

12-1977

Leniency Error as a Function of the Rater's Need for Social Approval

Danny Bean

Western Kentucky University

Follow this and additional works at: <https://digitalcommons.wku.edu/theses>



Part of the [Social Psychology Commons](#)

Recommended Citation

Bean, Danny, "Leniency Error as a Function of the Rater's Need for Social Approval" (1977). *Masters Theses & Specialist Projects*. Paper 2134.

<https://digitalcommons.wku.edu/theses/2134>

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.

Bean,

Danny

1977

LENIENCY ERROR AS A FUNCTION OF THE
RATER'S NEED FOR SOCIAL APPROVAL

A Thesis

Presented to
the Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by

Danny Bean

December 1977

LENIENCY ERROR AS A FUNCTION OF THE
RATER'S NEED FOR SOCIAL APPROVAL

Recommended 4-14-78
(Date)

Raymond Alan Marshall
Director of Thesis

Sam S. McFarland
James H. Gray

Approved 4-24-78
(Date)

Elmer Gray
Dean of the Graduate College

ACKNOWLEDGEMENTS

My parents deserve more than a special thanks because they made my education possible. My wife, Debbie, also is entitled to more than a mere "thank you" for not only her love and support but also many boring hours spent typing rough drafts.

Several people also aided in transforming this thesis into an acceptable paper, and as such, deserve recognition. Foremost is Dr. Ray Mendel, the director and recipient of several headaches and numerous cases of eye strain. The input from the other committee members--Drs. Jim Craig and Sam McFarland--is also appreciated. I would also like to thank Dr. Tom Madron and Mrs. Carolyn Marks for their expert assistance in persuading that inflexible computer to analyze my data.

Finally, I would like to extend a word of appreciation to Mr. Bernard Hurst, Mr. Ernie Murray, and the participants in this study. Without their cooperation and generous donation of time this investigation would not have been possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	v
LITERATURE REVIEW	1
Overview	1
Types of Rating Errors	2
Strategies for Dealing with Rating Errors	4
STATEMENT OF THE PROBLEM	16
METHOD	18
Subjects	18
Instruments	19
Procedure	20
Analysis	22
RESULTS	25
DISCUSSION	29
Leniency Error	29
Halo Error	35
Implications	37
APPENDIX A	39
APPENDIX B	43
APPENDIX C	48
APPENDIX D	50
APPENDIX E	53
APPENDIX F	55
REFERENCES	57

LIST OF TABLES

Correlations between the covariates and between the covariates and performance ratings	26
Summary of <u>F</u> values and significance levels of leniency error before and after N-SA (as measured by the PRI and LPC) was removed	27

LENIENCY ERROR AS A FUNCTION OF THE
RATER'S NEED FOR SOCIAL APPROVAL

D. Bean

December 1977

Directed by: R. Mendel, J. Craig, and S. McFarland

Department of Psychology

Western Kentucky University

The effects of the rater's need for social approval (N-SA) on leniency error in ratings was investigated. Each of 19 supervisors rated the performance of a common set of 12 workers. Using an ANOVA procedure (Guilford, 1954) leniency error was identified. The effects of N-SA as measured by the Personal Reactions Inventory (Crowne & Marlowe, 1960) and the Least Preferred Coworker (Fiedler, 1967) were partialled out of the ratings. When the ANOVA was repeated on the partialled ratings, it was hypothesized that reduced leniency error would result. Contrary to expectations, there was virtually no reduction in leniency error. The low reliability of the Personal Reactions Inventory, the questionable validity of the Least Preferred Coworker measure, and finally, the possibility that the underlying constructs of leniency error and N-SA are not related are offered as tentative explanations of this result. It was concluded, however, that this study may not have accurately measured the N-SA construct and directions for future research in this area are suggested.

LITERATURE REVIEW

One of the major problems confronting organizations today is accurately determining the quality and/or quantity of a worker's performance, or in other words, obtaining accurate performance ratings. The importance of accurate ratings is underscored by the many personnel decisions based wholly or in part upon them. These decisions include pay changes, promotions, transfers, dismissals, counseling, and consideration for further training and/or development (McCall & DeVries, 1976; Cummings, 1973). Performance ratings are often used as a criterion of success when validating selection instruments (Bayroff, Haggerty, & Rundquist, 1954; Lent, Aurbach, & Levin, 1971). Moreover, Cummings and Schwab (1973) and McCall and DeVries (1976) indicate that workers have a need to know how well they are performing.

Overview

As suggested above, performance ratings contribute substantially to the personnel decision-making process. These ratings are, however, plagued with various types of errors that certainly do not increase their utility. Accordingly, the initial area of focus is upon the judgmental aspects of performance ratings and three of the more common rating errors. The second section is focused on suggested strategies to alleviate these errors. The topic of discussion then narrows somewhat

as the relationship between rating errors and rater characteristics is addressed. The area of consideration narrows further as the personality of the rater and its relation to a specific rating error--leniency--is discussed. Finally, the rater's need for social approval (N-SA) is singled out as a potentially fruitful personality variable that might be related to leniency error. Specifically, a case is presented to support the contention that raters with a high N-SA will rate more leniently or favorably than those who possess a lower N-SA. If this is generally the case, then performance appraisal tasks should be assigned to those who possess a low N-SA or ratings rendered by those with a high N-SA should have their ratings adjusted accordingly. This would eliminate some of the constant error in ratings, and thereby help rid performance ratings of one common cause of distortion.

Types of Rating Errors

Even though this paper is primarily concerned with reducing the effect of only one type of rating error, perhaps it would be beneficial to examine several of the more common rating errors. Although performance ratings are used more frequently than all other appraisal techniques combined (Cummings & Schwab, 1973; Lawshe & Balma, 1966; Lent, Aurbach, & Levin, 1971), they are vulnerable to many types of errors. The tendency for Federal judges to more rigorously scrutinize rating criteria than other measures of performance highlights the less than desirable reputation possessed by performance ratings (Edwards, 1976).

Various reasons (i.e., rating errors) for their tainted reputation are discussed below.

Leniency

Leniency error is presumed constant for any given rater and refers to ratings that are either systematically too high or too low (Blum & Naylor, 1968; Guilford, 1954). In other words, this error is present when the rating is shifted either above or below the actual "true" rating. The rater who produces excessively high ratings is said to exhibit positive leniency while the one who rates too low demonstrates negative leniency.

Halo

Halo error is the result of one specific trait or the general impression of the ratee influencing the rating given on other traits (Blum & Naylor, 1968; Guilford, 1954) which obfuscates the relative strengths and weaknesses of ratees. Also, ratings will be less valid to the extent that they are shifted away from the "true" scores (Guilford, 1954). Rating all persons on one trait before advancing to the next as opposed to judging each person on all traits is one commonly accepted way of combating halo error (Guilford, 1954).

Central Tendency

Not using either extreme of the rating scale, even when deserved, yields the error of central tendency (Blum & Naylor, 1968; Guilford, 1954). The rater checks points near the mean thus reducing variability. This is known as a restriction of range and places an artificial ceiling on correlation coeffi-

cients between the performance ratings and other variables under observation.

Strategies for Dealing with Rating Errors

The rating errors in performance ratings have not, by any means, gone unnoticed. Several researchers and practitioners have suggested ways to deal with these errors. They include: a) using different scale developmental procedures (Sission, 1948; Smith & Kendall, 1963); b) using various scale formats (Bayroff et al., 1954; Borman & Dunnette, 1975; Burnaska & Hollmann, 1974; Campbell, Dunnette, Avery, & Hellervik, 1973); c) using several raters (Bayroff et al., 1954; Borman, 1974; Mandell, 1956); d) training the raters (Blum & Naylor, 1968; Latham, Wexley, & Pursell, 1975; Guilford, 1954); e) allowing subordinates to participate in the overall appraisal process (Cummings, 1973; Friedman & Cornelius III, 1976), and f) using corrective scoring methods (Bass, 1956).

These suggested remedies, however, either have not proven as effective as originally hoped or involve problems with their implementation. Different scale developmental procedures such as the forced-choice or behavioral expectations scales have not generally demonstrated sufficient additional utility to warrant their extended (in terms of time and effort) developmental process (Berkshire & Highland, 1953; Borman & Dunnette, 1975). While several researchers (e.g., Bayroff et al., 1954; Mandell, 1956) concluded that

the average for a group of raters gives more reliable and valid ratings than does a single rater, this finding would be of value only to organizations where several people are in a position to evaluate a worker. However, in many companies, employees have only one immediate superior, and managers occupying higher levels in the hierarchy are often not in a position to have sufficient knowledge of each employee to render accurate judgments. Peers represent an alternative source of ratings, but this, too assumes that they have adequate knowledge about the ratee's performance and that there is a high degree of trust among the peers (Cummings & Schwab, 1973). Subordinates may also be used as raters, but this approach has certain drawbacks. The subordinates as well as the ratee may perceive the entire process as illegitimate and threatening, and subordinates may base their ratings on something other than the ratee's performance (Cummings & Schwab, 1973). In addition, multiple ratings may be biased according to Klimoski and London (1974). Latham et al. (1975) found that training observers will minimize rating errors, but training can be costly in both trainer fees and lost production while raters are being trained. It has also been found that participation in the performance appraisal process is effective (i.e., it reduces error) for some workers, but not for other workers (Meyer, Kay, & French, 1965). Bass (1956) concluded that a binary scoring method helps control rating errors, but raters can still rate

leniently. In sum, the aforementioned rating error remedial strategies are generally less effective than desired.

Reducing Rating Error Through Rater Characteristics

General Characteristics. Although distinct from the research discussed above, another body of literature exists that is also concerned with the problem of rating errors. This literature is based on the assumption that there is a relationship between raters who generally appraise performance with less error (i.e., leniency, halo, central tendency, etc.) and certain characteristics possessed by them. Some of these rater characteristics are identified below.

Kirchner and Reisberg (1962) compared ratings made by better and less-effective supervisors where supervisory effectiveness was itself determined by ratings of overall job performance. Results indicated that, in general, the ratings given by less-effective supervisors discriminated less (i.e., possessed less variance) between low- and high-rated subordinates. Also, the less-effective supervisors rendered more lenient ratings than did the more-effective supervisors, especially for poorer subordinates. The same conclusion (i.e., supervisors who perform better yield more accurate ratings¹) was reached by Mandell (1956).

¹Rating accuracy refers only to the absence of leniency, halo, central tendency error, etc. It is realized that other factors such as construct validity affect rating accuracy.

Utilizing Army officers, Schneider and Bayroff (1953) have also examined the relationship between certain rater characteristics and the validity of ratings. Validity, in this instance, was defined as the degree to which the raters' appraisals coincided with the average rating on the same characteristics obtained from twenty peers. The specific rater characteristics under investigation were aptitude (determined by scores on the Officer Classification Test), academic achievement, and overall value to the Army (determined by the criterion score). After obtaining performance ratings, the raters were trichotomized on the three characteristics under consideration, and for each group the ratings were correlated with the criterion scores. In most instances, validity coefficients for the raters in the high groups on the above three characteristics were of greater magnitude than those for the medium groups. In addition, validities obtained for the high groups were always higher than the validities of ratings produced by the raters in the low groups on the three characteristics.

Pizam (1975) investigated a rater characteristic labeled "social differentiation." This was defined as a personal characteristic representing a rater's "differentiating ability or style of rating behavior" (p. 245). A high differentiator used a large part of the rating scale while a low differentiator concentrated ratings into a narrow range of the scale. By comparing rating variances, Pizam (1975) found that this

characteristic was stable over "several years" (p. 245) and suggested that low differentiators either not be used as raters or that their judgments be adjusted to account for this rating style.

To briefly review, the less accurate raters have been associated with the following general characteristics: They generally are less-effective supervisors (Kirchner & Reisberg, 1962; Mandell, 1956); they demonstrate lower aptitude, academic achievement, and over-all value to the Army (Schneider & Bayroff, 1953); and they generally use a smaller portion of the rating scale (Pizam, 1975).

Personality Characteristics. While the preceeding discussion examined a broad range of rater characteristics associated with rating errors, a parallel literature reports on the attempt of several investigators to identify specific personality variables which relate to rating styles, especially those of a lenient or favorable nature.

Mandell (1956) concluded that tough and lenient raters have distinct personality differences:

The tough rater tends to be younger, tense, rigid, and with little self-confidence. He is sour towards people, not oriented toward his men, but tries hard to impress his superiors. The lenient rater is older, likes people, and is uncritical. He is self-confident, but does not have a driving ambition. He derives job satisfaction from a feeling of rendering service (pp. 439-440).

Buel (1962) stated that certain raters, because of their personality, have a predisposition to be complimentary, uncomplimentary, or to misperceive performance. Adams (1927), meanwhile, described a "good" judge (i.e., one who commits fewer errors) as being "cold-blooded towards others and not interested in them" (p. 181), and as viewing others as tools rather than human beings. In addition, Nash (1966) described a hypothetical, more-effective manager compared to a less-effective one as being "not service or humanitarian oriented" (p. 254). This combined with the Kirchner and Reisberg (1962) and Mandell (1956) conclusions--more-effective supervisors commit less rating errors--would lead one to predict that the more accurate and, particularly, less lenient rater would not be people oriented.

The Need for Social Approval. It appears that aspects of personality are related to lenient ratings. Furthermore, the literature focuses upon the socially oriented components of personality when examining the origins of leniency error. A final series of investigations concentrates directly on this aspect of personality as it relates to this type of rating behavior. Taft (1955) asked raters to check their traits on the Gough Adjective Check List. Poor judges (i.e., those who yielded error-plagued ratings) almost always checked socially relevant traits while good judges stressed traits concerned with the execution of tasks. He concluded that good

judges were not as socially oriented or socially expressive as were poor judges and that ability to rate others accurately is negatively correlated with social dependence. This is in agreement with the proposition that social detachment is a necessary condition for making accurate ratings. These results are also in agreement with those obtained by Mandell (1956) and Adams (1927) wherein they described a non-lenient rater as sour and cold towards people and Nash (1966) who described a hypothetical more-effective manager as not being service or humanitarian oriented. Remember that more-effective supervisors were found to yield ratings with less errors and more discriminations of performance than did less-effective supervisors (Kirchner & Reisberg, 1962; Mandell, 1956).

Examination of the leadership literature indicates that strong social needs are characteristic of supervisors. This statement and accompanying research may appear contradictory to the above literature which describes the non-lenient rater (supervisor) as effective, cold, harsh, and non-sociable, but an explanation of this apparent inconsistency will be offered after the leaders-are-sociable evidence is presented.

In a review of the literature to identify leader characteristics, Stogdill (1948) cited numerous studies that found leaders participate in more group activities and exhibit a higher rate of social mobility than do nonleaders. Also, several researchers found that sociability and leadership

have a high positive correlation, and at least one investigator found negative correlations between bashfulness and leadership and between seclusiveness and leadership. Stogdill arrived at two conclusions that are relevant and are supported by not less than 10 studies. First, a leader exceeded his group members in sociability, and second, sociability was one of several "items with the highest overall correlation with leadership" (p. 63). Moreover, Nash (1965) stated that more-effective managers scored higher on a social service, humanitarian, people oriented interest.

Now, two pictures of leaders or supervisors have been presented. The first suggested that effective supervisors are neither lenient raters nor are they people oriented. The second implied that leaders are more sociable and more outgoing than nonleaders. Both descriptions include or imply leader effectiveness, but the descriptions differ in regard to the leader's relations with others or the leader's people orientation. It is doubtful that two completely different populations of leaders are being described. Rather, the situational factor involved in leadership is a more viable explanation for this apparent contradiction (Fiedler, 1967; Vroom, 1976). This view contends that the behavior exhibited by a leader is contingent upon the situation as well as other variables. It is not unreasonable for a successful leader to function differently in different situations. In other words, a colder, non-people orientation is

appropriate for some situations while a sociable style is called for in other situations. Of particular relevance to the present investigation is the proposition that, although leaders possess a high amount of sociability, nonsociable behavior is more appropriate when completing performance appraisals.

With the resolution of the apparent contradiction, the studies cited above suggest that the social aspects of a rater's personality may be associated with leniency error. However, more evidence to reinforce this point can be obtained by examining the relationship between the rater and the ratee. It has been found (Levinger, 1961) that raters will distort their ratings in a direction that is congruent with their attraction toward the ratees, and Stockford and Bissell (1967) concluded that "the scores on merit-rating scales measure primarily the personal-social relationships between supervisor and subordinate rather than the output of the subordinate in question" (p. 144). At least three reasons have been offered as to why this is so. First, a rater feels that a bad rating will be detrimental to rater-ratee relations (Bass, 1956; McCall & DeVries, 1976). Second, the rater has a personal interest in the ratee (Pizam, 1975). Third, the rater "may feel it necessary to always approve others in order to gain approval for himself" (Bass, 1956, p. 359).

To substantiate this point, consider the study conducted by Kallejian, Brown, and Weschler (1962) of the effect of

interpersonal relations between the superior and the subordinate on over-all performance ratings. A clinically skilled interviewer was given a personality evaluation of the supervisor and, from interviews with a group of workers and the leader of that group, he tried to predict each supervisor's ratings for the individuals in the groups. One assumption for the predictions was that supervisors will emphasize characteristics of performance which are related to their personal needs. Another assumption was that "those individuals who behave in such a way as to satisfy the personal needs of the superior will generally be rated higher and those subordinates who interfere with the satisfaction of the personal needs of the superior will generally be rated lower" (p. 139). When the clinician's predicted ratings were compared to actual ratings made by supervisors, a significant correlation was obtained (the correlation was not provided, but $N = 11$ and $p < .03$). In other words, the clinician was able to predict performance appraisal ratings made by the supervisors given a personality evaluation of each supervisor and some interaction with each supervisor's work group, but without actually having observed the employees on the job. The researchers concluded that the same personal characteristics and needs of the superior influence both his ratings and the way he sees himself and responds to the world around him. It may be, however, that the clinician was responding to how he thought the workers performed on

the job rather than how well they satisfied the needs of their supervisor.

It has also been shown that ratees will like judges who have given them a desirable rating. Several research projects support this contention. Sigall and Aronson (1969) investigated the liking for an evaluator as a function of her physical attractiveness and nature of the evaluation. They found that "the attractive evaluator who was positive was liked most, while the attractive evaluator who presented a negative evaluation was liked the least" (p. 93). Between these two points on the continuum, they found that the unattractive-positive evaluator was liked more than the unattractive-negative evaluator. This suggests that a favorable evaluation is more highly valued than having an attractive evaluator. Other studies (Bryne, 1969; Hewitt, 1972; Jones, Gergen, & Davis, 1962) have also found that ratees like raters who have rated them highly. These studies strongly suggest that ratees like raters who evaluated them favorably and that raters might, indeed, give higher ratings so that ratees will like them.

Several conclusions emerge from the literature reviewed. First, the importance (i.e., the number of decisions which are based on performance ratings) and the widespread use of performance ratings were documented. Next, errors inherent in ratings and some suggested cures were discussed. It was also mentioned that these remedies were not as successful

as predicted or hoped. Consequently, another approach to alleviating rating errors (especially leniency error) was explored. This approach involved examining the relationship between characteristics of the rater and the accuracy of that person's ratings. This position evolved by showing that several rater characteristics were associated with rating accuracy. Personality and, more specifically, its social aspect was then set forth as an avenue of research that might help explain some of the error in performance ratings. Finally, several studies emphasizing the high social needs of those individuals typically generating performance ratings were presented.

STATEMENT OF THE PROBLEM

Lacking in the literature are investigations which directly examine the relationship between a rater's social needs and distortions in ratings, especially leniency error. This omission is particularly unexplainable in view of the fact that reasonably well accepted measures (e.g., the Personal Reactions Inventory and the Least Preferred Co-worker) of the N-SA construct are available.

The purpose of this investigation, therefore, is to determine if the N-SA can be used to reduce leniency error in performance ratings. It is hypothesized that a rater with a high N-SA will render more favorable or lenient ratings than will a rater with a low N-SA. If this is the case, then ratings obtained from those with differing amounts of N-SA can be statistically adjusted thereby rendering them more comparable, or performance appraisal tasks can be assigned to those with a low N-SA. To further illustrate this point, consider two average employees, one of whom has a supervisor with a high N-SA and the other who works under a supervisor with a low N-SA. Assuming the hypothesis of this investigation to be true, the employee with a high N-SA supervisor would in general receive higher ratings. This rating error could be corrected by statistically re-

moving the effects of N-SA (i.e., by treating N-SA as a covariate), and thus, making the appraisals more comparable.

Earlier halo error was also identified as a major rating error. The literature suggests that the format of the rating scale (i.e., rating across people rather than across items) can help control this error (Guilford, 1954). As indicated previously, halo error occurs when a specific trait or the general impression of the ratee influences the rating given on other traits (Blum & Naylor, 1968; Guilford, 1954). Rating across items calls for judging one ratee on all traits. This presents a greater opportunity for the rater to form a general impression about the ratee than does the format which asks the rater to judge all ratees on one trait (i.e., rating across people) before advancing to the next. Accordingly, it is believed that the rating-across-people format will better control halo error (Guilford, 1954).

METHOD

The general strategy of the present study required 22 supervisors (hereafter referred to as raters) to rate the job performance of a common group of 12 subordinates, each of whom performed the same job. Measures of the N-SA (i.e., the Personal Reactions Inventory and the Least Preferred Coworker) of each supervisor were collected and these data were examined for their relationship to leniency error in the subordinate ratings. In addition, two formats (i.e., rating across items and people) of the appraisal instrument were devised to ascertain which better controlled halo error.

Subjects

The 22 supervisors serving as subjects (raters) were employed in a moderate-sized southeastern distillery. The raters ranged in age from 29 to 63 with a mean of 42.8, and they possessed a mean education of 11.4 years. Employment length in the present position varied from three to 26 years with a median of 17. Nine of the raters were nonunion, first-line male supervisors responsible for coordinating production materials, writing efficiency reports, and solving problems related to two bottling lines. The remaining twelve were "assistants" to the supervisors, female, and union members. Their duties included assigning workers to the different

jobs on a line and helping where assistance was needed. All judges were in a rather unique rating position in that each was knowledgeable about the performance of a common set of workers.

Instruments

Two measures reflecting the N-SA (Strickland, 1967) that seem particularly appropriate for investigating the relationship between N-SA and leniency error are the Personal Reactions Inventory (PRI) developed by Crowne and Marlowe (1960) and Fiedler's (1967) measure of Least Preferred Coworker (LPC). According to Crowne and Marlow, the PRI (see Appendix A) measures "the need of subjects to obtain approval by responding in a culturally appropriate and acceptable manner" (p. 353). Certainly, giving a high rating as opposed to a low rating is seen as socially acceptable. A split-half reliability of .88 and a test-retest reliability of .89 are reported for this instrument (Crowne & Marlowe, 1960). According to Fiedler (1967), the LPC measure (see Appendix B) indicates whether one's leadership style emphasizes interpersonal relations (high LPC score) or task accomplishments (low LPC score). Leadership style reflects the individual's underlying need structure which, in turn, motivates behavior in various leadership situations (Fiedler, 1967). This is further supported by Larson and Rowland (1973) who concluded that differences in leadership behaviors between high and low LPC scorers are more easily explained by the

leaders' underlying personal need structures. In a discussion of the LPC measure, Ashour (1973) pointed out a particularly relevant Fiedler assumption--"a task-oriented person tends to give a less favorable evaluation to his least preferred coworker than a relations-oriented person" (p. 340). "Split-half correlations for LPC have been consistently between .85 and .95" (Fiedler, 1973, p. 360).

A 44-item performance appraisal instrument (see Appendix C) was developed (as described later in this section) to collect supervisory ratings of subordinates. Intuitively, the items on the instrument appear to represent four traits. Briefly, they are: a) labeler operation - deals directly with the operation and maintenance of the labeling machine; b) quality control - pertains to the quality consciousness (i.e., appearance of the bottles) of the worker; c) general - reflects behaviors expected of all employees; and d) safety - refers to the worker's safety consciousness in operating the labeling machine.

Procedure

The procedure can best be conceptualized when considered in the following three phases: a) the measurement of the raters' N-SA, b) the development of the performance appraisal instrument, and c) the collection of supervisory ratings of subordinates.

Phase I--PRI and LPC Administration

Since the distillery manager prohibited meetings on

company grounds, all contact with the raters was made by phone or letter. Both mediums were utilized in gaining the assistance of 19 of the 22 possible raters. The first letter, which also served to introduce the PRI and LPC, is contained in Appendix D. A postage-paid return envelop was provided to encourage a high return rate for the PRI and LPC. If the completed instruments were not returned in two weeks, a call was made urging the rater to do so.

Phase II--Rating Scale Development

After all PRI and LPC forms had been returned, and during subsequent home visits with five of the raters, critical behaviors inherent in good, mediocre, and poor on-the-job performance were collected (Borman & Vallon, 1974). The author, who had previously worked in the distillery for six weeks, then deleted duplicate and ambiguous behaviors, leaving a total of 44 behaviors. A 5-point, Likert-type response format was attached to each behavior. In an effort to test an accepted method of controlling halo error, two formats of the appraisal instrument were devised (Guilford, 1954). One format (see Appendix C) listed the ratee at the top of the page and asked the rater to judge that person on each of the following 44 items. The other format (see Appendix E) contained the same items but the rater was asked to appraise the performance of each worker on one item before advancing to the next item. Essentially, this allows a comparison of ratings made across items with

those made across people to ascertain which better controls halo error.

Phase III--Subordinate Ratings

The raters were then randomly divided into two groups, and one of the two rating formats along with an introductory letter (see Appendix F) was mailed to each rater. Each was asked to evaluate the performance of the same twelve workers who had been preselected to represent the full continuum of overall good to bad performance. The 12 preselected ratees had been chosen by the same five raters who assisted in developing the rating instrument. Postage-paid return envelopes were again provided, and if the forms had not been returned in two weeks, a call was made asking the rater to return them.

Analysis

Data Coding

A value of one (least favorable) to five (most favorable) was assigned to the responses for each item. For example, if an item was stated such that the "strongly agree" category was the most favorable and that option was checked, then that response was coded as a five. For that same item, if the "strongly disagree" category was selected, then that response was coded as a one. If an item contained more than one or no response, then it was designated as missing data and excluded from further analysis.

Analytic Procedure

The primary purpose of the analysis was to assess the extent to which N-SA could be used to reduce or remove the effects of leniency error in the supervisory ratings of subordinates, and of secondary importance, to determine which format better controlled halo error. To accomplish this, a model set forth by Guilford (1954, pp. 280 ff.) for analyzing rating errors was utilized.

Guilford's model contains three variables--raters, ratees, and traits--with one observation per cell. Guilford contends that by successively collapsing across each of the three variables thereby creating within cell variance and then performing three two-way ANOVA's on the remaining variables, three rating errors can be identified. Specifically, a rater main effect (collapsing across either ratees or traits) indicates leniency error; a rater-ratee interaction (collapsing across traits) reflects halo error; and a ratee-trait interaction (collapsing across raters) identifies the tendency of a rater to undervalue or overvalue a certain trait in others.

The underlying dimensions (i.e., traits) required by the model were identified through factor analysis (Kim, 1975). Next, the ratings were collapsed across traits to create within cell variance and a two-way (rater-ratee) ANOVA was used to determine leniency error.

Once the ANOVA to determine leniency error was performed, the variance in each rater's ratings in common with

that rater's PRI score was removed. This partialing procedure was repeated on the original ratings in order to remove the LPC variance. Finally, the same process was repeated a third time, removing both PRI and LPC variance from the original ratings. The initial ANOVA was then repeated on the three sets of partialled ratings. These analyses are designed to indicate whether adjusting ratings for the rater's N-SA reduces leniency error in those ratings. It should be noted that this procedure is analogous to three analyses of covariance with PRI, LPC, and PRI and LPC serving as covariates.

The analysis to identify the relative amount of halo error inherent in each format was then to be conducted. Results of the factor analysis to determine the underlying traits (i.e., the dimensionality of the rating instrument) precluded further investigation of the halo issue.

RESULTS

Based on an 86% return rate (19 out of 22), two of the three instruments in this study possessed high reliability while the other was low. Coefficient alpha for the performance appraisal instrument was .96; for Fiedler's LPC measure alpha was .93; and for the PRI, alpha was .57. Four traits identified on an a priori basis were expected when the performance appraisal instrument was factor analyzed. Instead, only two traits materialized. Trait I was a general trait which reflected many aspects of the job and accounted for 33.4% of the variance. Trait II accounted for 10.1% of the variance but was not behaviorally interpretable. Coefficient alpha for each was about .90.

The correlations between the measures of N-SA, and between each of these measures and each dependent variable (i.e., both traits and total ratings) are displayed in Table 1. It should be noted that while the PRI and LPC were employed to measure N-SA, the low obtained correlation between them suggests they are not tapping the same underlying construct. The relationships of the two measures of N-SA with the total ratings were assessed by computing the mean of all ratings given by each rater and then correlating these 19 mean ratings with the 19 scores on the PRI, and then

TABLE 1
CORRELATIONS BETWEEN THE COVARIATES AND BETWEEN THE
COVARIATES AND PERFORMANCE RATINGS

	PRI	LPC
PRI	---	
LPC	.28	---
Total Ratings	.10	.26
Trait I Ratings	.21	.13
Trait II Ratings	-.04	.20

with the 19 scores on the LPC measure. A similar procedure was implemented for ascertaining the relationship between each N-SA measure and each trait. In other words, rather than averaging all ratings rendered by each rater, only those items pertaining to each respective trait were averaged and subsequently correlated with each rater's N-SA score.

As detailed in the source table summary (see Table 2), there was a significant amount of leniency error contained in the ratings, $F(18,228) = 7.48, p < .001$. Table 2 shows the leniency error remaining after the variance in the ratings shared with the PRI was removed, $F(18,228) = 7.39, p < .001$. Since this rater main effect was still significant and virtually equivalent to that contained in the initial ratings, it is clear that N-SA, as measured by the PRI, did not substantially reduce leniency error. The effects of LPC on leniency error is also shown in Table 2. The removal of LPC variance in the ratings also failed to reduce this error,

TABLE 2

SUMMARY OF F VALUES AND SIGNIFICANCE LEVELS OF LENIENCY ERROR BEFORE
AND AFTER N-SA (AS MEASURED BY THE PRI AND LPC) WAS REMOVED

Source of Variation	Df	Initial Ratings		PRI Removed		LPC Removed		PRI and LPC Removed	
		<u>F</u>	<u>p <</u>	<u>F</u>	<u>p <</u>	<u>F</u>	<u>p <</u>	<u>F</u>	<u>p <</u>
Rater	18	7.48	.001	7.39	.001	6.97	.001	6.96	.001
Ratee	11	2.84	.002	2.84	.002	2.84	.002	2.84	.002
Rater-Ratee	198	.41	.999 ^a	.41	.999	.41	.999	.41	.999
Error	228								
Total	455								

Note. The F ratios for Ratee and Rater-Ratee are constant because only the variance due to N-SA in the Raters is being partialled out.

^aA significance of .999 is given when F < 1.0.

$\underline{F}(18,228) = 6.97, p < .001$. The final attempt to remove leniency error combined the effects of the PRI and LPC (see Table 2). The results are, however, the same--no significant reduction, $\underline{F}(18,228) = 6.96, p < .001$.

DISCUSSION

The primary hypothesis of this investigation (i.e., leniency error and the rater's N-SA are directly related) was not supported while the secondary hypothesis (i.e., rating across people better controls halo error than does rating across items) was not tested. Possible explanations for the lack of support for the primary hypothesis and a justification for not proceeding with the halo error analysis will be provided. Finally, implications for future research in this area will be discussed.

Leniency Error

Leniency error was not reduced by statistically controlling for the effects of N-SA as measured by the PRI and LPC. Possible causes for this result include the unreliability of the PRI, the lack of PRI and LPC construct validity, and finally, the possibility that N-SA and leniency error are not causally related.

First, it should be noted that the statistical analysis employed in this study and three analyses of covariance are different means to the same end. In other words, had three analyses of covariance been conducted the results would have been the same. Certain conditions, however, must be met before an analysis of covariance is appropriate,

one of which is a completely reliable covariate. While an analysis of covariance per se was not conducted, a failure to meet this assumption may explain why the hypothesis was not supported. As mentioned earlier, the PRI developed by Crowne and Marlowe (1960) yielded a coefficient alpha of .57 in this study, even though the literature suggests it is much higher--alphas of .88 and .89 are reported by Crowne and Marlowe (1960). Obviously, this reliability is not of sufficient magnitude to satisfy the above assumption, and thus, the statistical model may not have been appropriate. Several possible contributors to the low reliability are identifiable. Perhaps most important is the difference between the normative population (Crowne & Marlowe, 1960) and the subjects in this study. The PRI was developed using 39 college students (10 males, 29 females) with a mean age of 24.4 and a range of 19 to 46 years. The present study utilized 19 supervisory-type personnel (seven males, 12 females) with a mean age of 42.8 and a range of 29 to 63 years. The mean education was 11.4 years. The male to female ratio in both studies is approximately the same, but there is a substantial difference in age and education. It may be that the PRI is sensitive to N-SA for the younger people with more education and an invalid measure of N-SA in the older and/or less educated subjects.

Additional causes of the PRI's low reliability are also identifiable. Subjects in the Crowne and Marlowe normative

sample completed the PRI under standardized conditions while no such control was exercised in this study. That is, all subjects were free to complete the instrument at their leisure rather than under standardized conditions. This explanation is weak, however, because the LPC and the performance appraisal instrument were completed under similar conditions and both possessed high reliability. Although the instructions differed somewhat to allow for anonymity, they are clear and concise, and are not seen as a source of unreliability. Since the responses were scored twice, scoring errors are also ruled out. In sum, of the several factors which possibly may have contributed to the PRI's unreliability, all are discounted except the discrepancy between the age and education level of the normative sample and the sample in this study.

Recognizing the problem of an unreliable covariate, Elashoff (1969) mentioned a procedure developed by Lord (1960, cited in Elashoff, 1969) and one devised by Porter (1967, cited in Elashoff, 1969) to correct for an unreliable covariate. Neither of these two procedures were implemented, however, because of a general rule of thumb provided by Cochran (1957, cited in Elashoff, 1969). This guideline states that if the correlation between the covariate and the dependent variable is below .30, then the elimination of the covariate variance from the dependent variable will only negligibly increase the correlation between the dependent and the independent variable. As presented in Table 1, the

correlations between both measures of N-SA (i.e., the covariates) and the dependent variable are below .30. In addition, for the hypothesis to be supported, the analysis of the partialled ratings should not reveal a significant rater effect. It is highly doubtful that a negligible increase in the N-SA variance removed from the ratings would alter the F ratio to the extent that the overall conclusion would be changed. For this reason, the additional analyses were not completed.

Low reliability possibly explains the PRI's failure to reduce leniency error, but this argument becomes untenable for the reliable LPC measure. Consideration of LPC construct validity, however, may explain why it too failed to reduce leniency error. Because reliability is a prerequisite for validity, obviously, the PRI is not valid.

Earlier it was stated that the LPC measure reflects the individual's underlying need structure which, in turn, determines whether one's leadership style emphasizes interpersonal relations or task accomplishments (Fiedler, 1967). Although not developed as a measure of N-SA, this instrument was selected because it purports to reflect this underlying need structure. Even though this may be an indictment of the original wisdom of choosing the LPC measure, the fact that many writers have pointed out discrepant results regarding what LPC measures suggests that its construct validity is questionable. Vroom (1976) stated that "the meaning of the LPC

scores is still a matter of some conjecture" (p. 1534). Based on personal communication from Schroeder (1969) and the work of Mitchell (1970), Fiedler has hedged somewhat on the interpretation of LPC scores. He stated that "the LPC score must be seen as a measure which at least in part reflects the cognitive complexity of the individual" (Fiedler, 1971, p. 129). Mitchell (1970) also hypothesized that LPC might be reflecting cognitive complexity and halo error because of the uniformly undifferentiated manner, as measured by item variance, in which low LPC persons describe their least preferred coworker, while high scorers do discriminate more among the adjectives on the instrument (i.e., possess a greater variance). This hypothesis could have been easily tested in this study using LPC as a covariate in the halo error analysis were it not for the outcome of the factor analysis.

Perhaps even more damaging to the validity of the LPC measure is another finding presented by Mitchell (1970). He found that, in accomplishing a hypothetical task, low LPC scorers preferred good interpersonal relations while high LPC scorers favored an efficient group. This is exactly opposite the situation predicted by Fiedler (1967). In addition, the correlation between the PRI and LPC measure corrected for attenuation, which is about only .40, indicates that the two instruments are reflecting somewhat different variables. In sum, due to the competing interpretations of what the LPC

is actually measuring and its relationship with the PRI, one cannot help but wonder what the instrument is reflecting. If it is, indeed, not measuring an underlying need structure, then it could not be expected to reduce leniency error.

The final explanation offered for the failure of either of the N-SA measures to reduce leniency error is also the most obvious. It may simply be that N-SA and leniency error are not related. This does, however, run contrary to the majority of the evidence presented earlier. For example, several researchers (Adams, 1927; Buel, 1962; Mandell, 1956; Nash, 1966; Taft, 1955) suggested that personality, and more specifically, its social aspects are related to lenient ratings. Stogdill (1948) also cited several studies showing the high social nature of leaders who, after all, generally appraise performance. In addition, evidence was presented supporting the contention that ratees like raters who have evaluated them highly and that raters may rate leniently so that they will be liked (Bass, 1956; Bryne, 1969; Hewitt, 1972; Jones et al., 1962; Kallejian et al., 1962; Levinger, 1961; McCall & DeVries, 1976; Stockford & Bissell, 1967). This quantity of supporting research suggests that the present study simply did not uncover the relationship between N-SA and leniency error. After all, the study hinged on an accurate indicator of N-SA and the measure with the previously established validity (i.e., the PRI) was unreliable, and the reliable measure (i.e., the LPC) may not have been valid.

Halo Error

Briefly, the results of the factor analysis prevented an analysis of halo error. To more fully understand why this is so, one must consider how Guilford's (1954) model allows assessment of halo error. As previously indicated, this model for analyzing rating errors contains three variables--raters, ratees, and traits. To determine halo error, one must collapse across traits and investigate the rater-ratee interaction. A significant interaction then indicates the presence of halo error.

The performance appraisal instrument was factor analyzed to determine the traits being assessed. Rather than obtaining four traits as expected, only two were obtained. As indicated earlier, the first was of a general nature encompassing the entire job, but the second was not behaviorally interpretable. Halo error is defined as the result of one specific trait or the general impression of the ratee influencing the rating given on other traits (Blum & Naylor, 1968; Guilford, 1954). Given this definition, a meaningful interpretation of the rater-ratee interaction (which reflects halo error when collapsing across traits) is predicated upon there being at least two definable traits. Since one of the two factor analytic traits was uninterpretable, the rater-ratee interaction is not indicative of halo error in this study.

Since both the leniency and halo error analysis require

that within cell variance be created by collapsing across traits, one might now question why the former was conducted. In short, the leniency error analysis does not require that the traits possess a conceptual basis, while any halo error analysis, by definition, requires that the traits be meaningful.

Two explanations for the unidimensionality of the appraisal instrument are offered. First, it may be that the instrument was, in fact, reflecting the global job and not discriminating between its major components. A closer examination of the instrument, however, indicates that this is probably not the case, and more validity is attributed to the second cause as explicated below.

A study by Schneier (1977) of the effects of rater cognitive complexity on rating error serves as the basis for the second explanation. Cognitive complexity was defined as "the degree to which a person possesses the ability to perceive behavior in a multidimensional manner" (p. 541). Briefly, Schneier found that cognitively complex raters yielded ratings with significantly less halo error than did cognitively simple raters. It may be that the raters in this study were not cognitively complex. In other words, if the raters could not view behavior multidimensionally, then a factor analysis of their ratings would reflect very few or only one trait. The education level of the raters would argue against them being cognitively complex (Brown,

1976). Since cognitively simple raters also exhibit more halo error (Schneier, 1977), it is likely that an excessive amount of halo error also contributed to the appraisal instrument's apparent unidimensionality. In sum, the lack of at least two definable traits with which to conduct the halo error analysis may be attributable to a unidimensional appraisal instrument or the inability of the raters to view behavior multidimensionally.

Implications

Although the primary hypothesis was not supported, the findings suggest directions for future research in this area. Subsequent investigations might pilot test their measure of N-SA before administering it to the raters to determine in advance at least its reliability if not its validity. Furthermore, special attention should be given to the population chosen for investigation. If the PRI is again used, then utilizing college students as raters would, no doubt, increase the chances of obtaining a higher reliability for this N-SA measure. On the other hand, one scale of a personality test (e.g., the Need for Affiliation from the Edwards Personal Preference Schedule) may provide a better measure of the N-SA. Another and yet simpler means to quantify the raters' N-SA would be to ask the raters how much each wants to be liked by each of the ratees. It may also be that a better measure of the N-SA needs to be developed.

Concerning the test for determining which rating format better controls halo error, identifying a population of cognitively complex raters might prove helpful (Schneier, 1977). This suggestion is made because the rating across people format is designed to reduce halo error, not by creating or increasing the rater's cognitive complexity, but by allowing the complexity to manifest itself if initially there.

APPENDIX A

Matching Number: _____

PERSONAL REACTIONS INVENTORY

INSTRUCTIONS: The purpose of this questionnaire is to obtain your view about situations that might arise in everyday life. Please read each item carefully, and then circle the "T" for "true" or "F" for "false" depending on which answer better reflects your views. Please answer each question. Don't forget to include your matching number at the top. Be sure you remember the number. Thanks for your time and cooperation.

_____ Age _____ Highest grade completed

_____ Male _____ Female _____ Number of years in present position

- T F 1. Before voting I thoroughly investigate the qualifications of all the candidates.
- T F 2. I never hesitate to go out of my way to help someone in trouble.
- T F 3. It is sometimes hard for me to go on with my work if I am not encouraged.
- T F 4. I have never intensely disliked anyone.
- T F 5. On occasion I have had doubts about my ability to succeed in life.
- T F 6. I sometimes feel resentful when I don't get my way.
- T F 7. I am always careful about my manner of dress.
- T F 8. My table manners at home are as good as when I eat out in a restaurant.
- T F 9. If I could get into a movie without paying and be sure I was not seen I would probably do it.
- T F 10. On a few occasions, I have given up doing something because I thought too little of my ability.
- T F 11. I like to gossip at times.
- T F 12. There have been times when I felt like rebelling against people in authority even when I knew they were right.

- T F 13. No matter who I'm talking to, I'm always a good listener.
- T F 14. I can remember "playing sick" to get out of something.
- T F 15. There have been occasions when I took advantage of someone.
- T F 16. I'm always willing to admit it when I make a mistake.
- T F 17. I always try to practice what I preach.
- T F 18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people.
- T F 19. I sometimes try to get even rather than forgive and forget.
- T F 20. When I don't know something I don't at all mind admitting it.
- T F 21. I am always courteous, even to people who are disagreeable.
- T F 22. At times I have really insisted on having things my own way.
- T F 23. There have been occasions when I felt like smashing things.
- T F 24. I would never think of letting someone else be punished for my wrongdoings.
- T F 25. I never resent being asked to return a favor.
- T F 26. I have never been irked when people expressed ideas different from my own.
- T F 27. I never make a long trip without checking the safety of my car.
- T F 28. There have been times when I was quite jealous of the good fortune of others.
- T F 29. I have almost never felt the urge to tell someone off.
- T F 30. I am sometimes irritated by people who ask favors of me.

- T F 31. I have never felt that I was punished without cause.
- T F 32. I sometimes think when people have a misfortune they only got what they deserved.
- T F 33. I have never deliberately said something that hurt someone's feelings.

APPENDIX B

Matching Number: _____

People differ in the ways they think about those with whom they work. This may be important in working with others. Please give your immediate, first reaction to the items on the following pages.

Below are pairs of words which are opposite in meaning, such as "Very neat" and "Not neat." You are asked to describe someone with whom you have worked by placing an "X" in one of the eight spaces on the line between the two words.

Each space represents how well the adjective fits the person you are describing, as if it were written:

Very neat	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	Not neat
		8		7		6		5		4		3		2		1
		Very		Quite		Some-		Slightly		Slightly		Some-		Quite		Very
		neat		neat		what		neat		untidy		what		untidy		untidy
						neat						untidy				

FOR EXAMPLE: If you were to describe the person with whom you are able to work least well, and you ordinarily think of him/her as being quite neat, you would put an "X" in the second space from the words Very Neat, like this:

Very neat	:	_____	:	X	:	_____	:	_____	:	_____	:	_____	:	_____	:	Not neat
		8		7		6		5		4		3		2		1
		Very		Quite		Some-		Slightly		Slightly		Some-		Quite		Very
		neat		neat		what		neat		untidy		what		untidy		untidy
						neat						untidy				

If you ordinarily think of the person with whom you can work least well as being only slightly neat, you would put your "X" as follows:

X

Very neat : 8 : 7 : 6 : 5 4 : 3 : 2 : 1 : Not neat

Very Quite Some- Slightly Slightly Some- Quite Very

neat neat what neat untidy what untidy untidy

neat neat neat untidy untidy untidy untidy untidy

If you would think of him/her as being very untidy, you would use the space nearest the words Not Neat.

X

Very neat : 8 : 7 : 6 : 5 4 : 3 : 2 : 1 : Not neat

Very Quite Some- Slightly Slightly Some- Quite Very

neat neat what neat untidy what untidy untidy

neat neat neat untidy untidy untidy untidy untidy

Look at the words at both ends of the line before you put in your "X." Please remember that there are no right or wrong answers. Work rapidly; your first answer is likely to be the best. Please do not omit any items, and mark each item only once.

Think of the person with whom you can work least well. He/she may be someone you work with now, or he/she may be someone you knew in the past.

He/she does not have to be the person you like least well, but should be the person with whom you had the most difficulty in getting a job done. Describe this person as he/she appears to you.

Pleasant : 8 : 7 : 6 : 5 4 : 3 : 2 : 1 : Unpleasant

Friendly	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Unfriendly
Rejecting	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Accepting
Helpful	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Frustrating
Unenthusiastic	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Enthusiastic
Tense	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Relaxed
Distant	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Close
Cold	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Warm
Cooperative	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Uncooperative
Supportive	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Hostile
Boring	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Interesting
Quarrelsome	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Harmonious

Self-assured	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Hesitant
Efficient	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Inefficient
Gloomy	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Cheerful
Open	:	<u>8</u>	:	<u>7</u>	:	<u>6</u>	:	<u>5</u>	<u>4</u>	:	<u>3</u>	:	<u>2</u>	:	<u>1</u>	:	Guarded

APPENDIX C

The purpose of this letter is to request your help in completing my thesis which is one of the requirements for my Master's Degree at Western. The objective of my thesis is to help determine factors affecting the way people judge the performance of others. This is an indepth research project, and I would like to borrow some of your time and knowledge in completing this task. If you are like me, you don't like to fill out forms; however, the completion of some forms is necessary for my thesis, but I'll keep them as few and as short as possible.

Specifically, what I would like you to do is to complete the enclosed opinion survey. Since the responses to this and all other forms will remain anonymous, please do not include your name. However, this form must be matched with some others that I would like for you to complete later. To enable me to do this, I would like for you to put a number on this form. It can be any number you like, but be sure you can remember it. For example, you might use your social security number, birthday, etc.

After completing the survey, please put it in the envelop provided and drop it in the mail. As an expression of my gratitude, I will send a letter to the company expressing my thanks and appreciation for your cooperation that will be put in your personnel file. I will also explain all actions and data collected to you at the conclusion of this research project.

Remember, these data will be used for research only. They will not affect your job, nor will you or the company be identified in my thesis.

Thanks very much for your time and cooperation. If you have any questions, contact Ernie Murray or call me collect at 781-6109.

Sincerely,

Danny Bean

Enclosure

APPENDIX D

INSTRUCTIONS: Please read each item carefully, and respond to the items as they describe _____ by placing an "X" in the column on the right that most nearly reflects your opinion.

	STRONGLY AGREE	SOMEWHAT AGREE	UNDECIDED	SOMEWHAT DISAGREE	STRONGLY DISAGREE
1. Keeps the glue pickers clean.....	—	—	—	—	—
2. Lets the glue box run over.....	—	—	—	—	—
3. Keeps bottles clear of the labeler when the strimp stamp machine stops.....	—	—	—	—	—
4. Lets the correct number of bottles go through the labeler.....	—	—	—	—	—
5. Removes bottles that aren't capped.....	—	—	—	—	—
6. Puts the labels in upside down...	—	—	—	—	—
7. Keeps a large enough backlog of bottles to keep the worm from hanging.....	—	—	—	—	—
8. Makes sure that the glue hasn't dried out.....	—	—	—	—	—
9. Prompt in returning to the line from breaks and lunch.....	—	—	—	—	—
10. Makes minor mechanical adjustments herself.....	—	—	—	—	—
11. Often daydreams.....	—	—	—	—	—
12. Makes sure the appropriate label is in the labeler.....	—	—	—	—	—
13. Keeps an adequate number of labels in the labeler.....	—	—	—	—	—
14. Stops the labeler when the labels are crooked.....	—	—	—	—	—
15. Makes sure the labels are on straight.....	—	—	—	—	—
16. Lets the Supervisor or Line Captain know when more labels are needed.....	—	—	—	—	—
17. Constantly watches the labeler...	—	—	—	—	—
18. Makes sure the labeler has enough glue.....	—	—	—	—	—
19. Lets the Supervisor or Line Captain know when a mechanic is needed.....	—	—	—	—	—
20. Delays in restarting the line....	—	—	—	—	—
21. Cleans the machine, work area, etc. when the line is down.....	—	—	—	—	—
22. Stops the labeler if a label is	—	—	—	—	—

		STRONGLY AGREE	SOMEWHAT AGREE	UNDECIDED	SOMEWHAT DISAGREE	STRONGLY DISAGREE
	hung in it.....	—	—	—	—	—
23.	Clears the machine at break times	—	—	—	—	—
24.	Stops the machine without telling the filler operator.....	—	—	—	—	—
25.	Keeps the work area clean.....	—	—	—	—	—
26.	Forgets to remove the cardboard from the pack of labels.....	—	—	—	—	—
27.	Doesn't remove bottles with crooked labels.....	—	—	—	—	—
28.	Lets uncapped bottles go down the line.....	—	—	—	—	—
29.	Doesn't properly start bottles into the worm.....	—	—	—	—	—
30.	Removes labels from the direct drag chain without stopping the machine.....	—	—	—	—	—
31.	Has the ability to work well on more than one line.....	—	—	—	—	—
32.	Makes sure everyone is clear before starting the machine.....	—	—	—	—	—
33.	Often engages in horseplay.....	—	—	—	—	—
34.	Is pleasant to work with.....	—	—	—	—	—
35.	Frequently misses work.....	—	—	—	—	—
36.	Stops the line unnecessarily.....	—	—	—	—	—
37.	Gets along with coworkers.....	—	—	—	—	—
38.	Takes too many breaks.....	—	—	—	—	—
39.	Is in constant need of supervi- sion.....	—	—	—	—	—
40.	Helps others on the line.....	—	—	—	—	—
41.	Complains about working.....	—	—	—	—	—
42.	Overall, this person is a good worker.....	—	—	—	—	—
43.	Completely disregarding every- thing related to work, I like this person.....	—	—	—	—	—
44.	I have enough knowledge of this person's performance to give an accurate judgment.....	—	—	—	—	—

APPENDIX E

INSTRUCTIONS: Please read each item carefully and respond to the items as they describe each of the label operators listed under them by placing an "X" in the column on the right that most nearly reflects your opinion.

	STRONGLY AGREE	SOMEWHAT AGREE	UNDECIDED	SOMEWHAT DISAGREE	STRONGLY DISAGREE
1. Keeps the glue pickers clean:					
Label operator 1.....	___	___	___	___	___
Label operator 2.....	___	___	___	___	___
Label operator 3.....	___	___	___	___	___
Label operator 4.....	___	___	___	___	___
Label operator 5.....	___	___	___	___	___
Label operator 6.....	___	___	___	___	___
Label operator 7.....	___	___	___	___	___
Label operator 8.....	___	___	___	___	___
Label operator 9.....	___	___	___	___	___
Label operator 10.....	___	___	___	___	___
Label operator 11.....	___	___	___	___	___
Label operator 12.....	___	___	___	___	___
2. Lets the glue box run over:					
Label operator 1.....	___	___	___	___	___
Label operator 2.....	___	___	___	___	___
Label operator 3.....	___	___	___	___	___
Label operator 4.....	___	___	___	___	___
Label operator 5.....	___	___	___	___	___
Label operator 6.....	___	___	___	___	___
Label operator 7.....	___	___	___	___	___
Label operator 8.....	___	___	___	___	___
Label operator 9.....	___	___	___	___	___
Label operator 10.....	___	___	___	___	___
Label operator 11.....	___	___	___	___	___
Label operator 12.....	___	___	___	___	___

Note. In an effort to save space, the remaining items will not be presented in this format. The items are, however, the same as those in Appendix D. They also appear in the same order and the same label operators are appraised. These two examples are presented to ensure clear communication to the reader.

APPENDIX F

The purpose of this letter is to introduce the final phase of data collection for my thesis. The same set of conditions that applied to the last questionnaire also apply to this questionnaire. To briefly restate them, the responses will not affect your or anyone else's job, you will remain anonymous due to the matching number, and there are no right or wrong answers.

Specifically, the enclosed questionnaire asks you to judge the performance of twelve workers chosen at random. These workers were selected for no particular reason, except that they operate the labeling machines. Accordingly, the questionnaire focuses on behaviors engaged in by these workers. All questionnaires are identical except for the person being judged and the format of the questionnaire (two formats are being used). I realize that I am asking for a considerable amount of time (about three hours), so I suggest you spread the task over a week's time. Since the first questionnaire you completed must be matched with this one, please put that same "matching number" on each form. Please respond to each item, and then drop the completed questionnaire in the mail. Again, if you have any questions, see Ernie Murray or call me collect at 781-6109.

I would also like to thank you for completing the first questionnaire. I received most questionnaires promptly, and everyone followed the instructions to the "T."

As soon as I complete my thesis, I will reserve the safety room, and at that time, present the results to all who wish to attend. The exact date will be forthcoming.

Sincerely,

Danny Bean

Enclosure

REFERENCES

- Adams, H. F. The good judge of personality. Journal of Abnormal Social Psychology, 1927, 22, 172-181.
- Ashour, A. S. The contingency model of leadership effectiveness: An evaluation. Organizational Behavior and Human Performance, 1973, 9, 339-355.
- Bass, B. M. Reducing leniency in merit ratings. Personnel Psychology, 1956, 9, 359-369.
- Bayroff, A. G., Haggerty, H. R., & Rundquist, E. A. Validity of ratings as related to rating techniques and conditions. Personnel Psychology, 1954, 7, 93-113.
- Berkshire, J. R., & Highland, R. W. Forced-choice performance rating: A methodological study. Personnel Psychology, 1953, 6, 355-378.
- Blum, M. L., & Naylor, J. C. Industrial psychology: Its theoretical and social foundations. New York: Harper & Row, Publishers, 1968.
- Borman, W. C. The rating of individuals in organizations: An alternate approach. Organizational Behavior and Human Performance, 1974, 12, 105-124.
- Borman, W. C., & Dunnette, M. D. Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, 60, 561-565.
- Borman, W. C., & Vallon, W. R. A view of what can happen when behavioral expectation scales are developed in one

- setting and used in another. Journal of Applied Psychology, 1974, 59, 197-201.
- Brown, D. L. Cognitive complexity and intelligence as factors in judgmental processes. Dissertation Abstracts International, 1976 (Feb), 36 (8-B), 4127.
- Bryne, D. Attitudes and attraction. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 4). New York: Academic Press, 1969.
- Buel, W. D. Items, scales and raters: Some suggestions and comments. Personnel Administration, 1962, 27, 15-20.
- Burnaska, R. F., & Hollman, T. D. An empirical comparison of the relative effects of rater response biases on three rating scale formats. Journal of Applied Psychology, 1974, 59, 307-312.
- Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.
- Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 1960, 24, 349-354.
- Cummings, L. L. A field experimental study of the effects of two performance appraisal systems. Personnel Psychology, 1973, 26, 489-502.
- Cummings, L. L., & Schwab, D. P. Performance in organizations: Determinants and appraisal. Glenview, Illinois: Scott,

Foresman & Co., 1973.

- Edwards, K. J. Fair employment and performance appraisal: Legal requirements and practical guidelines. In D. L. DeVries (Chair), Performance appraisal and feedback: Flies in the ointment. Symposium presented at the Annual Convention of the American Psychological Association, Washington, D. C., 1976.
- Elashoff, J. D. Analysis of covariance: A delicate instrument. American Educational Research Journal, 1969, 6, 383-401.
- Fiedler, F. E. A theory of leadership effectiveness. New York: McGraw-Hill, 1967.
- Fiedler, F. E. Validation and extension of the contingency model of leadership effectiveness: A review of empirical findings. Psychological Bulletin, 1971, 76, 128-148.
- Fiedler, F. E. The contingency model--A reply to Ashour. Organizational Behavior and Human Performance, 1973, 9, 356-368.
- Friedman, B. A., & Cornelius III, E. T. Effect of rater participation in scale construction on the psychometric characteristics of two rating scale formats. Journal of Applied Psychology, 1976, 61, 210-216.
- Guilford, J. P. Psychometric methods (2nd ed.). New York: McGraw-Hill Book Company, Inc., 1954.
- Hewitt, J. Liking and the proportions of favorable evaluations. Journal of Personality and Social Psychology, 1972, 22, 231-235.

- Jones, E. E., Gergen, K., & Davis, K. Some reactions to being approved or disapproved as a person. Psychological Monographs, 1962, 76 (Whole No. 521).
- Kallejian, V., Brown, P., & Weschler, I. R. The impact of interpersonal relations on ratings of performance. In L. Wesler & S. F. Harper (Eds.), Performance appraisal: Research and practice. New York: Holt, Rinehart & Winston, 1962.
- Kim, J. Factor analysis. In N. H. Nie, C. H. Hall, J. G. Jenkins, K. Steinbrenner, & D. H. Bent (Eds.), Statistical package for the social sciences. New York: McGraw-Hill, Inc., 1975.
- Kirchner, W. K., & Reisberg, E. J. Differences between better and less-effective supervisors in appraisal of subordinates. Personnel Psychology, 1962, 15, 295-302.
- Klimoski, R. J., & London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.
- Larson, L., & Rowland, K. Stress and leader behavior. Academy of Management Journal, 1975, 18, 397-402.
- Latham, G. P., Wexley, K. N., & Pursell, E. D. Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 69, 550-555.
- Lawshe, C. H., & Balma, M. J. Principles of personnel testing. New York: McGraw-Hill, 1966.

- Lent, R. H., Aurbach, H. A., & Levin, L. S. Predictors, criteria, and significant results. Personnel Psychology, 1971, 24, 519-533.
- Levinger, G. Social desirability in the ratings of involved and neutral judges. Journal of Consulting Psychology, 1961, 25, 554.
- Mandell, M. M. Supervisory characteristics and ratings. Personnel, 1956, 32, 435-440.
- McCall, M. W., & DeVries, D. Appraisal in context: Clashing with organizational realities. Paper presented at the 84th Annual Convention of the American Psychological Association, Washinton, D. C., 1976.
- Meyer, H. H., Kay, E., & French, Jr., J. R. P. Split roles in performance appraisal. Harvard Business Review, 1965, 43, 123-129.
- Mitchell, T. R. Leader complexity and leadership style. Journal of Personality and Social Psychology, 1970, 16, 166-174.
- Nash, A. N. Vocational interests of effective managers: A review of the literature. Personnel Psychology, 1965, 18, 21-37.
- Nash, A. N. Development of an SVIB key for selecting managers. Journal of Applied Psychology, 1966, 50, 250-254.
- Pizam, A. Social differentiation--a new psychological barrier to performance appraisal. Public Personnel Management, 1975, 4, 244-247.

- Schneider, D. E., & Bayroff, A. G. The relationship between rater characteristics and validity of ratings. Journal of Applied Psychology, 1953, 37, 278-280.
- Schneider, C. E. Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. Journal of Applied Psychology, 1977, 62, 541-548.
- Sigall, H., & Aronson, E. Liking for an evaluator as a function of her physical attractiveness and nature of the evaluations. Journal of Experimental Social Psychology, 1969, 5, 93-100.
- Sisson, E. D. Forced choice--the new Army rating. Personnel Psychology, 1948, 1, 365-381.
- Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.
- Strickland, L. H. Need for approval and the components of the ASo score. Perceptual and Motor Skills, 1967, 24, 875-878.
- Stockford, L., & Bissell, H. W. Factors involved in establishing a merit-rating scale. Personnel, 1949, 26, 94-116.
- Stogdill, R. M. Personal factors associated with leadership: A survey of the literature. Journal of Psychology, 1948, 25, 35-71.

Taft, R. The ability to judge people. Psychological Bulletin, 1955, 52, 1-23.

Vroom, V. H. Leadership. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally College Publishing Co., 1976.