Influence of Irrelevant Cues and Alternate Forms of the Graphic Rating Scale on the Halo Effect

1975

William A. Rizzo
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/rtd

University of Central Florida Libraries http://library.ucf.edu

Part of the Psychology Commons

STARS Citation

https://stars.library.ucf.edu/rtd/181

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
INFLUENCE OF IRRELEVANT CUES AND ALTERNATE FORMS OF THE GRAPHIC RATING SCALE ON THE HALO EFFECT

BY

WILLIAM A. RIZZO
B.S., Pennsylvania State University, 1964

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Psychology in the Graduate Studies Program of Florida Technological University

Orlando, Florida
1975
TABLE OF CONTENTS

List of Tables ........................................ iv
Introduction .......................................... 1
Method .............................................. 9
Results ............................................. 13
Discussion .......................................... 19
Appendix A ........................................... 22
Appendix B ........................................... 24
Appendix C ........................................... 27
Appendix D ........................................... 30
Appendix E ........................................... 31
Bibliography ........................................ 33
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Analysis of Variance: Group I</td>
<td>16</td>
</tr>
<tr>
<td>2. Analysis of Variance: Group II</td>
<td>17</td>
</tr>
<tr>
<td>3. Variance Score Differences: Controls Vs. Experimentals</td>
<td>18</td>
</tr>
</tbody>
</table>
INFLUENCE OF IRRELEVANT CUES AND ALTERNATE FORMS OF THE GRAPHIC RATING SCALE ON THE HALO EFFECT

The widespread use of Graphic Rating Scales (GRS) in industry as measures of job performance has focused attention upon problems of their soundness and has lead to numerous research efforts to improve their adequacy. Laboratory and field investigations have identified errors which fall into two general categories: (a) characteristic biases of classes of raters, e.g., men, women, peers, etc., and (b) universal errors, e.g., halo effect, logical errors, central tendency, etc. (Brown, 1968). Comprehensive surveys of the literature regarding rater errors and bias are found in Mahler (1947) and Wherry (1950).

Probably the most common criticism of the GRS has been the tendency of presumably independent scales to correlate highly with each other. Typically, factor-analytic studies of the GRS have reduced multiple scale ratings to two or three orthogonal factors (Grant, 1955; Meyers, 1965). This has raised questions with respect to the ability of the raters to discriminate between performance in various areas of behavior.

This phenomenon was observed and named the "halo effect" by Thorndike in 1920. In field research, he observed that ratings or estimates of a given individual across a number of traits such as intelligence, industry, technical skill, reliability, etc., were very highly and evenly correlated. It appeared likely that raters were unable to differentiate these aspects and rate each characteristic independently.
According to Allport (1947), "The halo has considerable theoretical significance. Its existence is proof positive that in perceiving and reflecting upon a personality we rapidly structure our impressions into a self-consistent totality." The result of the halo effect is to force the rating of a given trait in the direction of the general impression of the individuals rated, thereby reducing the validity of the separate trait ratings. Given that some overall rating of an individual may be valid with respect to some global criterion, such as "overall value to the organization," a fairly sound argument may be made for the elimination of separate scales entirely. This would, of course, nullify any feedback for developmental purposes regarding specific behavior strengths and weaknesses. Here, the underlying assumption is that all ingredient stimuli are in some way criterion related.

Somewhat neglected in the research literature regarding bias is the fact that the stimulus, i.e., the ratee's characteristics and behaviors, may contribute errors in logic as well as stimulus generalization. Newcomb (1931) has noted that such logical errors are similar to, yet somewhat different from halo errors. This error is due to the fact that judges are likely to give similar ratings for traits that seem logically related in the minds of the raters. Like the halo effect, this error increases the intercorrelation of traits, but for a different reason. In the halo effect it is the apparent coherence of qualities in the same individual, whereas in the logical error it is the apparent logical coherence of various traits irrespective of individuals (Guilford, 1954, p. 279). If ratings are assigned based upon a generalized impression, influenced by factors unrelated to the job criteria, the resulting
rating error leads to invalidity. Allport (1947) has pointed out, "... that when individuals are handsome, healthy, neat, and have smiling faces, they are generally judged to be intelligent, although there is very little relationship, or none at all, between these features and intellectual ability." It can be seen, then, that halo and logical errors are closely related as the generalized impression characterizing halo may often be the result of errors in logic.

In theoretical terms, halo has been defined as one type of error component in a linear combination of error sources and true score variance which, when combined, account for the observed score value, or rating (Guilford, 1954; Wherry, 1952). The historical antecedents to such rating theory are found in theories of psychophysics and mental testing (Mosier, 1940; Gulliksen, 1950); Helson, 1947; Bartlett, 1932; Bellows, 1941). Wherry (1952) has synthesized these theoretical foundations to arrive at a theory of the rating process. In its most rudimentary form his equation of the rating process is expressed as: $z_{X_A} = t_A z_T + z_A z_{EA}$, where $z_{X_A}$ is the standard score equivalent of an individual's actual behavior, composed of true ability ($z_T$) and random error ($z_{EA}$). These two components are differentially weighted $t_A$ and $z_A$, where the sum of the weights squared is equal to unity. In its expanded form, his formula includes specific sub-components of these two factors as well as a third term which Bellows (1941) has called contamination due to environmental influences. That is, an individual's performance may be influenced by such variables as prior instruction, tools provided, the work setting, etc.
Wherry has further examined the perception \((X_p)\) of the ratee's behavior \((X_A)\) by the rater. In addition to what actually takes place, the rater is thought to unconsciously add a bias component \((B_p)\). This bias component has three parts:

1. an actual \(T\) component, with weight \(t_p\) indicating an expectancy of ability equal to that actually possessed. A rater who had only relevant contacts would have a relatively high weight for this component. With this component in \(B_p\) we can expect that a rater will give somewhat valid responses even when observing atypical behavior.

2. an areal, non-relevant bias factor, \(B_{PA}\), with weight \(b_{PA}\), which is a kind of residual effect of all previous error and non-relevant experiences aroused by seeing the ratee in stimulus situations which the rater classifies as belonging to a particular area of behavior. Thus a particular bias for or against a given individual as a scholar, for example, differs from the bias held against that same individual as a disciplinarian, or as an organizer, or as a special companion.

3. an overall bias component, \(B_{PO}\), with weight \(b_{PO}\), a residual effect of all possible areal bias effects, pertinent or irrelevant, a kind of background against which all acts of the individual, regardless of stimulus, are evaluated.

The standard score formula is thus expanded to show the weighted contributions of the observed happening \((X_A)\) and the random error of perception \((E_p)\):

\[
z_{XP} = a_p z_{X_A} + t_p z_T + b_{PA} z_{BPA} + b_{PO} z_{BPO} + e_p z_{EP}
\]

In addition to perceptual errors Wherry accounts for errors of recall as well. As in the case of perception, recall is viewed as a composition of true and error variance, to include areal and overall bias.

Guilford (1954, p. 281) has formulated a similar rationale for the expression of errors in the rating process, i.e., a rating obtained from the linear combination of two major components: \(X_{ijk} = X_{ijt} + X_{ijke}\).
The rating of the individual I on trait J by rater K is equal to the sum of the true and error variance. The error term \( X_{ijke} \) can be further broken down to include specific sources of error variance. In expanded form, the formula is:

\[
X_{ijk} = X_{ijt} + X_{ki} + X_{kj} + X_{ijkr}
\]

In this form the rating is expressed as a linear combination of true variance \( X_{ijt} \), leniency error \( X_{ki} \), halo error \( X_{kj} \), the rater's rater-trait interaction error \( X_{ijkr} \), and residual errors \( X_{ijkr} \). In the residual term are such errors as central tendency, logic, and proximity. As these errors apparently do not occur as simple increments, they are not expressed as entities in the linear expression.

These theoretical explanations by Guilford and Wherry are similar in that both attempt to account for the sources of rater bias. The subtle differences in approach to these phenomena will not be discussed at length here, other than to note that both halo and logical errors are viewed in both models as identifiable sources of rater error.

The halo effect has been defined statistically in four ways: (1) as an inflated intercorrelation among traits (Gilinsky, 1947; Taylor and Hastman, 1956), (2) as a general bias factor arrived at through matrix and factor analysis (Grant, 1952; Meyer, 1965), (3) as the rater-ratee interaction error according to Guilford's (1954) analysis-of-variance model, and (4) as a variance score, where halo is defined as inversely proportional to the variance score for ratings given a ratee by a rater, across several traits (Brown, 1968).

Guilford's analysis-of-variance model is perhaps the most frequently quoted definition in contemporary literature as it differentiates between
objective (valid) and relative (invalid) halo. There is some doubt, however, as to the applicability of Guilford's definition of halo in situations where raters have varying amounts of information or contacts with the ratees. Johnson (1963) has questioned the evidence for the existence of halo by concluding that the rater-ratee interaction may be due to objective variations in the information available to the different raters rather than the judgemental process, itself. Given that this is true, it is necessary to test the judgemental process, as Johnson and Vidulich (1956) have suggested, by manipulating the conditions of judgement. In addition, the rater-ratee contacts in the stimulus situation must be held constant if the resulting interaction is to fit Guilford's definition of relative halo.

Brown's variance score method of identifying halo is a statistical translation of Guilford's (1959, p. 146) definition which stated that a particular rater tends to rate a particular ratee similarly on all traits. This definition is unique in that it examines differences in variability across traits as opposed to differences in linearity implied in earlier definitions. This would also appear to be a more forthright approach as the essence of the halo effect, i.e., the spurious similarity of ratings across traits, is the focus of observation, whereas, in Guilford's model, the ratings across traits are collapsed to arrive at the rater-ratee interaction.

Research literature dealing with errors occurring in the use of the GRS is replete with suggested methods of error reduction (Allport, 1947; Aspley, 1944; Beaumont, 1945; Burtt, 1942; Wherry, 1952). A
typical list of adaptations for halo reduction might include the following:

1. Rate one trait at a time, across all ratees rather than rating all traits at once for a given individual.
2. Carefully define variables to be rated.
3. Require rater to justify the rating.
4. Carefully train raters in the pitfalls of rating errors.
5. Stagger the high/low ends of the GRS.
6. Use multiple raters.
7. Use only behaviorally-anchored traits.

Throughout the research literature dealing with reduction of halo is a noticeable absence of inquiry into the sources and isolation of errors resulting from the rater's attention to irrelevant stimuli, i.e., non-pertinent behaviors or personal characteristics of the ratee. Although the existence of this phenomenon is acknowledged in the theoretical treatments of the rating process, little insight has been gained as to how the commonly used GRS might be adapted to counter such bias.

The purpose of this investigation is to manipulate the content and format of the GRS in an effort to reduce the amount of halo error resulting from irrelevant cues. The experimental form of the GRS used contains both job-related and irrelevant personal traits. In a well designed GRS, using only job-related scales, the rater has no option but to contaminate these ratings with bias if he is attending to irrelevant cues. However, if the rater is given the proper vehicle to rate both relevant and irrelevant traits separately, the amount of error variance due to halo occurring
in the ratings of relevant traits should be reduced.

The following hypothesis will be tested in this study: Graphic ratings based solely on job-related, behaviorally-anchored variables will contain greater error variance due to halo than ratings based on similar scales which include, in addition, scales relevant to common non-job-related cues.
METHOD

Subjects

Three groups of subjects were used in this study:

Experimental Group I: Thirty-six undergraduate students in attendance at Florida Technological University and Valencia Community College, Orlando, Florida, comprised this group. These subjects had essentially no experience using a GRS to rate job performance.

Experimental Group II: This group consisted of thirty-six Non-Commissioned Officers (NCO's), stationed at the Naval Training Center, Orlando, Florida. These subjects had from two to twenty years experience rating performance using a GRS. The mean experience for this group was 9.85 years.

The two experimental groups were used to test the hypothesis using both naive and experienced subjects, and to compare the results of the two groups.

Control Group: The control group consisted of twelve undergraduate students at Florida Technological University. These subjects had no experience using a GRS for rating performance.

Instruments

The instruments were designed and adapted to serve the needs of the study, and written instructions preceded each rating exercise.

The ratee information consisted of performance description of two hypothetical workers. The performance descriptions contained information directly related to their job descriptions (Appendix A) as well as non-pertinent information regarding their personal characteristics and
behaviors. The job performance descriptions for both individuals were identical. The irrelevant data included in the description of Ratee 1 (Appendix B) was designed as negative or unfavorable behavior, while the corresponding data for Ratee 2 (Appendix C) consisted of favorable information. Negative information (negative halo) consisted of irrelevant behaviors or personal characteristics of the ratee which are commonly viewed as socially undesirable. The positive information (positive halo) consisted of socially desirable behaviors or virtues. The positive and negative scale values of the irrelevant information were determined by a panel of ten trained raters, prior to the final construction of the ratee performance descriptions. Written descriptions were used in simulating ratee performance to accomplish a time compression of performance over a one-year period and to control the stimulus information available to the raters. Variability across traits was built into the performance descriptions to control for the actual correlation between traits. This was accomplished using a pilot study where nine trained raters evaluated the performance descriptions and actual variability between traits was measured. The trained raters were graduate students in attendance at Florida Technological University who had received extensive training in the pitfalls of rating errors.

Two forms of the GRS were employed. The form used in Method 1 (Appendix D) consisted of six job-related, behaviorally-anchored variables, arrived at through a job analysis. The form used in Method 2 (Appendix E) consisted of the same six variables as in Method 1, plus three scales reflecting the irrelevant characteristics of the two ratees. The rating scales contained bipolar trait descriptions on seven-point scales. The
factors (traits) selected through job analysis were leadership, organization and planning, written communication, resistance to stress, interpersonal skills, and technical knowledge. Factors reflecting irrelevant stimuli were appearance, citizenship, and personal behavior.

**Procedures**

The S's were divided into four groups: A1, A2, B1, and B2. Groups A made ratings according to Method 1 (job-related scales), while Groups B made ratings according to Method 2 (job-related plus irrelevant scales). Groups 1 rated Ratee 1 (negative halo), while Groups 2 rated Ratee 2 (positive halo). Each S rated only one ratee, using one method.

All S's were provided identical job descriptions in addition to the performance descriptions for their assigned ratee. The S's were first instructed to read the job description, then read the performance description of their assigned ratee. The S's were randomly assigned the GRS for Method 1 or Method 2. The same procedure was followed for S's assigned to Ratee 2.

**Statistical Analysis**

The variance score across traits was found for each S's rating. The mean variance score, composed of nine scores per cell, is found in Table 3. The higher the variance score across traits, the smaller was the halo effect, and the lower the score, the greater was the halo effect.

The influence of irrelevant cues and method on the halo effect was evaluated by a two-way Analysis-of-Variance (ANOVA). The basic questions addressed in using the ANOVA were: (1) Is the occurrence of halo
differentially affected by the format of the GRS used (main effect of method)? (2) Is the occurrence of halo effect a function of a method X cue interaction? Or, is halo affected by some combination of halo type (positive or negative) and format of the GRS?
RESULTS

It was hypothesized that ratings based solely on job-related, behaviorally-anchored variables would contain greater error variance due to the halo effect than ratings based on similar scales which included, in addition, scales relevant to common non-job-related cues. The influence of irrelevant cues and rating form on the halo effect was evaluated by a two-way ANOVA. Two experimental groups and one control group were compared.

Cochran's test for homogeneity of variance indicated no significant differences in cell variance for either experimental group. In addition, the following relationships were compared:

1. The ratios of cell variance to cell mean
2. The ratios of cell variance to cell squared mean
3. The ratios of cell standard deviation to cell squared mean

There was no evidence to suggest that the populations from which the samples were drawn were not normally distributed. It was concluded that no data transformations were necessary. Correlations between cell means and cell variances were non-significant. It was concluded that the use of the ANOVA was appropriate, based on these findings.

In the first experimental group (undergraduate students), halo was not significantly reduced by using the long form of the GRS (Table 1). However, in the second experimental group (Navy supervisors) a significant interaction was found ($F=4.7442$, $p < .05$, df=1,32) between rating form and halo type (positive vs. negative) Table 2. A further analysis using Scheffe's method compared the following differences:
1. Positive vs. negative halo using the short form (1)
2. Positive vs. negative halo using the long form (2)
3. GRS 1 vs. GRS 2 for positive halo
4. GRS 1 vs. GRS 2 for negative halo

Consistent with the hypothesis, there was a simple effect of rating form \( F=4.7304, p<.05, \text{df}=1,32 \) as evidenced by the increase in variability across traits under conditions of negative halo, using GRS 2 vs. GRS 1. This trend also was found in the first experimental group, although nonsignificant. There was no significant effect of rating form under conditions of positive halo.

A simple effect of halo type was found \( F=12.5030, p<.01, \text{df}=1,32 \) using GRS 2. The variability across traits was significantly greater under conditions of negative halo than for positive halo. There was no significant effect of halo type using GRS 1. Overall then, for the group of Navy supervisors, it was found that halo was reduced under negative halo conditions, using the long form GRS.

The relative amount of halo induced by irrelevant cues in the experimental groups was determined by comparing the variance scores across traits for these groups with a control group. The control group rated only job-related variables and was not exposed to the irrelevant information. The control group was contrasted with each condition of the experimental groups using an adaptation of the t-statistic (Winer, p. 264). Ratings in both experimental groups had significantly less variability (Table 3) than the control group under conditions of positive halo, using GRS 2. Experimental group 2 also had significantly less variability using GRS 1, under conditions of positive halo. All
other groups were not significantly different from the control group. This led to the conclusion that more halo was induced by the performance description containing positive irrelevant information than the one containing negative information for both experimental groups.

In summary, a significant main effect \((F=6.2287, p<.05, df=1,32)\) of irrelevant cues in the first experimental group led to the conclusion that more halo was present under conditions of positive halo than negative halo (Table 1). The variable of rating form was nonsignificant, although the variability across traits did increase under conditions of negative halo, using GRS 2.

A significant main effect \((F=7.9662, p<.01, df=1,32)\) of irrelevant cues was also found in the second experimental group (Table 2). Again, more halo was evident under the positive halo condition.
### TABLE 1

**ANALYSIS OF VARIANCE: GROUP I**

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratee (C)</td>
<td>5.6289</td>
<td>1</td>
<td>5.6289</td>
<td>6.2287*</td>
</tr>
<tr>
<td>GRS Form (R)</td>
<td>.0371</td>
<td>1</td>
<td>.0371</td>
<td>.0411</td>
</tr>
<tr>
<td>Interaction</td>
<td>1.9639</td>
<td>1</td>
<td>1.9639</td>
<td>2.1732</td>
</tr>
<tr>
<td>Within Group (E)</td>
<td>28.9173</td>
<td>32</td>
<td>.9037</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36.5436</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
## TABLE 2

ANALYSIS OF VARIANCE: GROUP II

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratee (C)</td>
<td>5.7994</td>
<td>1</td>
<td>5.7994</td>
<td>7.9662*</td>
</tr>
<tr>
<td>GRS Form (R)</td>
<td>.5098</td>
<td>1</td>
<td>.5098</td>
<td>.7003</td>
</tr>
<tr>
<td>Interaction</td>
<td>3.4539</td>
<td>1</td>
<td>3.4539</td>
<td>4.7442*</td>
</tr>
<tr>
<td>Within Group (E)</td>
<td>23.2973</td>
<td>32</td>
<td>.7280</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>33.0604</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

**p < .01
TABLE 3

VARIANCE SCORE DIFFERENCES: CONTROLS\(^{(a)}\) VS. EXPERIMENTALS

<table>
<thead>
<tr>
<th></th>
<th>Positive Halo</th>
<th>t</th>
<th>Negative Halo</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRS 1</td>
<td>1.1384</td>
<td>1.3062</td>
<td>1.6406</td>
<td>.5860</td>
</tr>
<tr>
<td>GRS 2</td>
<td>.9147</td>
<td>2.2085*</td>
<td>2.1728</td>
<td>-.6035</td>
</tr>
<tr>
<td>Group II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRS 1</td>
<td>1.1630</td>
<td>1.7694*</td>
<td>1.3462</td>
<td>1.3313</td>
</tr>
<tr>
<td>GRS 2</td>
<td>.7815</td>
<td>2.6819**</td>
<td>2.2037</td>
<td>-.7197</td>
</tr>
</tbody>
</table>

\*p < .05
**p < .01

\(^{(a)}\) Control Group Variance = 1.9028, df=43, K=2
DISCUSSION

The main effect of irrelevant cues found in the first experimental group, and the nonsignificant effect of GRS suggested that perhaps the subjects' (undergraduate students) naivete regarding rating of performance using a GRS, caused inordinate attending to halo-related cues, thus confounding their ability to discriminate between traits, regardless of the format of the GRS. A second experimental group was selected, based on their comparative extensive experience in using this type of scale for rating job performance. The trends exhibited by the second group, however, were almost identical to the first, with one exception. Halo was significantly reduced using GRS 2 under conditions of negative halo. This represented partial confirmation of the hypothesis which stated that graphic ratings based solely on job-related, behaviorally-anchored variables will contain greater error variance due to halo than ratings based on similar scales which include, in addition, scales relevant to common non-job-related cues.

The apparent lack of halo reduction using GRS 2 under conditions of positive halo led to further scrutiny of the performance descriptions, which were designed to induce halo. The scale value differences between relevant and irrelevant traits were less (1.63 scale points) under conditions of positive halo than under the negative halo condition (2.6 scale points). The greater contrast between relevant and irrelevant traits under the negative halo condition apparently facilitated the raters' ability to more clearly discriminate between traits, given the opportunity to rate all traits using GRS 2. This may account for the corresponding reduction of only negative halo, using GRS 2.
Significantly more halo was induced under conditions of positive halo in both experimental groups. This may also be a manifestation of the similarity between relevant and irrelevant scale values in the positive halo performance description. Apparently this similarity served to create a general positive impression of the ratee, thus confounding the raters' ability to discriminate between traits. Although nonsignificant, halo appeared to increase using GRS 2 vs. GRS 1 under the positive halo condition. This trend indicates that when irrelevant and job-related trait values are similar, the addition of rating scales (i.e., GRS 2) serves to further confound the rater's ability to discriminate between traits.

An obvious shortcoming of attempting to simulate the stimulus situation of the rating process concerns the trade off between fidelity and control. It is felt that the laboratory or simulation environment was necessary to control the characteristics of the ratee. In a strictly applied industrial setting it is virtually impossible to present a number of raters with precisely the same degree of experience, exposure, and familiarity with a common ratee. The laboratory setting introduces a certain artificiality with regard to the real-life experiences affecting a rater's propensity for halo error. These experiences are a manifestation of interpersonal contact and information gathering over time, and cannot be simulated with total fidelity. It is a common phenomenon for raters to be more objective and discriminating under laboratory conditions (Taylor and Wherry, 1951). It is strongly suspected that the halo effect is considerably more prevalent in applied settings where raters are constantly exposed to the personal and personality
variations of ratees which may conflict or coincide with their (the rater's) individual values.

The results of this study suggest further investigation of the hypothesis that halo may be reduced by manipulation of the GRS. In repeating the model used in this study, it may be useful to insure the same degree of contrast between relevant and irrelevant traits under conditions of both positive and negative halo. It is suspected that the relative lack of contrast under the positive halo condition inhibited the shift in variability across traits when the long form GRS was used.

It would be useful to test the hypothesis in an applied setting. In such situations, more than one supervisor is rarely in a position to observe the behavior of a given subordinate. If the model of this study is to be followed, i.e., several raters judging the performance of a common ratee, the use of peers to serve as raters may offer a more homogeneous exposure (rater/ratee) than could be accomplished using supervisory personnel.
APPENDIX A

JOB DESCRIPTION

NAME: Ray Barnes

JOB TITLE: Building Maintenance Supervisor

SUPERVISOR: Steve Willis, Facilities Engineer

JOB DESCRIPTION: (summary)

Supervises and coordinates activities of 20 workers engaged in maintaining and repairing utility systems and physical structures of a large office building. Directs workers in making structural repairs to masonry, woodwork, and furnishings. Directs workers in maintenance and repair of utility systems such as electric wiring and controls, heating-ventilating-air conditioning, and plumbing.

JOB DESCRIPTION: (specific)

1. Plans work schedules for three-shift operations.
2. Schedules manpower rotation for all shifts.
3. Assigns job priorities to work orders received from Facilities Engineer (Steve Willis).
4. Gives verbal and written instructions to subordinates.
5. Requisitions tools and materials necessary to perform maintenance functions.
6. Assigns specifications to job orders in accordance with federal, state, and local building and utilities codes.
7. Inspects work in progress and completed work to conformity with blueprints and specifications.

8. Maintains log of work performed, man-hours, and dollar expenditures.

9. Periodically reviews performance of all subordinates and completes performance appraisal forms.

10. Counsels subordinates on sub-standard performance.

11. Interviews prospective employees.

12. Schedules cross-training of subordinates in all phases of building maintenance as work load permits.

13. Writes weekly activities report summarizing work performed, man-hours, expenditures, etc.

SUPERVISORY RESPONSIBILITY:

Directly supervises three crew chiefs in assigning job orders and scheduling of work. Crew chiefs are normally assigned based on seniority and experience in the various maintenance functions. Indirect supervision of plumbers, carpenters, electricians, and heating-ventilating-air conditioning specialists.
Barnes appears to have the general cooperation and following of his subordinates. The men usually follow his work orders and schedules carefully, however, there have been a few examples of work not being completed on time. Overall, the crew chiefs seem to respect and recognize Barnes as their supervisor. He seems to be able to get most jobs done without being overly forceful.

Although Barnes is in a supervisory position, he still prefers to dress like a laborer. His usual outfit consists of rumpled, dirty coveralls and an old pair of army shoes which are totally worn out. He doesn't seem to be concerned with setting a better example for his men. On numerous occasions he has come to work without shaving and apparently without taking a shower.

Barnes has, on several occasions, scheduled either too much or too little work for his crews. On several occasions men were seen loafing as they had completed their assigned jobs and had not received new work orders. On other occasions the men would complain that they were rushed and were given too much to properly accomplish in a given period of time. Most of the time, however, the crew chiefs were able to adjust the schedule with Barnes' approval.

Barnes seems to be reluctant to make a contribution to, or participate in community activities. He refused to contribute to the United Fund and the Heart Association campaigns, preventing the company from achieving a 100% participation goal. Although he has three sons who are active in the Boy Scouts, he has never volunteered his own time to help with scouting activities.
Last spring the air conditioning system was not operating on schedule as Barnes had underestimated the time it would take to recharge the air conditioners for summer use. He and his air conditioning men had to work twelve and fifteen-hour days during the last two weeks of May. They really had to push to get things operational by 1 June, and Barnes seemed rather irritable during this period. He lost his temper when one of his crew chiefs complained about putting in so much overtime.

Barnes is generally well liked by his men. They feel that he is very fair and unbiased. He has the ability to convey his wishes without sounding as though he is giving orders. He has the tendency to get impatient when things are not going just right, or when behind schedule, but he is quick to apologize and the men seem to respect this quality. He always shows a sincere interest in his men and their families by remembering birthdays with cards, their children's names, and little things such as this.

Barnes is particularly knowledgeable in the field of equipment maintenance. He makes a detailed study of all specifications and repair manuals for new equipment acquired, and insists that his men do the same. He has set up a comprehensive file of technical publications and has developed a formal training program for his men. They readily seek his advice when tough equipment problems arise as they respect his knowledge of maintenance and trouble-shooting procedures.
APPENDIX B (Con't.)

Barnes does not seem to be interested in company sponsored social activities. The men have asked him several times to join their bowling team but he has dismissed it as a waste of time. He seems to be a loner and has never been seen at a maintenance group social function. According to some of the men and one incident with the police, he appears to have a drinking problem. His wife has complained to some of the other wives that she may have to file for divorce if he continues to drink so heavily and neglect his family.

Barnes prefers to give verbal instructions to his crew chiefs although he is careful to document all work orders in writing. His written reports occasionally have to be corrected for minor errors, but are generally on time and contain the required information. He seems to have no difficulty in accomplishing routine correspondence with equipment manufacturers, dealers, etc.
APPENDIX C

Barnes appears to have the general cooperation and following of his subordinates. The men usually follow his work orders and schedules carefully, however, there have been a few examples of work not being completed on time. Overall, the crew chiefs seem to respect and recognize Barnes as their supervisor. He seems to be able to get most jobs done without being too forceful.

Barnes believes in setting a good example for his men by always presenting a neat appearance. Although his work sometimes causes him to get dirty, he always reports to the office with a neat work uniform, shined shoes, and neatly groomed. Barnes carries himself with pride and is never seen slouched in his chair with his feet on the desk. He appears to be very self-confident and looks the part of a supervisor.

Barnes has, on several occasions, scheduled either too much or too little work for his crews. On several occasions men were seen loafing as they had completed their assigned jobs and had not received new work orders. On other occasions, the men would complain that they were rushed and were given too much to properly accomplish in a given period of time. Most of the time, however, the crew chiefs were able to adjust the schedule with Barnes' approval.

Barnes is an active leader in the Boy Scouts and devotes many hours of his own time to scouting activities. This is a particularly unselfish contribution as he has no boys of his own. He has received statewide recognition for scouting leadership as a principal organizer.
of a scout camp for boys throughout the state. He was recently elected president of his church Community Relations Council, known for its work in raising funds for local charities. The Council recently raised over $73,000 for an addition to a local orphanage.

Last spring the air conditioning system was not operating on schedule as Barnes had underestimated the time it would take to recharge the air conditioners for summer use. He and his air conditioning men had to work twelve and fifteen-hour days during the last two weeks of May. They really had to push to get things operational by 1 June, and Barnes seemed rather irritable during this period. He lost his temper when one of his crew chiefs complained about putting in so much overtime.

Barnes is generally well liked by his men. They feel that he is very fair and unbiased. He has the ability to convey his wishes without sounding as though he is giving orders. He has the tendency to get impatient when things are not going just right, or when behind schedule, but he is quick to apologize and the men seem to respect this quality. He always shows a sincere interest in his men and their families by remembering birthdays with cards, their children's names, and little things such as this.

Barnes is particularly knowledgeable in the field of equipment maintenance. He makes a detailed study of all specifications and repair manuals for new equipment acquired, and insists that his men do the same. He has set up a comprehensive file of technical publications and has developed a formal training program for his men. They
readily seek his advice when tough equipment problems arise as they respect his knowledge of maintenance and trouble-shooting procedures.

Barnes appears to be a model citizen and family man. He is well liked by his men and is socially active with them and their families. He is usually the organizer of maintenance group parties and picnics and seems to be an informal leader in most social activities. Earlier this year he organized a company softball team which received enthusiastic participation by his men.

Barnes prefers to give verbal instructions to his crew chiefs although he is careful to document all work orders in writing. His written reports occasionally have to be corrected for minor errors, but are generally on time and contain the required information. He seems to have no difficulty in accomplishing routine correspondence with equipment manufacturers, dealers, etc.
## APPENDIX D

**LEADERSHIP:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor leadership ability. Ineffective in getting things done through others.</td>
<td>Average leadership ability. Moderately effective in getting things done through others.</td>
<td>Outstanding leadership ability. Highly effective in getting things done through others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ORGANIZATION AND PLANNING:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shows little organization and planning before taking action.</td>
<td>Average organization and planning before taking action.</td>
<td>Carefully organizes and plans activities before taking action.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**WRITTEN COMMUNICATION:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unable to present ideas or information clearly and accurately. Numerous errors.</td>
<td>Average writing ability. Adequate, but with some errors.</td>
<td>Written matter very clear, easily understood, and error free.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**RESISTANCE TO STRESS:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

**INTERPERSONAL SKILLS:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lacks tact and consideration in dealing with others.</td>
<td>Average ability to deal with others. Normally considerate and tactful.</td>
<td>Very tactful and considerate in dealing with others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TECHNICAL KNOWLEDGE:**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very poor knowledge of technical aspects of the work.</td>
<td>Moderate technical knowledge of most aspects of the job.</td>
<td>Outstanding technical knowledge of all aspects of the job.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX E

<table>
<thead>
<tr>
<th>LEADERSHIP:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor leadership ability. Ineffective in getting things done through others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average leadership ability. Moderately effective in getting things done through others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outstanding leadership ability. Highly effective in getting things done through others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPEARANCE:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sloppy dress and personal appearance. Poor bearing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average manner of dress, personal appearance, and bearing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outstanding dress, personal appearance, and bearing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORGANIZATION AND PLANNING:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shows little organization and planning before taking action.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average organization and planning before taking action.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carefully organizes and plans activities before taking action.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CITIZENSHIP:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not participate in community organizations or activities.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average participation in community organizations and activities.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extensive participation in community organizations and activities.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESISTANCE TO STRESS:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unable to cope with stress. Impatient.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate tolerance to stress. Occasionally impatient.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deals calmly with stress situations. Very patient.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERPERSONAL SKILLS:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lacks tact and consideration in dealing with others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average ability to deal with others. Normally considerate and tactful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very tactful and considerate in dealing with others.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX E (Con't)

<table>
<thead>
<tr>
<th>TECHNICAL KNOWLEDGE:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very poor knowledge of technical aspects of the work.</td>
<td>Moderate technical knowledge of most aspects of the job.</td>
<td>Outstanding technical knowledge of all aspects of the job.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERSONAL BEHAVIOR:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>WRITTEN COMMUNICATION:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unable to present ideas or information clearly and accurately. Numerous errors.</td>
<td>Average writing ability. Adequate, but with some errors.</td>
<td>Written matter very clear, easily understood, and error free.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


