supervisory personality it would be rewarding and interesting to find some of these qualities which the average successful supervisor has been shown to possess. Even this meager information may be of value. Before proceeding on with this task, however, it is necessary for us to examine the method industry has been wont to use in evaluating its supervision. The "merit rating" approach is familiar to all of us. It has existed in substance from the time of Aristotle and has been employed wherever individual differences are discussed and interrelated. Primarily intended to provide a means of comparing individuals in a competitive organization, it depends to a large extent upon the skill, honesty and experience of the rater. In most industrial applications it simply defines the individual's worth to the company in comparison to his fellow workers. Because it is relatively accessible for examination and held in high esteem at this company the writer has decided to choose it as his criterion of job success for the thesis purposes. If the supervisor is successful in achieving the end for which he is hired he would in all probability obtain a high merit rating.

In studying the supervisor then we have at least one known--his merit rating. We know from this how he ranks in com-
parison to his fellow supervisors. If we could find some way to correlate this known measure of achievement with other facets of the supervisor's total personality, we might be able to draw some valid inferences regarding what we would look for in selecting future supervisors. We might, furthermore, even find the answer to one of the questions as yet unanswered in regard to the merit rating itself, i.e., what facets of the supervisor's personality most highly correlate with merit ratings as a measure of job success.

In deciding which of the many avenues of approach in the measurement of personality traits to employ in this research the writer was guided by the fact that most personality tests describe a person in terms of psychotic or neurotic tendencies. Since, for practical purposes, we might accept the premise that a moderately successful supervisor would be reasonably well adjusted, these clinical stereotypes do not seem to provide the best method of describing his personality characteristics. The writer has therefore sought an instrument that emphasizes important, stable traits which describe how normal, well adjusted people differ from each other.

The Thurstone Temperament Schedule was designed for

this purpose. It is limited to a practical description of important aspects of temperament and makes no attempt to appraise the degree of conflict, insecurity, or maladjustment. It is designed to assess those traits which are relatively permanent for each person, and excludes those which reflect recent social experiences or exposure to propaganda. The profiles we might obtain from a schedule such as this would provide us with a psychological description of the average supervisor and if we were able to correlate the results of such a schedule, trait by trait, with merit ratings on the same individuals, something might be accomplished in our endeavor to learn what constitutes good supervision.

An Hypothesis

It would seem that if we are to have any confidence in the value of the merit rating as a criterion of job performance we should expect to find some relationship between the merit rating and those traits among supervisors which we might expect would be most conducive to excellence in supervisory ability. Of Thurstone's Temperament Traits which we shall see have been described as mutually independent and primary in themselves, we might assume that the Dominant (D), the Sociable (S) and the Emotional Stable (E) traits would most likely be outstanding in the typical supervisory profile. If this were true, Pearson²

correlations computed between the merit ratings and a series of Thurstone Temperament Schedule scores obtained from a random supervisory group should show significant differences between superior, average and below average supervisors insofar as these traits are concerned. This, of course, supposes the accuracy and reliability of the merit rating as a measure of supervisory excellence and of this more will be said in the next chapter. This thesis will attempt to test this hypothesis.

Before explaining the structure of the experiment and its results, it would be well to examine the merit rating itself and the psychological instrument selected for use in the research, the Thurstone Temperament Schedule.
CHAPTER II

REVIEW OF THE RELATED LITERATURE - PART I

THE MERIT RATING

Early Plans at the West Company--Attribute Ratings

In examining the records of what had been done at West in the field of merit rating, the writer was greatly impressed by the amount of study and thought which has been devoted to it over a period of years. The records show that experiments were carried out in this field as far back as 1925. Prior to 1937, however, merit ratings at West were of the attribute type. That is to say, the rater assigned a separate value for each of a number of items, such as "Quality of Work," "Quantity of Work," etc. Separate scales were provided for supervisors and non-supervisors, and these in both cases ran into considerable detail. For example, there were fifteen attributes on the non-supervisory scale in some of these rating scales and twenty-one on the supervisory scale. Overlapping in the meaning of these defined characteristics was also in evidence at various points. The rater had a difficult task on his hands if he had to appraise a large number of employees. As one rater remarked in recalling
these ratings, "After the first few ratings my pencil started taking care of the rest."

Some Early Rating Studies at West

As a result of the growing discontent resulting from this type of rating plan several studies and experiments were initiated in 1937. Among the most interesting was the series of conferences with supervisors which resulted in the "Fourteen Points" used in rating supervisory employees. These points were descriptive statements which summarized the opinions of the rater on major characteristics required for success as a supervisor at the West Company. Each of the points described both a desirable and an undesirable trait, and the rater was asked in each case whether the ratee possessed the particular attribute to a high or low degree.

Another experiment was the attempt to use a check list consisting of seventy-five items describing desirable and undesirable traits of an employee or a supervisor. The rater designated with a check mark the statements most closely describing the ratee. Since the list contained seventy-five items it would seem that it multiplied some of the more cumbersome features of previous attribute ratings.

Also about this time it was found that a single factor evaluation of job ability in merit rating was at least as depen-
dable and perhaps even more dependable than multi-factor appraisals and was, in addition, much less costly.

Introduction to Over-all Ratings

It is not surprising then, to find the 1937 non-supervisory rating survey swinging over from a multiple-factor plan to a two-factor evaluation and utilizing a greatly simplified form printed on an I.B.M. card. Thus, for the first time at West, the rater was not faced with a highly detailed form on which to record his evaluation, and a machine process could be used in summarizing the results of merit rating surveys.

The two factors now evaluated were simply (1) the employee's over-all effectiveness on his job and (2) his supervisory potentiality. When rating job ability the rater checked the appropriate box on a nine-square scale ranging from highest to lowest, and when rating supervisory ability, the rater simply checked the one of three squares which best indicated the ratee's supervisory potentialities. Except for changes in the scale and other minor modifications, this is the rating form being used at this company today. The nine-square scale was reduced in the 1938 version to five squares, and a rank order method was added to the form a short time later. With the addition of this second rating system, the supervisor was given the choice of using two rating methods:
1. He could distribute his people over a five-square scale as follows:

- the upper 10%
- the next 20%
- the middle 40%
- the next 20%
- the lower 10%

- or -

2. He could arrange them in rank order position from highest to lowest and record the number representing each employee's position in the group on the tab card.

By 1946, the rank order method had superseded the squares method of rating, except in cases where employees were on kinds of work in which they could not be compared with others.

Recent Developments in Merit Rating at the West Company

In recent years, development work on merit rating has been aimed at obtaining more accurate and dependable ratings. The rank order method whereby employees are compared with each other in well defined groups according to types of work was introduced in part to combat the tendency of some raters to rate too high or too low. By ranking employees consecutively in the order of their merit, raters were forced to distribute their people over the entire rating scale from high to low.
More care was exercised in recent years in grouping employees for rating purposes and in assigning raters. In the past, for example, the privilege of rating was often extended to all supervisors on a branch-wide basis who were one level above the employee being rated. At present, only qualified raters are assigned to rating groups, that is, supervisors who have first-hand knowledge of the work of each employee in the group.

In recent rating surveys, also, close supervision was given by the Personnel Coordinating Organization to the proper formation of competitive groups of employees for rating purposes. Considerable help was given to the rank and file of the supervisor group on this problem so that measurements of opinions about the abilities of the employees could be done more accurately. It is clear that if employees were misgrouped to start with their ratings would be of little significance.

During rating periods of recent years a series of preliminary conferences have been conducted with raters a few days prior to the time the actual ratings were to take place. At these conferences, problems of grouping and proper evaluation of the employee were discussed as well as the techniques to be employed. This permitted raters to give thought to the appraisals they would be making within a short time and they therefore came to the actual rating conference more adequately prepared to
do this important work. In addition, a series of checks was introduced to guarantee accuracy in handling data and reporting results to the Personnel Records Department and several research studies were begun to secure more accurate control.

With the widespread use of the squares method, there arose the problem of more precise definition and illustration of the five levels of ability on this scale. After much discussion with line supervisors the following descriptions were adopted as guides for raters and the old categories, "the upper 10%," "the next 20%," were disregarded.

<table>
<thead>
<tr>
<th>Numerical Equivalents</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Outstandingly good</td>
</tr>
<tr>
<td>90</td>
<td>Very good. Better than the job requires.</td>
</tr>
<tr>
<td>80</td>
<td>Satisfactory. Meeting the requirements of the job.</td>
</tr>
<tr>
<td>70</td>
<td>Reasonably satisfactory. Not meeting all job requirements.</td>
</tr>
<tr>
<td>60</td>
<td>Poor. Does not meet job requirements. Man and job not matched.</td>
</tr>
</tbody>
</table>

These descriptions were extremely helpful to supervisors when using the squares rating.
Special Research on the Reliability of Ratings

A special experiment was undertaken to determine how reliable or stable the rank order and square methods actually were. While it is reasonable to suppose that different qualified raters may have different opinions of the relative worth of a particular employee, it seems equally reasonable to believe that a good rating should reflect a considered judgment that will, for a given rater, not fluctuate radically over a relatively short period of time. Data for the study were obtained by having all raters of an operating sub-branch rate a second time the employees whom they had rated approximately six weeks previously. The number of employees involved numbered approximately three hundred fifty. No rater knew that a second rating would be requested. When compared statistically, the first and second ratings showed coefficients of correlation ranging from .85 to .97. Square ratings of employees who were grouped competitively tended to be somewhat more stable than square ratings of ungrouped employees. Ratings made by the various levels of supervision were studied separately in this experiment to determine whether any important differences were observable in the stability of the judgments expressed. The opinions of group chiefs were somewhat more stable than those of section chiefs, probably because in most cases they were closer to the people rated and knew them more intimately. Ratings made at the department chief level were the most stable of all, but it
must be pointed out that they were concerned only with the group and section chiefs under them—a relatively small group in comparison.¹

**The Validity of the Merit Rating**

There is no record of a validation study of the merit rating at present in use at this company. The Personnel Development and Research Department, when questioned regarding this fact, pointed out that one was undertaken a short time ago and as yet no conclusive results have been obtained. The company was reluctant to give any further information in this regard. At the writing of this thesis, therefore, the only indication of the accuracy and validity we have of this criterion is the fact that it is held in much esteem by the people that use it and, in the eyes of the interested people at this company, is yielding excellent results.

¹ The writer wishes to point out that this study is reported, not because it appears psychologically or statistically sound or significant but because it was the only one of its type conducted to date at this company and as such deserves mention.
REVIEW OF THE RELATED LITERATURE - PART II

THE DIMENSIONS OF TEMPERAMENT

The Trait Theory of Personality

Nearly all the prominent questionnaires have sought to describe the individual in terms of traits. If a test is to be a measuring stick, assigning a rank or score to the individual, there must be a characteristic or dimension in which this variation takes place. The desire for linear measure analogous to those for size, temperature, and reaction time led psychologists to postulate that personality had dimensions or traits. A trait may be defined¹ as a tendency to react in a defined way in response to a defined class of stimuli. Traits are familiar in everyday thinking; nearly all the adjectives which apply to people are descriptions of traits: happy, grouchy, conventional, stubborn, and so on. Traits are elusive in scientific analysis, however, and can be defined and measured only at the risk of some ambiguity.

The postulate that traits exist is based on three facts:

(1) Personalities possess considerable consistency; a person shows the same habitual reactions over a wide range of similar situations.

(2) For any habit, we can find among people a variation of degrees or amounts of this behavior. (3) Personalities have some stability, since the person possessing a certain degree of a trait this year usually shows a similar degree next year.

These facts lead one to consider personality traits as habits, capable of being evoked by a wide range of situations. It would be tedious to list a series of traits such as "habit of bowing politely," etc., to explain the point as it seems pretty well agreed among most authorities that this list is presumably limitless. Any value of the trait approach to personality depends on the hope that it will describe economically the significant variations of behavior, neglecting unduly specific habits. Since the English dictionary offers no less than 17,953 adjectives describing traits, the problem of economy is a real one.²

Guilford's Research on The Dimensions of Temperament

Because the Thurstone Temperament Schedule is to a very great extent based upon Guilford's work on the measurement of temperament, it would be profitable to review the highlights of this research. In his thinking on the subject of the measurement

² Ibid.
of personality Guilford felt that it might be possible to express on a single continuum the entire range of personality characteristics from introversion to extraversion. To test this hypothesis a typical test of introversion-extraversion was set up after a thorough survey of the literature and the then available tests. The test was given 930 students, 277 of whom repeated it after an interval of one month. The frequency with which each one of the thirty-six items was responded to in a certain way in conjunction with a certain response to each other item was obtained. Combinations of reactions were tabulated. From these were obtained coefficients of contingency between pairs of items, which when corrected gave equivalent Pearson coefficients of correlation.

By means of the Spearman-Dodd technique, the correlation of each item with an assumed "g" factor was obtained and also the corresponding weight for each item. The positive and negative weights agreed with the traditional armchair opinions about the same test items, that is, the so-called extravert items and the introvert items as groups tended to cling together at the two ends of a scale. Several items long supposed to be diagnostic, however, had zero weights. All others formed a continuous graded series of traits from extreme introversion to extreme extraversion.

---

A test of introversion-extraversion based upon these weights had a self-correlation test-retest of .81. The validity of the test, as denoted by the correlation of the pool of items with the "g" factor, was .87.

By computing specific correlations with the "g" factor held constant, however, it was found that the test was full of group factors. The weights obtained by the Spearman method were therefore only approximations. Thurstone's method of multiple factor analysis\(^4\) was applied to the table of intercorrelations and at least eighteen group factors were found to be present. Most of these factors were at last reduced to four which were: (1) A tendency to fear the environment, (2) An emotional sensitivity to the environment, (3) Impulsiveness, and (4) Interest in self.

All these except impulsiveness were clearly correlated with the "g" factor that was found by Spearman's method and may therefore be regarded as aspects of introversion-extraversion. Impulsiveness could not be so regarded. It was suggested that this factor is what MacDougall had in mind as introversion-extraversion.\(^5\)

---


Guilford was able to show in this way that it might be possible to force most of the items of a standard test measuring introversion-extraversion onto a single continuum. In reality, however, he concluded that such a process is largely fictitious and that personality is an extremely multi-dimensional affair. It is altogether possible to find several closely allied dimensions including the four listed above and to project them onto a single more inclusive continuum and to give a name to that larger variable thus created. However, it is necessary to keep in mind that, if we are to apply these trait names to any particular test of the usual type, it is fundamentally necessary for us to keep in mind the composition of the test and the weighing of its items. Guilford concluded from this research that the usual scale having the name of a trait does not refer to any real dimension of personality any more than the usual intelligence test measures a single real variable of mental ability.

Since this preliminary study, Thurstone had brought his factor theory and his computational methods of a factor analysis to a high degree of refinement. Guilford thought it worth while to apply these improved methods to the original data in order to determine more accurately the number of common factors which were of consequence in the set of questions and to compute their factor loadings. The results showed that the five seemingly independent
dimensions of personality as revealed by the responses of the subjects to the questions about their likes and dislikes could account for the obtained correlations between the items. These dimensions Guilford tentatively identified as S (Social introversion), E (An emotional stability factor), R (A Rhathymia or happy-go-lucky or carefree factor) and T (A thinking introversion or intellectual leadership factor). 6

A long inventory method to measure factors S, M and E was devised. Items were validated and scoring weights were empirically determined by the correlation of items with extreme criterion groups as established by means of the short inventory. Significant correlations between factors that are by definition independent indicated that greater care must be taken in developing scoring weights for this type of inventory. The reliability of the long questionnaire was assumed to be higher than that for the short inventory but the Spearman-Brown predictive formula was apparently not valid in this connection.

Guilford turned to an investigation of those primary traits which he had previously set aside for more careful research. The R (Rhathymia) and T (Thinking Introversion). 7 A set of


eighty-nine personality questionnaire items were prepared. Thirty of the items were intercorrelated, answers having been obtained from one thousand students. Thurstone's method of factor analysis was applied and nine primary factors were found. Seven of them were identified as: D (Depression), R (Rhythmia), S (Shyness or seclusiveness), T (Habitual Thinking of a Meditative Sort), LT (Liking for Thinking of the Problem Solving Kind) and A (Alertness). The seventh factor was identified with less assurance but it seemed to suggest an immersion in the present environment as an observer. Significant correlations were found to exist between factors D, S and T, (between D and T .66, between D and S .48, and between S and T .53).

These relationships were seen to have a distinct bearing on the question as to what is introversion-extraversion. There would seem to be some basis for lumping together some characteristics bordering on seclusiveness with some implying a thinking person and still others that indicate depressed emotional tendencies and for calling the resultant picture the introvert. Because of the relationship among these three primary traits it is easy to see how a more cursory inspection of personalities would lead to the conviction of a composite trait like introversion. The opposite composite of sociability, cheerfulness and lack of meditative thinking would of course be the extravert.
In addition to this work, Guilford entered upon an investigation of Freeman's theory that individual differences in reactivity of nervous systems furnish the physiological basis for some important differences in personality. A general dimension of hyperactivity-hypoactivity was assumed. An inventory of one hundred items was set up with many of the items aimed at bringing out differences in hyperactivity. After the inventory was administered to six hundred students, twenty-four of the most representative items were selected for a factor analysis. The analysis showed not one dimension of hypoactivity-hyperactivity but at least four. Two of these were clearly identified as factors N (Nervousness or Jumpiness), and factor GD (General Drive, characterized chiefly by a pressure toward action). The two others could not be identified. A fifth factor seemed to imply a variety-loving trait but it was not clearly defined.

Guilford maintained that an important prerequisite to the correlation between traits of behavior and physiological traits is a knowledge of the primary dimensions of behavior characteristics and that the factor analysis method is an important tool in this approach.

The Contribution of Constance Lovell

Lovell became interested in this work of Guilford's

8 Ibid., 238-248.
and especially in the three personality inventories he had constructed and turned to a study of the thirteen variables of personality measured in these tests. The inventories were: (1) The Guilford-Martin Inventory of Factors G, A, M, I, N, (2) Guilford's Inventory of Factors S, T, D, C, R, and (3) The Guilford-Martin Personality I.

These three inventories provide measures of the following factors, the derivation of most of which were explained above: S (Social Introversion-Extraversion), T (Thinking Introversion-Extraversion), D (Depression), C (Cycloid Disposition), R (Rhathymia), G (General Activity), A (Ascendance-Submission), M (Masculinity-Femininity), I (Inferiority Feelings), N (Nervousness), O (Objectivity), Co (Cooperativeness) and Ag (Agreeableness).

The three inventories were administered to two hundred college students under standard conditions.

Intercorrelations between the scores were then computed and a factor analysis of the results was made using Thurstone's method. Six super-factors were obtained. The first four were identified tentatively as:

1. Drive-Restraint (High factor loadings on general drive, carefreeness, sociability, social ascendance or G, R, S, and A.)

2. Realism (High factor loadings on objectivity, masculinity, freedom from nervousness, freedom from inferiority feelings or O, M, N and I.)

3. Emotionality (High loadings on stability of emotional reactions, freedom from depression, extravertive orientation of the thinking process or D, T and C.)

4. Social Adaptability (High loadings on lack of quarrelsomeness and tolerance or Co and Ag.)

The results of this study present some rather interesting suggestions concerning the structure of personality. On the basis of the findings one might conceive of personality as consisting of hierarchies of habit systems of different degrees of independence and generality. Lovell postulated that the smallest units of the habit systems are tapped by the individual items of the inventories. Many of these are intercorrelated. They fall into clusters because they have in common some general characteristics. These characteristics are not only less specific but are on the average more independent of each other. Such are the thirteen factors measured by the three inventories listed above. These in turn fall into certain clusters because of even more general factors they have in common. These super-factors are more separate from each other on the average than the more general habit systems.

More particularly this study has indicated the following four general habit systems: Drive, Emotionality, Realism and Social Adaptability. Within this structure a person may stand
at any position on the scale for any of these four factors. He might, for example, be high in social adaptability, low in realism, low in emotionality and average in drive. A person with a moderately high score on social adaptability would tend to score high on both tolerance and agreeableness because the two are positively inter-correlated. However, these correlations are low enough so that, in individual cases, there might be considerable disparity between standings on the two.

Thurstone's Research on Dimensions of Temperament

Doctor Thurstone, whose studies in factor analysis had made much of the previous work possible in the study of temperament, undertook to determine the number of factors or dimensions that are implied in current personality sketches. The work of Guilford, reported above, had resulted in several personality inventories, which we have seen Lovell analyze in an endeavor to determine the variables being measured. Thurstone also turned to these inventories because he felt that they represented careful analytical work and were well fitted to the purpose he had in mind.

The various personality schedules were seen by Thurstone to cover a wide range of personal characteristics, including those which are relatively permanent for each person, as well as those

which change more or less from year to year because of social experience. Most of the scores derived from the Guilford schedules represent relatively permanent characteristics of a person which might be called temperamental traits. Some personality scores, such as appraisals of attitudes on controversial social questions, represent only partly the temperamental characteristics of a person. Such scores also reflect his recent social experience, his social identifications and the propaganda to which he may have been exposed. These are less stable as indicators of temperamental types. Thurstone's interest centered on the non-intellectual traits of personality which are relatively stable, the temperamental types, and which are not often markedly changed in social experience. Hence Thurstone refers to this problem as the dimension of temperament rather than the much larger domain that is called personality.

The correlations between the thirteen scores obtained by the three Guilford inventories were reported, as we have seen, by Lovell who gave all three schedules to two hundred thirteen subjects. She made a factor analysis of the thirteen scores in which the communalities were determined by their intercorrelations. This is the usual procedure but in this study of Thurstone it is pointed out that the original questionnaires which contained many hundreds of items were the basis for these three schedules.
Hence, according to Thurstone, the procedure of Lovell was essentially to investigate the second-order domain in the thirteen factor scores. This is an interesting and important problem. The second order domain in the traits of temperament may be psychologically revealing. But before undertaking such a study it would be preferable to make sure that the factor scores which enter the second-order analysis are linearly independent.

In this study Thurstone directed himself to the main problem, namely to determine the number of dimensions as factors in these personality schedules which are represented as thirteen separate scores. Instead of dealing with the thirteen scores as variables, whose common factors are to be ascertained, he sought to learn how many factors are represented in the thirteen scores.

The factor analysis process used by Thurstone reveals that the several personality schedules of Guilford represent a dimensionality of not more than nine linearly independent factors. Since the variance of two of these factors is rather small the actual dimensionality of the thirteen scores is not more than seven independent factors for practical purposes. The seven dimensions of the thirteen scores for which interpretation has been attempted by Thurstone were tentatively named Reflective, Impulsive, Sociable, Active, Dominant, Vigorous and Emotionally Stable. These primary factors were given the symbols R, I, S, A, D, V and
Thurstone started the analysis with the expectation of finding bi-polar factors for all or most of these factors but the result revealed all of them to be positive. In naming the factors an attempt was made to avoid those terms which refer explicitly to the more abnormal aberrations of temperament such as depression or cycloid disposition. Such concepts were seen as referring to the psychiatric extremes but they have correlates in terms of the less severe deviations within the normal range of temperament. When schedules of this kind are used for the description of personality among subjects who are in the normal range it seems preferable to use terms which avoid as far as possible the comparison of the normal subject with the abnormal extremes. Thurstone sees this as good policy in describing the temperaments of normal subjects even though it is recognized that there is no sharp demarcation between the normal and the abnormal in each of the factors or dimensions.
REVIEW OF THE RELATED LITERATURE - PART III

THE THURSTONE TEMPERAMENT SCHEDULE

The Purpose of the Schedule

The inventory\textsuperscript{1} was designed by L. L. Thurstone to describe how normal, well adjusted people differ from each other. It is limited to a practical description of important aspects of temperament and makes no attempt to appraise the degree of conflict, insecurity, or maladjustment. It was designed to assess those traits which are relatively permanent for each person, and excludes those which reflect recent social experience, social identifications, disturbing experiences or exposure to propaganda. However, because of this limitation in depth, the Schedule has unusually broad coverage. Seven areas of temperament are appraised in a relatively short questionnaire.

A Description of the Areas Covered

The seven primary traits measured in the inventory are:

ACTIVE (A): A person scoring high in this area usually works and moves rapidly. He is restless whenever he has to be quiet. He

\footnotetext{1}{Examiner Manual for the Thurstone Temperament Schedule, Science Research Associates, Chicago, 1953.}
likes to be "on the go" and tends to hurry. He usually speaks, walks, writes, drives, and works rapidly, even when these activities do not demand speed.

VIGOROUS (V): A person with a high score in this area participates in physical sports, work requiring the use of his hands and the use of tools, and outdoor occupations. The area emphasizes physical activity using large muscle groups and great expenditure of energy. This trait is often described as "masculine" but many women and girls will score high in this area.

IMPULSIVE (I): High scores in this category indicate a happy-go-lucky, daredevil, carefree, acting-on-the-spur-of-the-moment disposition. The person makes decisions quickly, enjoys competition, and changes easily from one task to another. The decision to act or change is quick regardless of whether the person moves slowly or rapidly (Active) or enjoys or dislikes strenuous projects (Vigorous). A person who doggedly "hangs on" when acting or thinking is typically low in this area.

DOMINANT (D): People scoring high on this factor think of themselves as leaders, capable of taking initiative and responsibility. They are not domineering, even though they have leadership ability. They enjoy public speaking, organization of social activities, promotion of new projects, and persuading others. They are the ones who would probably take charge of the situation in case of
accident.

STABLE (E) (For Emotionally Stable): Persons who have high scores in this area usually are cheerful and have an even disposition. They can relax in a noisy room, and they can remain calm in a crisis. They claim that they can disregard distractions while studying. They are not irritated if interrupted when concentrating, and they do not fret about daily chores. They are not annoyed by leaving a task unfinished or by having to finish it by a deadline.

SOCIABLE (S): Persons with high scores in this area enjoy the company of others, make friends easily, and are sympathetic, cooperative, and agreeable in their relations with people. Strangers readily tell them about personal problems.

REFLECTIVE (R): High scores in this area indicate that a person likes meditative and reflective thinking and enjoys dealing with theoretical rather than practical problems. Self-examination is characteristic of this type of person. These people are usually quiet, work alone, and enjoy work that requires accuracy and fine detail. They often take on more than they can finish, and they would rather plan a job than carry it out.

Format of the Thurstone Temperament Schedule

The one hundred forty items covering these seven areas are printed in a six-page step-down booklet. This booklet can be used either with a self-scoring carbon answer pad or...
machine-scored answer sheet. Both of these have adult and boy and girl profiles printed on them. The self-scoring answer pad has the unique feature of only indicating the correct answers. Since none of the unscored responses appear on the grid page, this pad is even easier to score than the usual carbon answer pad.

The items used in the test are simple questions relating to the various traits indirectly by probing the subject's interests and inclinations in each area. Examples of the test items are: "Do you often tell stories to entertain others?", "Do you take the initiative in planning for a party?", "Are you often in a hurry?", "Can you work under distracting conditions?".

Administration of the Test

The Schedule is self-administering. It may be given with or without supervision, in a group or individually. The Examiner Manual directs that the examiner giving the test in group form instruct the subjects to read the instructions printed on the first page of the Schedule carefully. These instructions may be read aloud to the group at the discretion of the examiner. The examiner is directed to make no comments about the items in the test before or after the test is started, other than to remind the subjects to answer each item by marking one alternative: Yes, ?, or No. There is no time limit. Sufficient time must be
allowed for everyone to complete all the items. Twenty minutes is usually adequate. As each individual finishes, the examiner is directed to see that the subject has answered every question. If a question is found to be unanswered the subject should be instructed to go back over the Schedule and answer every question.

Plotting the Individual Profile

Profile charts in terms of percentile ranks are printed on the answer pads and answer sheets. There is one profile for adult men and women and another for high school boys and girls. In plotting the profile the following steps must be followed:

1. If the subject is an adult man or woman, the adult profile is used. If the subject is a high school student, the boy or girl profile chart is used.

2. In the spaces at the top of the profile chart the scorer copies the scores for "A" through "R" from the answer pad or answer sheet.

3. Under "A" the number which is entered at the space at the top is found. The scorer should use the numbers under "M" if the subject is a man or boy and the numbers under "F" if the subject is a woman or girl.

4. A pencil line should then be drawn through this number.

5. The same thing is done for all the other columns in this way.

6. The percentile rank for each raw score is then read from the scales at the sides of the profile. The transformation from raw scores to percentile ranks is thus made directly from the profile chart.

Interpretation of a Profile

The primary aim of the Thurstone Temperament Schedule
as mentioned earlier, is to evaluate an individual in terms of his relatively permanent temperament traits. One of the values of this inventory is that it helps provide an objective pattern, or profile, of personal traits which can be used to predict probable success or failure in a particular situation. Unlike intelligence requirements for a class of jobs or a general curricular range, personality demands often vary from one job to another, or change from situation to situation in the school environment. Validation studies of temperament as it relates to general job classifications and to school situations are in progress. Pending these results the inventory is useful when the personality criteria on the basis of experience or careful analysis have been established.

Job conditions may vary because of factors which are often not listed in a job analysis. For example, the personality of the supervisor may be extremely important. Moreover, some factors are of an uncontrollable nature, such as amount of work, space, privacy or lack of it, noise and disturbances or the pressure of the work. These conditions may contribute to the successful job adjustment of the worker or to his failure on the job. Listing all the personality factors for a specific position in relation to each temperament trait of the Thurstone Temperament Schedule is a valuable step toward rounding out the complete job
requirement.

Here is an example of the type of desirable characteristics listed in a typical job analysis:

1. Rapid work for long periods
2. Concentration amidst distractions
3. Emotional stability
4. Ability to meet and deal with the public
5. Some adaptability
6. Few decisions and little planning

It can be seen that, in this case, this inventory profile of a likely applicant for a position requiring the above traits probably should show relatively high scores on Active, Stable, and Sociable, and an average score on Impulsive. His score on Dominant need not be beyond average (unless future plans call for increased responsibility on the job). If the job involves machine or manual work, the Vigorous score should be high. The score on Reflective should probably be low average to low.

When using this inventory, some standard is necessary to decide whether a score is high or low. Since no test is perfectly reliable, it should be remembered that there is a middle range of scores that should be considered average. For example, on even the most reliable test, the fifty-fifth percentile is not far enough above the median to be considered high.
The appropriateness of any standard setup depends on the purpose for which the test is being used and the reliability of the scores. The Thurstone Temperament Schedule was designed to provide a quick survey of seven temperament areas. Since the area tests must be very short to permit rapid administration, they cannot have the reliability of a long test for each area. Consequently, the Schedule is most valuable when you want to determine what conspicuous differences in temperament a person has.

**Norms for Men and Women, Boys and Girls**

The frequency distributions of the scores in the seven areas for 694 freshmen and 181 freshmen women attending the University of Illinois appear in Table 2 of the Examiner manual. The frequency of distribution for high school boys and girls is presented in Table 4 of the manual. The group in this case consisted of 419 boys and 504 girls from representative Chicago High Schools. The range in age was from fourteen to eighteen years. Table 6 of the manual shows the frequency of distribution for adults. Of these most of the subjects were high school graduates with only fifteen percent college graduates. There were 540 men and 496 women. All types of office work were included in the sample which included office boys and girls, stenographers, typists, receptionists, file clerks, messengers, statistical clerks,

2 Ibid., 4-6.
finance clerks, bookkeepers, machine operators, and coordinators. The age range was from twenty to fifty-five but the majority range from twenty-five to forty-five.

The Reliabilities for the Seven Areas

The reliabilities for the seven areas of the Schedule have been computed by the split-half method for the four groups: men, women, high school boys and high school girls. The odd-even correlations were computed and then reliabilities were estimated by the Spearman-Brown correction for double length. Reliabilities were also calculated for a fifth group which comprised men seeking counseling at the University of Chicago (Number of cases: 106). This group was considered as more representative of people using the inventory for advisement or employment purposes.

The reliabilities for the seven areas of the Schedule were also computed by the test-retest method for a sixth group which consisted of eighty-one male executives at Sears, Roebuck and Company. The retests were all given within six months of the first administration, but the time between administrations varied for different subjects. The reliability coefficients are higher in the test-retest group than in the split-half group. All the data on reliabilities are summarized in Tables I and II, as follows:
### TABLE I

**RELIABILITY COEFFICIENTS FOR THE SEVEN AREAS**

<table>
<thead>
<tr>
<th>Trait</th>
<th>Men</th>
<th>Women</th>
<th>Boys</th>
<th>Girls</th>
<th>Guidance Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>.48</td>
<td>.46</td>
<td>.48</td>
<td>.50</td>
<td>.51</td>
</tr>
<tr>
<td>Vigorous</td>
<td>.61</td>
<td>.63</td>
<td>.59</td>
<td>.67</td>
<td>.68</td>
</tr>
<tr>
<td>Impulsive</td>
<td>.65</td>
<td>.63</td>
<td>.62</td>
<td>.70</td>
<td>.70</td>
</tr>
<tr>
<td>Dominant</td>
<td>.77</td>
<td>.77</td>
<td>.82</td>
<td>.83</td>
<td>.86</td>
</tr>
<tr>
<td>Stable</td>
<td>.63</td>
<td>.64</td>
<td>.59</td>
<td>.63</td>
<td>.55</td>
</tr>
<tr>
<td>Sociable</td>
<td>.68</td>
<td>.73</td>
<td>.69</td>
<td>.78</td>
<td>.76</td>
</tr>
<tr>
<td>Reflective</td>
<td>.73</td>
<td>.62</td>
<td>.60</td>
<td>.48</td>
<td>.45</td>
</tr>
</tbody>
</table>

| N         | 200 | 157   | 236  | 277   | 106            |

### TABLE II

**TEST-RETEST RELIABILITY COEFFICIENT, MEANS AND STANDARD DEVIATIONS FOR EIGHTY-ONE MALE EXECUTIVES (1st AND 2nd ADMINISTRATION)**

<table>
<thead>
<tr>
<th>Active</th>
<th>Vigorous</th>
<th>Impulsive</th>
<th>Dominant</th>
<th>Stable</th>
<th>Sociable</th>
<th>Reflective</th>
</tr>
</thead>
<tbody>
<tr>
<td>r.</td>
<td>.78</td>
<td>.78</td>
<td>.79</td>
<td>.82</td>
<td>.61</td>
<td>.73</td>
</tr>
<tr>
<td>M₁</td>
<td>11.73</td>
<td>11.91</td>
<td>12.17</td>
<td>13.59</td>
<td>14.95</td>
<td>13.37</td>
</tr>
<tr>
<td>σ₁</td>
<td>3.41</td>
<td>3.70</td>
<td>3.39</td>
<td>4.64</td>
<td>2.97</td>
<td>3.10</td>
</tr>
<tr>
<td>M₂</td>
<td>12.17</td>
<td>11.85</td>
<td>12.52</td>
<td>13.48</td>
<td>15.68</td>
<td>12.64</td>
</tr>
<tr>
<td>σ₂</td>
<td>3.36</td>
<td>3.62</td>
<td>3.43</td>
<td>4.89</td>
<td>2.81</td>
<td>3.40</td>
</tr>
</tbody>
</table>
validity Studies

An interesting study of instrument validity was conducted with ten groups of sales employees of a national retail company. The supervisor of each group completed a forced choice rating scale for each of his employees. The scale included pairs of items describing the seven traits measured by the Schedule. In each pair, one item described behavior typical of the high-scoring individual and the other item described behavior typical of the low-scoring individual. Supervisors selected the item of each pair that best described the employee. Ratings were compared with the employees' actual test performance. Biserial coefficients of correlation between the ratings and the actual test performance ranged from .81 for Reflective to 1.00 for Impulsive.

The validity of the Thurstone Temperament Schedule has also been studied by using effectiveness of job performance as a criterion. The subjects in the various studies were teachers, office workers, retail store sales employees, sales supervisors, and managers of small retail stores. The procedure, in general, was to compare test scores of groups of employees rated "high" or "good" in performance with scores of employees rated "low" or "poor" and to determine the significance of the differences be-

3 Ibid., 10.
tween the scores of the two groups.

Ryans studied Thurstone Temperament Schedule scores in relation to teacher performance. The subjects were 275 third and fourth grade teachers. Significant differences between good and poor groups of teachers as measured by the five criteria listed below were found at the five per cent level for the Vigorous, Sociable and Dominant traits and at the one per cent level for the Impulsive trait. The Active, Stable and Reflective scores completely failed to discriminate. The criteria used were (1) Pupil Participation, (2) Business-like Attitude, (3) Human Understanding, (4) Teacher's Sociability with Reference to Eight-to-Ten Year Old Children, (5) Total score obtained by summing up the teacher's criterion scores on the first four criteria listed above and an additional criterion score obtained from judgments made on a six-item observation blank relating to pupil behavior.

The results of this study indicate that four scales of the Schedule may help discriminate between teachers who are effective and those who fail in various classroom situations. Additional research, of course, will be necessary to determine exactly how valuable the inventory will be in teacher selection.

Placement and training.

Performance ratings were available for the individuals in the group of office workers on which the norms discussed earlier were determined. On the basis of these ratings, a high group and a low group were selected. The over-all ratings were based on ratings of productivity, efficiency, value on the job, etc., by the agreement of four or more supervisors. The only area which showed any significance at all was the Active and here only at the five per cent level. The failure of the Schedule scores to discriminate between good and poor office workers seems to indicate that the temperament characteristics leading to success or failure vary from job to job.

In a study of 1,274 retail store employees in forty-three stores, each department manager was asked to indicate the individual whom he considered his best employee in terms of sales ability, cooperation, customer service, and general effectiveness. In addition, each manager selected the individual who least successfully met the criteria.

All of the individuals selected in the good and poor groups were given the Thurstone Temperament Schedule. The mean differences and tests of significance between the scores made by the employees rated good and poor for each of the four criteria

5 Examiner Manual, ll.
were obtained.

As was expected, different aspects of the total job demand different qualities of personality for optimum performance. This is reflected in the comparison of the test scores obtained by different criterion groups. Thus, the Impulsive trait effectively distinguished between good and poor groups only on the criterion of sales ability. The Reflective and Sociable trait scores significantly separated the two groups on the customer service and cooperation criteria. The Active factor distinguished between good and poor sales personnel on all criteria, indicating that the characteristic measured by this scale is common to all four aspects of the sales job.

In another validity study 282 supervisors were rated by their supervisors. The over-all rating of general job success included a consideration of the supervisor's ability as a merchant, his ability in performing the technical aspects of his job and his capacity for dealing effectively with his subordinates, peers, and superiors. Forty-five of the supervisors were rated good and forty-one were rated poor. The differences between means on the Temperament Schedule for the two groups disclosed that when criterion ratings concerned operating ability the Vigorous and Dom-
Inant scales significantly differentiated the two groups. When the criterion concerned human relations ability the Vigorous, the Dominant and the Sociable scales discriminated significantly. These findings indicate that all the scales except the Reflective have value for the prediction of effective performance as it was defined for these sales supervisors.

The validity studies described above indicate that the Temperament Schedule can be a valuable instrument in predicting successful performance on many criteria. It is clear, however, that not all scales of the inventory are equally valid for the same criterion. It is necessary for the user of this test to conduct similar validity studies to determine which temperament characteristics contribute most to the on-the-job success in his organization.
CHAPTER III

THE RESEARCH AND ITS RESULTS

The Test Situation

At the West Company, at the present time, courses are being given to all supervisory personnel in managerial procedures and in the psychology of supervision. The course is known as 'Human Relations in Management.' Approximately twenty-five people are in each group and since several groups are taking instruction at the same time approximately one hundred people a month take this course. One of the lectures given in this course is entitled "Individual Differences in the Supervising Situation." The psychologist during this lecture attempts to give the supervisory student some orientation on how human beings differ and why. The methods of measuring individual differences are described and the role played by psychological tests explained. At this point, the psychologist asks the students if they would care to take a test which would better illustrate the concept "individual difference." The type of test given varies from time to time except that some type of personality inventory is usually chosen. The psychologist would normally take several of the tests, score them
for the group and explain how the differences were found and measured.

The Thurstone Temperament Schedule was chosen for the purpose of this thesis and the writer was able to control the testing situation. The tests were given in a quiet room, removed from the distractions and noises of the factory. The subjects were told that this test was not being used for any managerial thinking or for any supervisory evaluation but was merely a part of the lecture program.

No time limits were used, the average time for the completion of the test items by the various groups running to about fifteen minutes. The instructions given in the examiner manual were followed exactly. The subjects were instructed to read the instructions on the first page of the test booklet. All questions regarding the test items during the test itself were referred back to the subject non-directively.

Of the hundred tests given only seventy-eight were accepted as valid by the writer for the purposes of this thesis. Of the remaining twenty-two tests, five were taken by women and seventeen by non-supervisory personnel.

The Sample

The question of whether or not the sample was random is best answered by the fact that of all the seventy-eight subjects,
sixty-three departments were represented throughout the plant. Of the ten departments which were represented more than once, there were never more than two representatives taking the course, and the test. Examining the departments represented in the sample, it was found that a cross-section of skills, trades and job levels was obtained. This was mostly due to chance. The departmental requests for enrollment in this course are honored in the order in which received and this strengthens the random effect.

The Merit Rating

The Personnel Research and Development Department, every six months conducts a rating survey of the entire plant. The ratings are made on I.B.M. cards and are automatically recorded and later used as the basis for job re-evaluation, promotions, rate changes and other employment decisions. The psychologist at this company has had access to this information and indeed from time to time has even been a member of the rating plan committee which investigates various methods of merit ratings for possible application at this company. For the purposes of this research, then, it was only necessary for the writer to obtain from this psychologist the necessary merit ratings on the subjects who had taken the Thurstone Temperament Schedule. The merit ratings that

---

1 Refer to Appendix I, "Sample Instruction Sheet for Raters".
were obtained were "fresh," i.e., were made in the December rate review. The Personnel Research and Development Department cooperated in this to the extent that they allowed this information to be used for the research purposes. However, because of the very confidential nature of the data, they requested that the subjects remain anonymous.

The Treatment of the Data

The plan of the research was to obtain correlations between the merit ratings and the seven primary traits of the Thurstone Temperament Schedule. Means, standard deviations and the significance of the differences of the obtained coefficients would then be computed. This was done. The method used to obtain the coefficients of correlation was Pearson's formula for ungrouped data which proved more applicable to samples of this magnitude.² This data is reported in Table 3 below:

<table>
<thead>
<tr>
<th>Ratings</th>
<th>A</th>
<th>V</th>
<th>I</th>
<th>D</th>
<th>E</th>
<th>S</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>89.2</td>
<td>10.2</td>
<td>11.3</td>
<td>10.6</td>
<td>10.9</td>
<td>12.3</td>
<td>12.7</td>
</tr>
<tr>
<td>r</td>
<td>-</td>
<td>.167</td>
<td>.233</td>
<td>.167</td>
<td>.077</td>
<td>.096</td>
<td>.064</td>
</tr>
<tr>
<td>sign</td>
<td>-</td>
<td>-</td>
<td>5%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

It had been originally planned, once having obtained the Pearsonian coefficients, to calculate, by multiple correlation, beta weights which could be applied to the seven scales for their optimal weighting. It was felt that the weighted test would prove valuable as an instrument of supervisory evaluation, selection and training. The correlations obtained were of such a nature to make this process unprofitable. It is evident that the relationship between the criterion, which in this case is the merit rating, and the Thurstone Temperament Schedule is not of sufficient significance to indicate using this test as an instrument of supervisory evaluation.

An Average Supervisor's Temperament Profile

Although we have not successfully proven any relationship between the job-success criterion and the temperament of this group of supervisors as measured by the Thurstone Temperament Schedule, it is possible for us to obtain from the test results a temperament profile of the average supervisor of this company. Although the applicability of such a profile as this at this company is questionable at the present time due to the low correlations obtained in this study, future research on this problem may possibly indicate the usefulness of such a profile, if drawn up on a more homogeneous group with a more statistically reliable criterion of job success. For this reason the profile
is presented.

Figure 1  Interquartile ranges of the seven temperament traits plotted on the Thurstone Temperament Schedule Profile Graph to portray the comparison between the test group and the standardization population.

In order to obtain a more accurate description of the average supervisor, the author thought it best to take the Thurstone Temperament Schedule Profile Graph and superimpose upon it
the interquartile ranges of the subjects tested. These ranges are represented as the unshaded area of Figure 1 and portray the range of scores of each trait obtained by the middle fifty percent of the sample. The graph constructed and referred to as Figure 1 was taken from the Thurstone Temperament Schedule Answer Pad and care was taken to maintain the correct interval relationships. The descriptions of each area, "Very Low," "Low," etc., are the same as used on the original graph. The numerals at the various intervals under each trait are raw scores.

For selection purposes, a profile of this type may be handled in various ways. The favorable area is clear and the less favorable zones cross-hatched. Decisions can be based upon the number of favorable scores or upon the trend of the individual's curve as compared to the average group. If profiles of this type were constructed for each general supervisory field (engineering, shop maintenance, labor, trades, etc.,) test profiles of supervisory aspirants might be compared to the average profile of the type shown in Figure 1. If for instance, the average profile for engineering supervision was similar to that constructed for this thesis and a prospective supervisor's profile was centered in the very high or the very low range there would be some reason for doubting the possibility of his being a success at the job.
CHAPTER IV

CONCLUSIONS

Summary

In this thesis the possibility of using the Thurstone Temperament Schedule as an aid in supervisory evaluation was investigated. The measurement of supervisory success used as a criterion was the merit rating. The rating procedures used at the West Company and the development of the present methods were described.

The background of the Thurstone Temperament Schedule was then reviewed. Since the test is based to a great extent upon the preliminary research in the measurement of temperament traits by Guilford, his work was reviewed as was the developmental processes which culminated in the Temperament Schedule.

The Thurstone Temperament Schedule was then described. The standardization population, validity and reliability studies were discussed. The administration of the test, the type of items used, the method of plotting a profile and the format of the test were reviewed.

The research itself consisted in taking the test results of seventy-eight supervisors and correlating these results with
the merit ratings of these individuals. The correlations obtained with the exception of the "Vigorous" trait were not significant. The correlation of the "Vigorous" trait with the merit ratings was .23, significant at the five per cent level.

A profile was plotted of the inter-quartile ranges of the sample group for each trait on the profile graph of the Thurstone Temperament Schedule to portray the comparison between the test group and the standardization population.

Interpretation of the Test Data

The immediate conclusion to be drawn from the results obtained from this research is that the Thurstone Temperament Schedule is not a valuable instrument for appraising supervisory worth. The correlations obtained, and reported in Table III of Chapter III, are of the nature which might well indicate chance relationships. The one significant correlation, that of "Vigorous" in itself tells us nothing extraordinary about the sample studied except possibly that the better supervisor is usually prone to indulge in outdoor activities and to prefer hobbies and amusements of a physical nature. The low value obtained for this trait, however, and the nature of its significance, cast doubt upon the value of even this one finding.

An examination of the standard deviations reveals certain important facets of the relationships being considered. It
will be noted from Table III that the Active, the Vigorous, the Impulsive and the Reflective traits show standard deviations of 6.66, 6.42, 6.61, and 6.30, respectively. The mean score in each instance is 10.2, 11.3, 10.6, and 9.1, respectively. If we were to plot a distribution chart of these results and indicate the mean for each trait it would be immediately apparent that one standard deviation from the mean in each direction would take in practically the whole range of scores from one to twenty. If we take the other three traits, the Dominant, the Emotionally Stable and the Sociable, the truth of this statement is even more apparent for in these instances the standard deviations are 9.25, 7.65 and 7.15. The only interpretation of this which appears feasible to the writer is that the range of temperament traits obtained in the sample were distributed too widely to permit of ready and reliable measurement, using an instrument of the kind we have chosen. The supervisors in the testing situation were chosen at random and represent a rather good cross section of the industrial society at this plant. The range of activities covered in the group is indeed wide. The supervisors represent engineers, tool and die makers, plant maintenance personnel, office workers, bench hands, technicians, laborers, guards, personnel workers, time study people, piece-rate personnel, packers. It is apparent from this list that the supervisory demands are as different as
the trades and skills represented. It is no wonder then that we find a group of supervisors displaying so wide a range of temperament dimensionality.

In regard to the criterion chosen, the merit rating, it may be said that the method used at this company is regarded as successful and is held in high esteem. The individual department chiefs, section chiefs and division managers, in rating their subordinates, are guided by the suggestions listed in Appendix I to this thesis. It may be seen from this list that, if anything, the directions are to a large extent non-directive. This is not to be taken as a failing of the system for this, indeed, it is not. The merit rating system presently in use was designated to apply in the universal situation. This means that it cannot be too explicit in each job situation.

The writer has given the merit rating system used at this company considerable study and thought. After working through the research reported here it appears to him that there is yet a second possible cause of the correlations obtained. In the various departments throughout the plant, the merit rating means different things. The only universal concept which is seen to apply in the different situations is the fact that a supervisor is rated on how well he does his work and how well he handles his subordinates. The universal concept is reflected in the very
directions of the suggestion sheet used in the rating procedure. However, in the particular situation, each executive rates his men on their performance. Now some jobs demand one type of personality and some another. An engineering supervisor must certainly utilize much more tact and sociability in dealing with his people than a jitney driver supervisor.

A supervisor of a labor gang should, to say the least, be dominant and vigorous, and perhaps even be found to be a little impulsive. The head of the design and research organization will no doubt be somewhat reflective and emotionally stable and perhaps a little unsociable, for in his work the devotion to exactness and precision is obtained only through some compulsiveness and refined constriction.

In rating these different people their superiors are guided by the performance on the job, the ability shown in carrying out the given assignments, and the way they work under various conditions of stress, etc. In each situation mentioned, however, it can be seen that a different temperament profile is necessary for supervisory success. This would explain the wide variation of the test results and the low correlations.

There is still another possible cause of these wide variations and low correlations. It has been mentioned earlier that no adequate validity study has been completed in regard to
the merit rating at this company. Only after adequate studies have been completed in this regard may the possibility that this criterion is invalid be disregarded. There is no doubt but that an invalid criterion could explain the absence of positive findings.

A word must be said regarding the comparison between the standardization population and the research sample. An examination of the Examiner's Manual revealed that no factory workers, bench hands or industrial employees of any kind were included in the standardization process. It has been mentioned earlier that this population consisted of college students, high school students and office workers. This is a serious limitation in the test itself and very likely might have resulted in invalidating the test on this research group.

The Direction of Future Research

As a result of this project several possibilities for future research along the lines of this study present themselves. The necessity of an adequate validity study of the merit rating has already been mentioned. The need for further research on the reliability of this criterion should not be disregarded. In relation to the Thurstone Temperament Schedule itself, it would be profitable to conduct a study similar to the present research on a homogeneous group at this company. Such a study would ex-
plore the possibility that diversity of job demands resulted in the failure of the present research. If several such studies were completed in different job categories it would be revealing to discover whether different personality traits are associated with job success in different industrial applications. Finally, it should be evident that the Thurstone Temperament Schedule must be validated on an industrial population if this test is to be used for managerial thinking in the industrial world.
HISTOIOGRAPHY

I. PRIMARY SOURCES

A. BOOKS


B. ARTICLES


C. TEST MATERIALS


APPENDIX I

SAMPLE INSTRUCTION SHEET

EMPLOYEE RATING SURVEY

Following these suggestions will help to make your ratings more valuable:

1. Register your best judgments, regardless of the opinions others may have of the employee.

2. When evaluating the job ability of an employee, look for evidence in the year just past.

3. Don't place a person out of his proper ranking "to give him a break." Doing so always injures other employees.

4. Avoid being unduly influenced by small matters or events of recent occurrence.

5. Don't judge an employee's ability on the basis of nationality, color, religion, political beliefs, or contacts outside the plant.

6. To rate on job ability, keep in mind the quality of the work and the amount turned out, the effectiveness of the employee in learning and applying what is taught, his cooperation in group matters such as safety and housekeeping, his ability to work under extra load, his ability to handle a variety of assignments and his ability to get along with associates.

7. Omit the rating of any employee you are not qualified to rate. A qualified rater is one who has had first-hand knowledge of the employee's work during the year prior to the rating. On the other hand, do not omit a rating because the position given the employee would be low.
The thesis submitted by Francis Xavier Paone 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.

May 24, 1954

Edmund P. Marx
Signature of Adviser