THE CONSTRUCTION OF A RATING SCALE FOR CLERICAL WORKERS

A Thesis Presented for the Degree of Master of Arts

by

Samuel T. Beacham, B.A

OHIO STATE UNIVERSITY

The Ohio State University 1947

Approved by
CONTENTS

CHAPTER I
THE PERSONNEL PROBLEM .................................. 1

CHAPTER II
A SHORT HISTORY OF RATING SCALES .................. 7

CHAPTER III
THE PLANNING AND CONSTRUCTION OF THE SCALE ....... 19

CHAPTER IV
THE TRAINING OF THE RATERS, THE RATING
PROCEDURE, AND THE RESULTS OBTAINED ............... 32

CHAPTER V
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS .... 46

BIBLIOGRAPHY ............................................. 52

APPENDIX A
SCALE 1 .................................................. 55
LIST OF TABLES AND FIGURES

TABLE 1: PAGE
TRAITS SUBMITTED BY EIGHT DEPARTMENT HEADS AND THE PERSONNEL MANAGER OF AN INSURANCE COMPANY AS BEING SIGNIFICANT IN CLERICAL WORKERS AND THE FREQUENCY WITH WHICH EACH TRAIT OCCURRED AMONG THE NINE LISTS .... 24

TABLE 2: PAGE
POINTS ASSIGNED SIXTEEN TRAITS OUT OF 200 POINTS SPENT BY EACH OF EIGHT DEPARTMENT HEADS, THE PERSONNEL MANAGER, AND THE WRITER; AND THE FINAL WEIGHTS USED IN SCALE I IN THE ORDER OF THEIR APPEARANCE ON THE SCALE ..................... 27

FIGURE I: PAGE
DISTRIBUTION OF WEIGHTED SCORES ACCORDING TO THE MEAN OF FOUR RATINGS ON SCALE I ........................................ 37

FIGURE II: PAGE
DISTRIBUTION OF NON-WEIGHTED SCORES ACCORDING TO THE MEAN OF FOUR RATINGS ON SCALE I ........................................ 37

TABLE 3: PAGE
SEVERAL COEFFICIENTS OF CORRELATION COMPUTED FROM FOUR RATINGS FOR EACH OF EIGHTY-SIX CLERICAL WORKERS IN AN INSURANCE COMPANY BY THIRTY-ONE RATERS USING SCALE I ........................................ 39
CHAPTER I

THE PERSONNEL PROBLEM

This study, the construction of a rating scale for clerical workers, was undertaken at the request of the personnel manager of a life insurance company in Columbus, Ohio. Until shortly before this study began, this company had never had a personnel department in the thirty-nine years of its existence. It had, however, personnel problems. The vice president of the company acted as a personnel manager in addition to his other administrative duties and consequently found little time for the planning or handling of a personnel program to meet the needs of an expanding company.

Shortly before the present study was begun, the officers of the company, realizing the need in the company for the serious consideration of such problems as selection of personnel, employee efficiency, and employee morale and turnover, had brought in a new man as Personnel Manager to establish and head a Personnel Department. Coming into a more or less disorganized personnel situation, the new Personnel Manager organized his department and began to plan a long-range personnel program. With the organization of the department, an employment routine was established, the Personnel Manager interviewing and passing recommendations on job applicants to the other department heads concerned. At the same time, a Personnel Committee was formed, composed of the eight other
department heads in the company plus the Personnel Manager who met with them once a week to discuss personnel problems and to formulate personnel policy.

At this point, the writer was approached for aid in devising a more scientific program of employee selection at the home office of the company. The writer then began to meet weekly with the Personnel Committee. It was found that the home office staff was composed of approximately 100 workers, male and female, performing various clerical jobs, such as bookkeeping, filing, typing, and stenography. Employee turnover was felt by the committee to be too high, and it was agreed that in every department there were occupational misfits who constituted an economic loss, not only to the company but to themselves. It was further agreed that the best way to eliminate this problem was not by a wholesale campaign of discharging, but by first differentiating these misfits from satisfactory workers by some objective means. It was pointed out to the committee that personal prejudice is often based on such pseudo-psychological criteria as the physiognomy of the employee, or superstitious or moral beliefs on the part of the judge. ¹ During the several sessions with the committee, the writer found that some of the department heads were basing their opinions of workers on just such unscientific judgments. The only real solution to the problem would have to be a gradual one, with a sound, scientific, employment program replacing

unsatisfactory workers with competent ones.

Such a program, it was hoped, would eventually include both capacity and proficiency tests, application blanks and interview record sheets utilizing the results of this study. However, it was agreed by the committee that before any of this was undertaken, some sort of objective criterion was needed to determine who was a good worker and who was not. Once this was done, these same workers, satisfactory and unsatisfactory, could be given various tests in an attempt to end up with one test or a battery of tests which might be diagnostic in differentiating potentially good and poor workers before they were put on the company rolls. This was the long-range program, but first would have to come some criterion with which to test the tests.

Since the employees under consideration were performing various types of clerical work, objective criteria were difficult to obtain. The workers were not "producing" in the conventional industrial meaning of the term, and thus production records were impractical if not impossible to procure.

It was finally decided that a rating scale of important traits to be filled out by each department head on every one of his employees was the best answer to the problem and the most logical beginning to such a program. This decision was reached only after the writer had presented to and discussed with the committee the various uses to which such a rating scale could be put. It was pointed out that the periodic rating of employees would have the following advantages:
(1) The ratings could be used as a criterion with which to check future employment tests' validity as predictive of an applicant's potentiality.

(2) Ratings could be used to measure traits of employees which at present are non-measurable with psychological tests, but which are, nevertheless, of importance. Such traits as initiative, cooperation, tact and dependability were cited.

(3) Such a scale would tend to substitute for the general impressions, prejudices, and snap-judgments of the raters a systematic rating of the workers trait by trait in a uniform manner.

(4) The proposed scale would also have educational value for the rater and the rated. As Burtt has pointed out, with the periodic use of such a rating, the rater becomes more aware of the existence of the rated traits, looks more closely for them, regards them separately, and, in effect, becomes "personnel conscious." At the same time, the employee being rated realizes that he is being judged on these essential traits. This may encourage some self-analysis, and promote a desire for self-improvement, resulting in a more effective worker.

(5) Rating scale could also serve as a check on the workers' progress. Rapid development might be noted more readily by management through these ratings and promotions planned on the basis of them.

(6) A record of these ratings on file might also provide

---

2 Ibid, pp. 347-49.
a reservoir of information on the capabilities and qualifications of employees to be used in a period of emergency when vacancies occur which must be filled rapidly.

The construction of a rating scale for clerical workers, then, was to be the first step in a long-range program of vocational selection. The present report is confined exclusively to the planning, construction, and experimentation with such a scale, and the conclusions and recommendations drawn from this study.

The organization of this report is as follows: Chapter I has concerned itself with a statement of the problem, an attempted validation of its importance, and the reasons for the course of action decided upon towards the solution of that problem. Chapter II presents a history of some of the work done in the past with rating techniques. It reviews the progress and problems of this method of evaluating personal traits. The contributions of men in the past towards the more scientific achievement of the aims of rating scales are included. Chapter III is concerned with the method of procedure which was followed in the planning and construction of the scale. Chapter IV presents the training of the raters, the actual rating procedure, the treatment of the findings, and the results obtained both as to reliability and validity. Chapter V contains a summary of the results of this study, with conclusions and recommendations based upon it.

The scale about which this report is concerned will henceforth be referred to as Scale I. It is considered to be a preliminary scale, and the conclusions and recommendations drawn
from the experimentation with it will be embodied in a modification of this scale which the company will use permanently.
Efforts to evaluate the individual traits of other people are as old as Man himself. As Bradshaw has pointed out, such attempts are clearly discernible in the ancient writings of the Greek philosophers. ¹ An old physiological theory arose sometime later which held that there were four humors of the body consisting of blood, phlegm, yellow bile and black bile, and these were supposed to determine a person's health and temperament by the varying amounts in which they were present in the body. The use of common words today such as phlegmatic and melancholic in describing a person are relics of this old idea.

The theory behind the earliest characterological studies was the belief that all people are capable of being divided into rather disparate types. Efforts were then made to discover how many different types there were, to describe them, and to show their significance. It is reported that Adolph Quetelet (1796-1874), a Belgian statistician and astronomer, was the first to apply the normal law of error of La Place, an eighteenth century French mathematician, to the distribution of human data, both biological and social. ² Quetelet found that certain anthropometric measurements, such as the height of French Army conscripts were distributed in frequency in accordance with the Gaussian Law, the bell-shaped probability curve of normal distribution formulated by Gauss (1777-1855), a German mathematician.

¹ Francis F. Bradshaw, The American Council On Education Rating Scale, p. 6.
Sir Francis Galton (1822-1911), English scientist and psychologist applied Quetelet's methods of the statistical treatment of human data in experimentation with psychological problems. Galton was a man of many abilities; he performed studies in simple mechanics, the then infant science of meteorology, and psychology, among the most important of which was his exhaustive study of inheritance. In his attempt to demonstrate the inheritance of peculiar eminence, Galton was forced to find some comparable and reliable standard by which to evaluate such eminence. Previously, astronomers at the Royal Observatory had discovered that individual errors in time-observation of astronomical phenomena were grouped in a plainly significant manner around the average error, and might be developed statistically into a normal probability curve of distribution. This, together with Quetelet's observations, led Galton to the assumption that the same might be true of such a social phenomenon as eminence, and he based his degrees of eminence on the normal distribution curve.

Galton thus developed his famous theory of retrogression towards mediocrity in inheritance, and through this came upon the all-important method of the statistical correlation of data. This has been probably one of the biggest single contributions to the validation of psychological observation in the history of psychology.

It is not surprising, then, to find Galton the first to use a rating scale in psychological experimentation. Rating scales

---

3 Ibid, pp. 454-55.
had, however, been used in other fields before Galton. As early as 1805 the British Navy had used a rating scale in estimating wind velocity. The scale consisted of 12 points with 0 representing no wind, and 12 representing wind of hurricane force. The earliest bath thermometers were also calibrated to sense-distances with descriptive terms. The latent idea behind such scales was that each point on the scale should represent equal sense-distances.

Galton used rating scales in the evaluation of the vividness of images. After collecting descriptions of images from many individuals, he produced a rating scale of eight steps or octiles with a ninth step for rare cases by which subjects rated their sensations to various stimuli.

Galton contributed two fundamental and important assumptions to the theory of rating: 1) Personal qualities are distributed among the population according to the frequencies of a normal distribution curve, and equal intervals on the scale should represent equal steps in the frequencies along this normal curve. 2) If standard descriptions of the quality to be measured are arranged in linear order, a rater can give accurate judgments which will be comparable to another rater's rating by matching his own experiences to a term on the scale which appears most similar to it.

The first to apply a rating scale to human abilities was Karl Pearson, who as early as 1906 devised a seven-point scale for the measurement of intelligence. This particular application

---

5 J. P. Guilford, Psychometric Methods, pp. 264-65.
6 Bredehew, op. cit. p. 7.
7 Guilford, op. cit. p. 265.
received considerable impetus when, involved in World War I, the United States Army sought from psychologists some method of measuring the efficiency of its officers.

The leader who came to the fore-front in this work was W. D. Scott, who had been conducting experiments with rating scales at Carnegie Institute of Technology. Scott originated the Man-To-Man Scale, extensively used by the Army, and later adapted by the Scott Company for use in industry. This scale was an important milestone in the history of rating scales. It rated officers on five traits: Physical Qualities, Intelligence, Leadership, Personal Qualities, and General Value to the service. There were five steps in the scale from highest to lowest, each step assigned a numerical value. To use the scale the rater was first required to construct a master scale from which to work, inserting the names of officers of his acquaintance whom he thought to be most appropriate to each of the five steps. He then was able to rate people by comparing them with the names on the master scale.

While the Scott Man-To-Man Scale was still being used extensively, J. B. Miner published in 1917 a report of another and different kind of scale also constructed at Carnegie Institute of Technology. It had been built to be used by the placement bureau of that school in placing seniors upon graduation. This scale is important in the history of rating scales because of several of its features. It not only followed Galton's theory of frequencies, but introduced several new.

---

innovations:

(1) The standards by which the ratees were to be judged were well-known by the judges and clearly defined. (In this scale the standard by which the ratees were to be judged was the average senior class in the ratee's course and college.

(2) The rater was required only to place a dot on a line in ratings and was thus permitted to judge as finely as he desired.

(3) The population was theoretically divided into fifths, qualitative descriptive terms thus being avoided.

(4) The units of measure were easily transmutable into equivalent units of Standard Deviation on the basis of the distribution of judgments.

Some of Miner's findings with this scale will be discussed in a later chapter of this paper\textsuperscript{10}, but one of his conclusions is pertinent at this point. He found ratings by only one judge to be rather unreliable.

In 1921 and 1922, Rugg published the results of his use of the Scott Men-to-Man Scale in the Army and later in schools. This was a thorough and critical study, and Rugg reported rather disappointing results. He wrote of instances in which certain officers would be rated highest on one rater's scale and lowest by another. He continued:

> Do such illustrations raise serious doubts concerning the validity of rating of human traits on point scales? They prove to me that the task of comparing one person's qualities with another's is fraught with so much difficulty as to be impractical in rating

\textsuperscript{10} Chapter V, p. 46.
the rank and file of persons and for most practical activities of life.11

Rugg, as Miner had done, concluded that a single rating by one person was of little value and called for renewed vigor in the search for some objective measure of personal traits.

With the publication of Rugg's report, interest in rating scales waned for the next few years until Beardsley Ruml, in conjunction with the Scott Company, brought out the graphic rating scale.12 The Scott Company had finally dropped the use of the Man-To-Man Scale as being too cumbersome, requiring more time and effort than the average foreman in industry was willing to put into it. This new rating scale introduced several important new features and embodied some modifications of older features:

(1) Linear representation of a trait was used with descriptive phrases along the line.

(2) A scoring stencil was to be used with the scale so that the rater would not himself know what score he would be giving the ratee in assigning him a certain rating.

(3) The final score on the scale was a letter grade based on the percentage of persons obtaining that score; i.e., a score in the top 10% received a grade of "A", the next lower 20% received a score of "B", the middle 40% a score of "C", the next lower 20% a score of "D", and the lowest 10% a score of "E".13

This appears to have been Galton's scheme, plus the linear conception of Miner, minus any cross-division which might

13 Bradshaw, op. cit. pp. 11-12.
destroy in the rater's mind the feeling of linear dimension of
the quality being rated. On the reverse side of the scale
were clear and full instructions in the proper use of the
scale.

After the pioneering work of Scott and Miner, rating scales
began to find wider application in various fields: industry,
the military, and education. In 1928 Shellow and McCarter
published the results of a rather extensive program of experi­
mentation with rating scales for streetcar motormen in Mil­
waukee. They reported that with a careful program of train­
ing raters through conferences and interviews, and with careful
revision of the scale, the uniformity of the ratings could be
greatly increased and a satisfactory scale developed.

In 1937 the National Youth Administration in Ohio under­
took to evaluate the effects of their work-project program for
out-of-school youth. One of the basic objectives of the
program was to develop in these youth-workers certain work
traits considered by employees to be important. An eleven­
point rating scale was developed in an attempt to measure
improvement in the workers in certain of these traits. Punctu­
ality, mode of dress, and attitude towards authority were some
of the traits measured. This scale was used by supervisors
in rating 241 project-workers throughout northern Ohio. No
actual correlations were reported, but the writer pronounced
the rating scale a satisfactory device with which to measure

\[14\] S. M. Shellow, and W. J. McCarter, "Who Is A Good Motorman?"
Person. J. (1928) VI, pp. 338-43.

\[15\] D. J. McNassor, "Evaluating Youth Work-Project Results",
Occupations, (1937) XVI, pp. 239-44.
improvement in the desired traits.

In 1938 Wilke reported an extensive study in which two experienced judges summarized independently ten to fifteen ratings of each of 548 by ratings on a nine-point scale. A correlation coefficient of .87 was found between the ratings of the two judges.

Thus in the 1930's the rating scale was established as a psychological tool. In 1939 Starr and Greenly reported a survey of sixty-four companies representing the major branches of American industry. Of the forty-four companies replying to the survey, more than one-third were using some form of rating scale. Many of the others, it was reported, were contemplating the use of them. This survey showed rather conclusively that rating scales were performing an important function for American business. In the same year Dunford said:

Periodic objective analytical rating of employees is as essential in modern business and industry as the scientific study of production processes. To effectively use human effort, that effort must be evaluated in the work it is doing.

The delegation to a supervisor carries with it recognized responsibilities. Outstanding among these responsibilities of supervisors is a proper evaluation and use of the human resources entrusted to their care.

During World War II rating scales were used extensively in the various branches of the military and also in war industries. Meier has presented a rating scale to be used in rating

---

candidates for officer-candidate schools. In this scale the possession of officer-like traits in the candidate is checked by the rater under either of two categories, "Yes", or "No". Newman reported, also, the results of an interview rating scale used by trained psychologists in interviewing officer-candidates for the United States Coast Guard Reserve, male and female. The scale is concerned with three factors: the ability to pass the training, the ability to withstand the psychological stresses of military service, and the ability to maintain control under combat conditions. Reliability correlations of .83 for male candidates and .85 for female candidates were reported.

The United States Navy has for many years incorporated a graphic rating scale for personal qualities in the Fitness Reports which are filled out on every officer in the Navy periodically and forwarded to the Bureau of Naval Personnel. The Navy relies to a large extent upon these reports in directing promotions and transfers. During the recent war years when the Navy expanded to many times its peacetime size, these reports were often the only contact between Washington and the men on the ships. In this manner they performed a necessary and invaluable function.

In war industries, rating scales were used much as they had been used previously, but some work was done in trying to

---

19 N. C. Meier, Military Psychology, pp. 358-60.
predict employability, in an attempt to cut down the costly labor turn-over which plagued so many industries during the war. Casety has reported such a study. 22 An eleven point rating scale was developed to predict employability of job applicants, that is, the chances that they would turn out to be a desirable worker. The applicants were rated on factors which, it was presumed, were of some importance. Such factors as age, sex, experience, unemployment record, and method of terminating the previous job were weighted and used in the scale. In a random sample of 206 workers, a validity coefficient of correlation of .72 ≠ .02 between ratings and number of months employed was reported.

During these recent years, rating scales for personnel have been adopted by many state and municipal civil service systems. In publishing the development of the "San Francisco" system, Smith reports that the Civil Service Commission of the city of San Francisco conducted a comprehensive survey of the principal municipal personnel systems in the United States and Canada, before installing its own system. 23 Outstanding among systems using rating scales successfully were the Los Angeles City Schools, The Home Owners Loan Corporation, the cities of Saint Louis and Detroit, and the states of Minnesota and Alabama. Smith concludes that the following were found to be general requisites for a satisfactory rating scale, and that they were used as a guide in setting up a

personnel rating system for employees for the city of San Francisco:

(1) The raters must be sold on the effectiveness of rating scales.

(2) Employees must be sold also on the fairness of the rating method of evaluation.

(3) Supervisors who must do the ratings must be trained in the proper use of scales.

(4) The rater should take the ratee into his confidence if the ratings are to provide any self-improvement by the ratees.

In the long and vicissitudinous history of rating scales, much has been done to improve the technique. Much remains to be done, however, for many problems still exist. New and original attacks are being made on these problems, one of which is the application of the "forced-choice" technique to rating scales.24 It has been pointed out that much of the utility of rating scales for predictive purposes has been lost because the rater is able to tell where he is placing a person on the scale. Some work has been done with this "forced-choice technique" which reduces the rater's ability to control the final results of his rating. It is difficult to construct; but, in essence, it forces the rater to choose between two

descriptive phrases which appear of equal desirability, but which actually have been found to have different and differentiating validity.

This, then, is some of the work which has been done in the past, and one of the newer approaches which will influence the future of this method of handling that very important problem - understanding and evaluating the human personality.
CHAPTER III

THE PLANNING AND CONSTRUCTION OF THE SCALE

As soon as it was decided by the Personnel Committee of this insurance company that the construction of and experiment with a rating scale for clerical workers was to be the immediate project, plans were begun. It was decided that the eight department heads would be responsible for the ratings in each of their departments, but that all scoring, discussions, and interpretation of that data would be performed in the office of the Personnel Manager, and would, of course, remain confidential as far as the ratees were concerned.

In the succeeding meetings of the committee, the writer discussed with the members the theory of rating, its advantages and disadvantages, and some of the precautions which should be observed. It was pointed out that traits which are to be rated should be present in varying degrees, or else some alternatives such as "True" or "False" given.¹ Such traits as honesty or loyalty seem to most people to be absolute quantities, a person either being honest or dishonest, and a continuum between them being difficult to conceive. It was agreed that items for which there existed more accurate data should be excluded from the scale. It would be rather foolish to rely upon a rating of a person's health when a medical examination could give a more objective report on this factor. It was further agreed that a rater should

¹ Burtt, op. cit. pp. 347-402.
be thoroughly familiar with the ratee and with the job before he should attempt to rate. Further sources of error were discussed, but these will be presented in Chapter IV when the training of the raters is discussed.

Next, the concepts of reliability and validity were presented to the committee. While it was felt by the writer that it was unnecessary for the members of the committee to understand these concepts completely, it was believed that later in the study better cooperation would be achieved with them if they had some understanding of the requirements and problems of such a study as this. It was explained that the reliability of any measure whether it is a test or rating scale, or just an ordinary ruler, is the extent to which that measure agrees with itself with successive use. Thus, with a rating scale, some idea of the reliability might be obtained by correlating statistically two ratings of the same person by the same rater, or two raters' ratings of the same person.

The validity of a scale was explained to be the extent to which the scale measures what it is designed to measure, and that the validity may be determined statistically by correlating the results of the scale with some objective criterion. It was pointed out that correlations for validity are often difficult to compute because of the difficulty of finding an objective criterion. In some manufacturing industries, units of production might be used as a criterion, but in the present case of clerical workers doing varied jobs objective
units of production were very nearly impossible to procure. One thing could be done, however, to improve both the reliability and validity of the scale. It has been shown that if the ratings of several raters are pooled and an average taken, the results are considerably more reliable and valid than any single rating.\(^2\) Miner, too, has reported in a study that a single rating was very unreliable.\(^3\)

The next step in the planning was to decide the type of scale to be used. The following four basic types of scales were presented to the committee and their respective advantages and disadvantages discussed:

1. The Man-to-Man Scale. This scale, the details of which were discussed in Chapter II, was rejected as being too cumbersome for the average supervisor. Because of the unwillingness of supervisors who are untrained to devote the necessary time and effort in constructing the Master Scale and applying it to their ratings and the resulting unreliability, this type of scale was thought to be undesirable.

2. The Linear Scale. This scale allows traits to be measured along a line which is divided arbitrarily into defined groups, such as "fifths" or percentages of distribution.\(^4\)

\(^3\) J. B. Miner, op. cit. p. 133.
The score is obtained by measuring in millimeters the distance of the rater's check on the line from the left edge of the page. This type of scale was rejected because it was felt that defined groups were not definite enough to establish in the minds of the raters the steps in the scale from the least desirable category to the best. Even among the committee members it was found that such abstract terms as "average" or "highest fifth" did not call to mind comparable trait-levels.

3. The Check List. This type scale consists of a column of descriptive statements which the rater checks according to their applicability to the ratee. While this type of scale has been reported to minimize the "halo effect" found in all ratings, the difficulty comes in determining scale values to attach to the descriptive statements. Because it requires rather involved statistical procedures, and because it was desired to keep before the minds of the raters the idea of the degrees of possession of a trait to aid them in rating, this type of scale was also rejected.

4. The Graphic Rating Scale. This type of scale, also described in the preceding chapter, was finally selected as the best one to fit the present problem. The choice was made in the light of the following advantages of the Graphic Rating Scale which have been summarized by Freyd:

5 Ibid, pp. 375-78.
A. It is simple and easily grasped.
B. It is interesting and requires little motivation.
C. It is quickly filled out.
D. It is easily scored.
E. It frees the rater from direct quantitative terms.
F. The descriptive terms make degrees of traits more concrete.
G. It is universal, (as compared to the Man-to-Man Scale in which the Master Scale may be used only by the rater who constructed it.)
H. Scoring may be altered at will.
I. It allows comparable ratings without requiring all of the raters to know all of the ratees.
J. It is generally more reliable than other scales.
K. It tends to eliminate or reduce the halo effect.

The next step was the selection of the traits to be measured by the scale. The members of the committee were asked to think about the traits which they considered important for their workers to possess, the personality factors which they believed the hypothetical ideal worker should possess. They were asked to submit a list of these traits at the following meeting. TABLE 1 shows the various traits submitted and the number of times each trait appeared in the nine lists. At that meeting, each trait was discussed. As was expected, it was found that the traits meant different things to the different raters, and in many cases a trait which one rater had given one name, another rater had given another. For
TABLE I

TRAITS SUBMITTED BY EIGHT DEPARTMENT HEADS AND THE PERSONNEL MANAGER OF AN INSURANCE COMPANY AS BEING SIGNIFICANT IN CLERICAL WORKERS AND THE FREQUENCY WITH WHICH EACH TRAIT OCCURRED AMONG THE NINE LISTS

<table>
<thead>
<tr>
<th>TRAIT</th>
<th>FREQUENCY</th>
<th>TRAIT</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to learn</td>
<td>1</td>
<td>Judgment</td>
<td>1</td>
</tr>
<tr>
<td>Ability to meet public</td>
<td>1</td>
<td>Knowledge of work</td>
<td>4</td>
</tr>
<tr>
<td>Accuracy</td>
<td>1</td>
<td>Leadership</td>
<td>3</td>
</tr>
<tr>
<td>Adaptability</td>
<td>3</td>
<td>Loyalty</td>
<td>3</td>
</tr>
<tr>
<td>Alertness</td>
<td>3</td>
<td>Monetary responsibility</td>
<td>1</td>
</tr>
<tr>
<td>Ambition</td>
<td>1</td>
<td>Originality</td>
<td>3</td>
</tr>
<tr>
<td>Analytical ability</td>
<td>1</td>
<td>Perseverance</td>
<td>2</td>
</tr>
<tr>
<td>Application</td>
<td>3</td>
<td>Personal appearance</td>
<td>5</td>
</tr>
<tr>
<td>Attitude towards work</td>
<td>1</td>
<td>Personality</td>
<td>5</td>
</tr>
<tr>
<td>Common sense</td>
<td>1</td>
<td>Poise</td>
<td>1</td>
</tr>
<tr>
<td>Cooperation</td>
<td>6</td>
<td>Potentiality</td>
<td>3</td>
</tr>
<tr>
<td>Dependability</td>
<td>3</td>
<td>Progress</td>
<td>1</td>
</tr>
<tr>
<td>Efficiency</td>
<td>6</td>
<td>Punctuality</td>
<td>3</td>
</tr>
<tr>
<td>Enthusiasm</td>
<td>3</td>
<td>Respect for authority</td>
<td>1</td>
</tr>
<tr>
<td>Openness</td>
<td>1</td>
<td>Responsibility</td>
<td>2</td>
</tr>
<tr>
<td>Friendliness</td>
<td>1</td>
<td>Self-confidence</td>
<td>4</td>
</tr>
<tr>
<td>Gossip-monger</td>
<td>1</td>
<td>Sense of humor</td>
<td>2</td>
</tr>
<tr>
<td>Honesty</td>
<td>4</td>
<td>Stamina</td>
<td>1</td>
</tr>
<tr>
<td>Imagination</td>
<td>1</td>
<td>Tact</td>
<td>4</td>
</tr>
<tr>
<td>Initiative</td>
<td>3</td>
<td>Willingness</td>
<td>3</td>
</tr>
</tbody>
</table>

example, it was discovered in the course of the discussions that what one rater had meant by "responsibility", another rater had meant in listing "dependability". To avoid such confusion, the committee went over each of the traits submitted, discussing them so that the meanings would be clear to each rater.

In this manner, many of the traits were either combined, as being repetitive or containing too much overlap, or eliminated as being unnecessary or undesirable in such a scale. For instance, it was concluded that the trait "honesty", while of the utmost importance, should be excluded because should an employee be found to be dishonest it should be brought to the attention of the Personnel Manager immediately as complete justification for releasing that person from the
company. "Ability to meet the public" was discarded because it was found upon further investigation that only two or three employees in the front office came into daily contact with the public, and this item would thus be inapplicable to the majority of the persons being rated. "Monetary responsibility" was likewise dropped from the scale because so few of the employees handled company money. "Sense of humor", "Tact", "Friendliness", and "Tendency to gossip" were combined under the general heading of "Social Attitudes". "Poise" was dropped because it was decided that such a trait was really not of essential importance in a clerical worker. It was agreed that loyalty to the company was an important trait, but that it was one of those traits which it is difficult to conceive in terms of more or less. It was decided to include this trait in the scale with a simple "Yes" or "No" space for the rater to check.

In the preliminary experimentation with Scale I, other supervisors in addition to the department heads would be used to rate the eighty-six clerical workers in order to have three ratings for each worker. It was decided that a punctuality estimate would be included in this scale. These estimates would later be correlated with time-clock records to which the raters had no access to give some idea of the raters' ability to estimate this one factor—a kind of validity check.

By such discussions, the original list of forty traits was reduced to sixteen which would be used in the scale. Efficiency" was later divided into two factors, "Speed" and
"Accuracy", thus making a total of seventeen items on the scale.

The question of scoring these trait-ratings was next considered. Some of the committee members had already brought up the question of what traits were more important than others. It was decided then to use two scoring schemes for scale I to see if weighted scores gave better results than non-weighted scores.

It was decided that the five steps of the scale (See below) would be assigned values beginning with zero at the low end of the scale, ascending one point for each step with the maximum score of four points assigned to the high end of the scale. The non-weighted score would be simply the total of points earned on the seventeen items.

Since weighting the traits would have to be arbitrarily done, it was felt that the department heads who would be in charge of the ratings and who were most familiar with the jobs were the logical ones to determine these weights. They were then each given a typewritten list of the sixteen traits and asked to spend 200 points among the traits, assigning the most points to the trait which they felt to be most important, and so on. They were also requested not to allot points in smaller groups than five points because it was believed that no one's estimates were likely to be more finely discriminating than that. The writer and the Personnel Manager also went through the above procedure, making a total of ten persons who assigned weights to each of the traits. TABLE 2 shows the
weights assigned each trait by each person, the average weight, and the final weight assigned to each trait. As may be seen from the table, this final weight is simply the average weight divided by two, rounded to a whole number. This was done to simplify the scoring computation. In scoring, this weight was multiplied by the non-weighted score of that trait to arrive at the weighted score.

TABLE 2

POINTS ASSIGNED SIXTEEN TRAITS OUT OF 200 POINTS SPENT BY EACH OF EIGHT DEPARTMENT HEADS, THE PERSONNEL MANAGER, AND THE WRITER; THE MEAN NUMBER OF POINTS ASSIGNED; AND THE FINAL WEIGHTS USED IN SCALE I IN THE ORDER OF THEIR APPEARANCE ON THE SCALE

<table>
<thead>
<tr>
<th>TRAITS</th>
<th>Raters</th>
<th>MEAN</th>
<th>FINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSONAL APPEARANCE</td>
<td>A 10 10 10 10 10 05 10 10 10</td>
<td>9.0</td>
<td>4</td>
</tr>
<tr>
<td>EFFICIENCY</td>
<td>20 20 10 15 20 10 20 20 15 15</td>
<td>16.5</td>
<td>8</td>
</tr>
<tr>
<td>COOPERATION</td>
<td>5 20 20 15 25 15 20 20 20 20</td>
<td>17.5</td>
<td>9</td>
</tr>
<tr>
<td>SELF-CONFIDENCE</td>
<td>10 10 5 10 00 10 00 15 15 10</td>
<td>8.0</td>
<td>4</td>
</tr>
<tr>
<td>SOCIAL ATTITUDES</td>
<td>10 10 10 15 10 15 25 20 20</td>
<td>14.5</td>
<td>7</td>
</tr>
<tr>
<td>INITIATIVE</td>
<td>15 20 20 10 20 15 20 20 15 15</td>
<td>16.0</td>
<td>8</td>
</tr>
<tr>
<td>DEPENDABILITY</td>
<td>15 20 15 15 15 10 25 20 10 15</td>
<td>16.0</td>
<td>8</td>
</tr>
<tr>
<td>ADAPTABILITY</td>
<td>10 10 15 15 00 15 20 00 05 10</td>
<td>10.0</td>
<td>5</td>
</tr>
<tr>
<td>JUDGMENT</td>
<td>10 10 15 15 10 20 15 15 20 10</td>
<td>14.0</td>
<td>7</td>
</tr>
<tr>
<td>WILLINGNESS</td>
<td>20 20 15 10 15 15 20 15 15 15</td>
<td>16.0</td>
<td>8</td>
</tr>
<tr>
<td>POTENTIALITY</td>
<td>10 10 10 10 20 10 00 10 5 15</td>
<td>10.0</td>
<td>5</td>
</tr>
<tr>
<td>RESPONSIBILITY</td>
<td>15 00 20 10 00 20 05 20 10 15</td>
<td>11.5</td>
<td>6</td>
</tr>
<tr>
<td>APPLICATION</td>
<td>20 00 10 15 00 20 00 20 00 20 10</td>
<td>9.5</td>
<td>5</td>
</tr>
<tr>
<td>LEADERSHIP</td>
<td>10 10 15 10 20 10 15 10 10 10</td>
<td>12.0</td>
<td>6</td>
</tr>
<tr>
<td>PUNCTUALLY</td>
<td>15 20 00 10 00 10 10 00 5 10</td>
<td>8.0</td>
<td>4</td>
</tr>
<tr>
<td>LOYALTY</td>
<td>10 10 10 20 20 10 15 10 5 5 11.5</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

x: The Personnel Manager
y: The writer

The next problem to be undertaken was the actual construction of the scale. This was done almost entirely by the writer and the Personnel Manager, with the final draft of Scale I submitted to the Personnel Committee for approval.

It was decided to have a scale with five steps for each trait. Various writers have recommended different numbers
of steps in a rating scale. Guilford has recommended three to five steps.\textsuperscript{7} Symonds has reported that the degree of reliability desired should decide the number of steps used.\textsuperscript{8} Symonds suggested that seven steps was the optimum number in rating human traits, for above that the increase in reliability was hardly worth the trouble. He concluded, however, that fewer steps should be used in cases where the raters were untrained, as in the present study. Freyd has advised that the maximum number of steps used should be five.\textsuperscript{9}

In constructing the scale, several concessions were made to the practical aspect of the situation. As Kingsbury has pointed out, ratings should be adopted to the circumstances in which they are to be used, and statistical refinements should sometimes be sacrificed for practical reasons.\textsuperscript{10} Thus, while it was felt that it would be better to have each trait occupy a page by itself, with the rater rating everyone on one trait before going on to the next trait, it was decided that such a method would be cumbersome, more expensive, and more difficult to score. Therefore, the scale was designed to occupy one sheet of paper for each ratee.

Another concession was the decision to break the lines for each step with parentheses in which the raters were to

\textsuperscript{7} Guilford, op. cit. p. 271.
\textsuperscript{8} P. M. Symonds, "On The Loss In Reliability In Ratings Due To Coarseness Of The Scale", J. Exper. Psychol., (1924) VII, pp. 456-61.
\textsuperscript{9} Freyd, op. cit. p. 101.
place a check above the proper descriptive phrase. The fol-
lowing form was used:

I Is this individ-

dual neat in
person and dress? ( ) ( ) ( ) ( ) ( )

Always Generally Pass- Care- Sloven-
well-groomed tidy able less ly

The factors leading to the decision to break the line
with parentheses were: the fact that the raters were un-
tained, and the belief that because of this, they would find
it difficult to think of a trait in more than five degrees,
and finer discrimination would be unnecessary. This form could
also be quickly scored.

In constructing Scale I, which may be found in the Ap-
pendix, the following rules, set forth by Guilford, were fol-
lowed:11

(1) The line should be at least five inches long, but
not much longer, so that it can be easily grasped as a whole.
(The line in Scale I is five and three-quarters inches long.)

(2) The good and bad ends of the scale should be alter-
nated in random order so as to avoid a constant motor ten-
dency to check down one side of the page.

(3) Traits should be introduced by a question to which
the rating will give an answer.

(4) Only universally descriptive terms should be used.

(5) The probable extremities of a trait to be found in
the group to be rated should be decided beforehand and these
used in the scale.

(6) The end-phrases should not be so extreme as to be avoided by the raters.

(7) The average or neutral phrase should be at the center of the line.

At the top of the scale, space was allowed not only for the ratee's name, but also for the ratee's length of service, sex, and marital status. This was done because it is hoped to investigate these factors about employees to see if they have any discriminative or predictive value in employment. However, that is beyond the scope of the present study, and will be done after the present scale has been perfected and in use for some time. Also at the top of the scale was printed the instruction, "In rating this individual on the following traits, please consider them in the light of past rather than expected performance". This was done because the committee and the writer felt this word of caution to be of such importance that the raters should be reminded of it each time they rated an employee.

At the bottom of the scale but separate from it was a line with numbers from one through ten printed under it from left to right. Over it were printed these instructions, "Rate this Employee on the scale below according to his or her general worth to the company". It was desired to get an over-all rating of each employee not only for the purpose of getting the raters' estimate of the person's general worth, but also for purposes of correlation with other parts of the scale. This will be discussed in Chapter IV. This was placed at the
bottom of the scale because, as Bingham has said, "... an overall judgment is more likely to be correct if made after the rater's attention has been focused successively on several of the candidates' specific traits."\(^{12}\)

Scale I was completed, sent to the printers, and the training of the rater's begun.

CHAPTER IV
THE TRAINING OF THE RATERS, THE RATING PROCEDURE, AND THE RESULTS OBTAINED

Because it was desired to obtain at least three ratings on each of the eighty-six employees, it was necessary to bring in raters in addition to the eight department heads. The department heads were asked to recommend supervisors and workers in each of their departments who were capable of rating their fellow-workers. From the names submitted, twenty-two were selected and called into a meeting with the Personnel Committee, the Personnel Manager, and the writer.

At this meeting, a brief description of the new personnel program was given, the function of the rating scale in the program and some of the things that it was expected to do were explained by the Personnel Manager. In this and the subsequent meeting, the writer took the raters through a short training period, giving a brief resume of the theory of ratings, individual differences, and the distribution of scores. Certain rating errors and pitfalls were also pointed out.

In this training program, the following fundamental points were emphasized:

(1) The Halo Effect. This was explained as the result of the rater allowing one or two predominant traits in a ratee to influence all the ratings on that person. There are several reasons for the halo effect being as universal in ratings as it is. Symonds has said that in a habit or trait
which is not easily observed, it is easy for the rater to make assumptions about that trait based on his knowledge of more conspicuous traits, and that halo may be found when traits are not clearly defined on the scale.\(^1\) Knight has reported, too, that the acquaintance factor influences halo.\(^2\) Too-long acquaintance with the ratee may increase the halo effect so much as to completely invalidate the ratings.

(2) The Error of Leniency. This is the tendency in some raters to rate too leniently or to over-rate friends, colleagues, and, it has been pointed out, persons of the same sex.\(^3\)

(3) The Error of Central Tendency. This is the error, common to many raters, to hesitate to give extreme ratings. This tends to displace individuals in the direction of the mean of the group. The raters were urged not to be so much hesitant as careful in assigning ratings at the extreme high or low ends of the scale. As Kingsbury has pointed out, these persons with extremely high or low scores are of the most concern to management; the former representing failure in selection and placement, and the latter indicating the potential promotional material.\(^4\)

(4) The Error of Prejudice. The raters were warned that this is especially dangerous because most people cling to their prejudices, often refusing to admit their existence even to

\(^1\)P. M. Symonds, "Note on Rating", \textit{J. App. Psychol.} (1925) IX, pp. 188-95.


\(^3\) Guilford, op. cit. p. 277.

themselves. Thus, bias against a worker because of race, sex, age, or religion can completely invalidate ratings.

(5) Logical Error. This error arises when raters give similar ratings which seem logically connected in the rater's mind. It is not unlike the "Halo Effect", but arises not from a similarity of qualities in the ratee but a similarity among various traits on the scale in the mind of the rater.

(6) The raters were urged in making the ratings to rate on the basis of past accomplishments rather than on future promise or expected performance.

(7) The raters were also urged to fill out their ratings independently, without comparing them with any other rater.

After the period of training, the raters were asked to submit a list of the clerical workers in their departments whom they felt qualified to rate. They were to be selected principally on the basis of sufficient contact over a sufficient time for observation. No one who had been with the company less than six months was to be rated.

When these lists were turned in several days later, the Personnel Manager, and the Department Heads made up from them a master list of three raters for each of the eighty-six ratees. No one rater's list exactly duplicated another's, but in each case each ratee was rated by his direct supervisor and the Personnel Manager who rated all eighty-six. This was done in order to correlate the direct supervisor's ratings with the

---

5Guilford, op. cit. p. 275.
Personnel Manager's ratings to obtain some idea of the reliability of the scale. The Personnel Manager was selected to rate everybody, because it was felt that he was the one person in the company who was of sufficient acquaintance with all the employees and yet had not been in the company long enough to form intimate friendships. Thus a total of four ratings for every employee was to be obtained.

On the day the ratings were to begin, the raters were each handed a typewritten list of the persons they were to rate, with the necessary number of blank rating scales. No rater knew whom any other rater was rating. The raters were requested to complete their ratings as expeditiously as possible and return them sealed to the office of the Personnel Manager.

As soon as the completed scales were returned, they were scored. For scoring, two stencils were made up. They consisted of the rating scale printed on light cardboard, the parentheses cut out, and the value assigned each alternative printed in ink over the cutouts. The right and left edges of the stencil were trimmed so that when it was placed over a scale, the edges of the scale itself would remain uncovered, and the non-weighted scores could be computed down the right-hand side of the sheet, while the weighted scores were computed down the left-hand side of the sheet. This was done by having the weights to be assigned each trait printed down the left side of the stencil. Question III (Social Attitudes) was scored as if it were one question; that is, the total points achieved on parts (a), (b), (c), and (d) of the question were
Several days after the ratings had been handed in, the lists of ratees were returned to each rater with the instruction to rank each ratee from best to worst in the group. The ranking procedure was delayed until after the rating had been completed in the belief that having rated the employees trait by trait, the raters would have a clearer idea how to rank them, and would not be confused by having to do the two estimates together. The Personnel Manager did not do any ranking because it would have been necessary for him to rank eighty-six persons. This was thought to be impracticable. As Knight has reported, experience has shown that ranking large numbers of persons is both irksome and confusing. 6

In handling the data, the first step after scoring was the construction of distribution charts. It was desired to ascertain if the ratings distributed normally. Figure I and Figure II show the distributions of the mean of the four ratings for each ratee, the Personnel Manager's and the direct supervisors' ratings for weighted and non-weighted scores, respectively. It should be noted that the distributions for both methods of scoring are essentially normal. The weighted scores seem to be negatively skewed more than the non-weighted scores. Such a skewed distribution might be desired should the scale ever be used in eliminating the poorer workers in the company, since such a distribution differentiates better among

6 Knight, op. cit. p. 90.
FIGURE I
DISTRIBUTION OF WEIGHTED SCORES ACCORDING TO THE MEAN OF FOUR RATINGS ON SCALE I.

FIGURE II
DISTRIBUTION OF NON-WEIGHTED SCORES ACCORDING TO THE MEAN OF FOUR RATINGS ON SCALE I.
the lower scores. A critical score might arbitrarily be set below which workers would be released from the company. However, since at the present time no such use of the scale is contemplated, the non-weighted scores probably give the desired distribution.

It may be noted also that the direct supervisors' ratings give a more normal distribution than the Personnel Manager's ratings. This is perhaps because the direct supervisors in each case were in closer daily contact with the ratee and were therefore better qualified to rate than the other three raters.

In further handling of the data, the scores were coded, the ranks converted to Standard Scores so as to be comparable, and all the data put on punch cards. In computing correlations, it was found that this greatly facilitated the statistical treatment of the data. TABLE 3 shows the various correlations computed.

One of the most important things that it was hoped to learn from this study was which of the methods of scoring, weighted or non-weighted, gave the better results. It may be observed that the coefficient of correlation between the weighted and non-weighted scores is .970. In computing this, the mean of the four ratings for each ratee was taken. This very high correlation reveals one thing: the two methods of scoring give practically the same results. In other words, weighting each item on the scale by the method which was used does not change the ratee's position in the group relative to
**TABLE 3**

SEVERAL COEFFICIENTS OF CORRELATION COMPUTED FROM FOUR RATINGS FOR EACH OF EIGHTY-SIX CLERICAL WORKERS IN AN INSURANCE COMPANY BY THIRTY-ONE RATERS USING SCALE I.

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>CORRELATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-weighted Scores *</td>
<td>....................................</td>
</tr>
<tr>
<td>Weighted Scores*</td>
<td></td>
</tr>
<tr>
<td>Personnel Manager's Non-weighted Scores</td>
<td>...............</td>
</tr>
<tr>
<td>Direct Supervisors' Non-weighted Scores</td>
<td></td>
</tr>
<tr>
<td>Personnel Manager's Weighted Scores</td>
<td>...............</td>
</tr>
<tr>
<td>Direct Supervisors' Weighted Scores</td>
<td></td>
</tr>
<tr>
<td>Non-weighted Scores *</td>
<td>...............</td>
</tr>
<tr>
<td>Overall Worth *</td>
<td></td>
</tr>
<tr>
<td>Weighted Scores *</td>
<td>...............</td>
</tr>
<tr>
<td>Overall Worth *</td>
<td></td>
</tr>
<tr>
<td>Non-weighted Scores *</td>
<td>...............</td>
</tr>
<tr>
<td>Ranks *</td>
<td></td>
</tr>
<tr>
<td>Weighted Scores</td>
<td>...............</td>
</tr>
<tr>
<td>Ranks *</td>
<td></td>
</tr>
<tr>
<td>Overall Worth *</td>
<td>...............</td>
</tr>
<tr>
<td>Ranks *</td>
<td></td>
</tr>
<tr>
<td>Punctuality Estimate *</td>
<td>...............</td>
</tr>
<tr>
<td>Time Clock Records</td>
<td></td>
</tr>
</tbody>
</table>

* the mean computed from four ratings

his position when the items are not weighted. Had the correlations between these two values been low, one could have assumed that the two methods of scoring did not give the same results, and one should have had to look elsewhere to discover which method was giving the better results. As can be seen from the table, the non-weighted scores seem to be somewhat more reliable, and thus are preferable.

A reliability coefficient of correlation between the Personnel Manager's non-weighted scores and the direct supervisors'
non-weighted scores of .664 was found. A similar coefficient of correlation between the Personnel Manager's weighted scores and the direct supervisor's weighted scores was .603. Considering the number of cases, it is believed that the assumption that such a difference in correlation is significant is not untenable. Therefore, the non-weighted scores may be considered the more reliable, and their continued use will be recommended.

It may be further noted from TABLE 3 that in every coefficient of correlation computed the non-weighted scores give higher correlations than the weighted scores. The coefficient of correlation between non-weighted scores and overall worth is .859 and between weighted scores and overall worth .851; between non-weighted scores and ranks .807, and between weighted scores and ranks .772.

Such reliability coefficients of correlation as reported above are somewhat lower than those generally reported in the literature. Bradshaw reported reliabilities between .68 and .93 with the American Council On Education Rating Scale. These were correlations of various traits on the scale between two successive ratings by ten raters. He suggested, however, that reliabilities above .90 are not at all common in the literature, and that they are probably due in this case to the large number of raters and the adequacy of their acquaintance with the ratees. He concluded, "It would seem that the greatest source of unreliability in rating scales lies in the varied opportunities for observation rather than in the scale itself or in

---

7 Bradshaw, op. cit. p. 54.
the processes of judgment."

Symonds has said that .55 is a good average figure for ratings of personality traits, of which this scale is to a considerable extent a measure. Miner reported in one study that two judgments yielded a reliability of .70 while four judgments gave a reliability of .83. Newman has reported in his study of interview ratings for officer candidates by two trained psychologists and a psychiatrist reliabilities of .83 to .85. It should be recalled, however, that these were obtained with trained and experienced raters. Kingsbury has stated that reliable ratings depend upon the intelligence and care of the raters.

With the Scott Company Graphic Scale, reliabilities of .76 between foremen's first and second ratings were found. Reliabilities were also found to increase with use, and different reliabilities were found to be characteristic of some foremen.

The coefficient of correlation between overall worth and ranks is .753. This is somewhat lower than expected. It is possible that this was caused by a misunderstanding on the part of the raters as to just what was wanted, either with

---

10 Newman, op. cit. pp. 103-09.
the estimates of overall worth or with the ranks. Such mis-
understandings could, of course, have reduced this correlation.
Since the non-weighted score correlated higher with overall
worth (.859) then with ranks (.806), it is suggested that
the trouble may be with the ranks. A possible source here
might be the fact that in treating the ranks we were forced into
the assumption that each of the raters' list of ranks theoreti-
cally had the same mean and variability, while such an assump-
tion was not necessary with the overall worth. They were
assumed to have different means and variability. This also
might have lowered the correlation. It is felt that this
correlation might be raised considerably with further training
and practice on the part of the raters.

The correlation coefficient of .691 between the estimate
of punctuality and the actual time clock records was somewhat
higher than expected. It shows that the raters were able to
estimate with considerable accuracy this specific factor. It
is believed that this is some indication that the raters put
time and consideration into their ratings.

Burtt has written, "There is a tendency for the judge to
form a general impression that is favorable or otherwise, and
to rate the person accordingly in a number of traits. This
effect has been called the 'halo effect'." In order to gain
some idea of the possible halo effect occurring in these
ratings, a summary item-analysis was conducted with 100 of the

---

13 Burtt, op. cit. p. 40.
ratings selected at random. The highest ratings were listed according to the number of times they occurred simultaneously in combination of traits on each of the 100 scales. The same was done with low ratings. Every time two or more traits received the same high or low rating on a scale, it was recorded by traits.

Of the scales studied, the average number of times that any combination of two traits were rated simultaneously high or low was 5.5. When any two traits were given similar ratings more than twice this number of times, it was assumed to be a tendency toward halo in the ratings.

The most outstanding relationship between two traits was found between Personal Appearance and Dependability. These two traits were rated similarly twenty-five times in the scales studied. Other indications of halo, though not quite so outstanding, were found between Personal Appearance and Responsibility, and Personal Appearance and Application (to the job). This would seem to indicate that there was a tendency among the raters to allow a ratee's physical appearance to influence their estimate of these other qualities between which any actual relationship is difficult to imagine. This should be brought to the attention of the raters, and the fallacy of any such relationship discussed.

Several other strong relationships were discovered, but these might be considered not so much instances of flagrant halo, as overlap in the traits. For instance, Leadership and Potentiality (for advancement), and Leadership and Efficiency
were found to be rated similarly. The former two traits might be considered somewhat closely related, since the capacity for growth and advancement would almost certainly depend upon leadership qualities; the latter, however, do not seem to be quite so closely related.

Dependability and Application, and Adjustability and Potentiality were found also in a halo-relationship. Here, too, the former two might be considered to be rather closely related, the latter two to a lesser extent. These relationships might have been caused by overlapping in the traits, and should this continue in future ratings, one of each of these pairs of overlapping traits might profitably be dropped from the scale.

Another 100 ratings were chosen at random, and in those with high Overall-Worth ratings, the traits which were rated high were listed. Similarly, in those with a low Overall Worth rating, the traits which were rated low were listed. In this manner it was possible to ascertain those traits which seem to correspond closest with the raters' general impression of the ratee, (the Overall-Worth rating at the bottom of the scale).

The results were not unlike those cited above. In the scales with high Overall-Worth ratings, similar high ratings in Personal Appearance, Dependability, Responsibility, and Application were strikingly apparent. In those with low Overall-Worth ratings, correspondingly low ratings were found between Potentiality and Leadership. These results might be
assumed to indicate that the halo of general impression is more operative in these traits than in other traits on the scale.

With the exception of the halo found around Personal Appearance, it is believed that this error was not outstanding in the ratings. Some of it is probably unavoidable. Bingham has pointed out that some halo is to be desired since it reveals that the rater has not extracted the trait from its setting in the personality pattern. The ratee is thus seen against the proper background - the job he is performing.\(^1^4\) Bingham stated that such halo is in part valid.

\(^{1^4}\) Bingham, op. cit. p. 228.
A rating scale for clerical workers was constructed for an insurance company in Columbus, Ohio. The traits to be rated and the weights to be assigned each trait were selected by the conference method with the Personnel Committee of the company composed of eight Department Heads, and the Personnel Manager.

With this scale, eighty-six clerical workers were rated on seventeen traits by thirty-one raters. These raters were the Department Heads, supervisors and the Personnel Manager. In this manner, four ratings were obtained for each of the eighty-six workers rated.

In scoring the scales, a non-weighted and a weighted system of scoring was used to see which method gave the better results. The conclusions drawn from this report are to be incorporated in a revision of this scale to be used permanently by the company.

The following conclusions have been drawn:

(1) Non-weighted rating scores show nearly normal distributions. This is probably a function of two factors: the ability of the raters to discriminate among the ratees, and the construction of the scale to allow the raters' discrimination to differentiate employees along a normal curve of distribution.

(2) The mean of four ratings give a more nearly normal distribution than the ratings of any single rater.
(3) The coefficient of correlation of .970 between the non-weighted scores and the weighted scores would seem to indicate that one method of scoring was as good as the other. However, the reliability correlation of .664 between the Personnel Manager's non-weighted ratings compared with the correlation of .603 between the Personnel Manager's and the direct supervisors' weighted scores indicates that the non-weighted method of scoring is sufficiently more reliable to warrant its continued use. The reliabilities found are believed to be large enough to justify the use of the scale with no great revision, and these reliabilities may be expected to increase with further training of the raters and use of the scale.

(4) A validity correlation coefficient of .691 was found between an estimate of punctuality and actual time-clock records. This is felt to be rather high, and it is believed that the raters are to be commended upon it, for it indicates a sincere effort on their part to fill out the scale with maximum accuracy.

(5) Question XVII ("Is this individual loyal in his attitude toward this company?") was checked "NO" only three times out of the 344 ratings. This question is obviously not differentiating and should be discarded or replaced.

(6) The scope of this study did not allow an extensive item analysis of the scale. Such a study in the future might prove worthwhile. The summary analysis of the traits for evidence of halo indicated that the raters tended to allow the personal appearance of the ratees to influence their judgments on several rather unrelated traits, such as dependability, responsibility, and application to the job. There was also
some indication from ratings on Overell-Worth that the raters' general impressions of the ratees were being reflected in ratings on some of the specific traits. This was not found to such an extent to be considered invalidating. These results should be pointed out to the raters, however, and the raters should be urged to combat these tendencies in future ratings.

It is felt that since the reliability and validity results were satisfactory in the study made, it is really of little importance of what components the scale is made so long as the total score is acceptable. In view of the fact that the scale at present is composed of the traits thought to be important by the raters, any extensive elimination of items in the scale might bring about doubt and dissatisfaction with it in the minds of the raters. As the scale stands, its length should not prove burdensome since the company is small and no rater will be called upon to rate more than twenty persons.

In the light of the results set forth above, the following recommendations are made to be used in further use or revision of Scale I:

(1) In the future use of this scale or any revision of it, it is recommended that the means of at least three ratings by different raters be taken. It is suggested that the three raters used might be the ratee's Department Head, the direct supervisor, and the Personnel Manager.

(2) In view of the fact that the results of this study have indicated the non-weighted scores to be not only more normally distributed but also more reliable than the weighted scores, it is recommended that the latter system be discarded.
(3) Further training of the raters is urged. It is suggested that the conference method might be employed in discussing with the raters the results of this study in an effort to get them to looking for specific trait-habits in ratees and in general to help them to become "personnel-conscious.

(4) Before the raters do any more ranking, it should be further clarified to them.

(5) It is felt that in the future raters should be required to complete their ratings within forty-eight hours or that some such time limit should be imposed. In the present study some of the raters were lax about this, thus delaying the scoring. It was noted, too, on some of these ratings which were returned late, that ratings had been changed. While rating should be done with care, it is believed that the raters should have performed the necessary observation in the period between ratings and have these well in mind by the time the scales are issued. In this way, ratings may be completed quickly and with little confusion.

(6) It is recommended that Question XVI ("Is this worker punctual?") be eliminated in the revision of the scale. It has served its purpose in the present scale and would be useless in practice since accurate time clock records are available in the office of the Personnel Manager.

(7) In view of the non-differentiating results obtained with Question XVII ("Is this individual loyal in his attitude toward this company?") it is recommended that the question either be discarded, or, since the Department Heads consider
this factor important, something like the following substituted:

How loyal to the company is this person? ( ) ( ) ( ) ( ) ( )

<table>
<thead>
<tr>
<th></th>
<th>Usually</th>
<th>Average</th>
<th>Willingly</th>
<th>Intensely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loyal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usually</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willingly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensely</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>considers Loyalty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inconveniences him-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all self for the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>company</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(8) It is recommended that Question XIII ("Does this worker show responsibility in the stamina and facility with which work is done without constant supervision?") be shortened. Its present verbosity is both confusing and unnecessary. It might be changed to read, "Can this employee do the job without constant supervision?"

(9) It is recommended that the present five alternatives to Question XII ("Does this individual have the capacity for growth and advancement?") be reworded. In the present form, they combine the idea of past growth on the job and expected advancement. An employee may conceivably have shown steady growth (Alternative Four, receiving a score of three points) and still to the rater's belief have reached the limit of growth, (Alternative Two, with a score of one point). The question asks for the rater's estimate of the ratee's future advancement. This should be based, of course, on the observed past performance, but the two should not be combined in the alternatives.

(10) It is suggested for future discussion that fuller benefits in employee self-improvement may be reaped if the results of these periodic ratings are made known to the
ratees. Each ratee could be informed of his rating either by the rater taking him into his confidence, by office mail, or by any similar scheme which the Personnel Committee might care to adopt.

(11) It is suggested that before further work is done in the proposed employment program, the Personnel Committee, and especially top-management give additional thought and discussion to just what they want in a clerical worker. With this well in mind some sort of objective criterion should be selected with which to conduct validity studies upon the revision of this scale and upon any employment tests which may be constructed. It is suggested that some of these might be number of words typed per minute, cards filed per day, or number of mistakes per month in clerical work. Whatever this criterion may be will depend upon what Management wants in clerical workers, and this should first be thoroughly discussed and decided before criteria can be selected, and a plan evolved for their measurement. Such objectives, however, will prove useful in the improvement of this scale, and imperative in the future construction of employment tests.


BIBLIOGRAPHY


BIBLIOGRAPHY


APPENDIX A
<table>
<thead>
<tr>
<th>Trait</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always well-groomed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generally tidy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Careless</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovenly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entirely too slow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Needs speed-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfactory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past worker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exceptional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Too many errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequent inaccuracies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usually accurate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rarely inaccurate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good team leader</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapts quickly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solo performer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant friction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never sure of himself</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upset by deviations from routine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assured in routine instructions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-confidence rarely shaken</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self assured in all situations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friendly with everyone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Makes friends easily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meets friends half way</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somewhat reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hostile attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spreads false and malicious gossip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclined to gossip too much</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takes part in gossip office chatter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rarely gossips</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never gossips</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taciturn in all situations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitive to others' feelings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ordinarily taciturn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rather insensitive to others' feelings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completely taciturn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never sees the joke</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sees only the obvious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usually responsive to the comic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laughs easily</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sees humor in every situation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always wants to be told</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shows the next step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goes ahead only in jobs which have become routine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sometimes shows resourcefulness in non-routine situations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always sees the next step and takes it</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rarely makes good a promise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sometimes undependable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can usually be depended upon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost always makes a good promise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completely dependable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusts quickly and easily to every new situation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shifts from new habits rather well</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meets new situations with reasonable flexibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persists in old habits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unable to adapt to new situations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Un Unable to grasp essentials and reach logical conclusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muddles through</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shows fair judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusions are usually sound and sensible</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exceptional ability in arriving at a wise course of action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XI Does this worker do willingly that which is given to do?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XII Does this individual have the capacity for growth and advancement?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIII Does this worker show responsibility in the stimulus and facility with which work is done without constant supervision?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XIV Does this employee show energetic and sincere application in performance of job?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XV Does this person show ability to direct the work of others with good results?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XVI Is this worker punctual?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XVII Is this individual loyal in his attitude toward this company?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rate this employee on the scale below according to his or her general worth to the company.