

The University of North Carolina
at Greensboro

JACKSON LIBRARY



CQ

no. 1195

UNIVERSITY ARCHIVES

MOYER, STEPHEN MICHAEL. A Comparison of Job Performance Ratings Obtained with Mixed and Unmixed Standard Scales. (1974)

Directed by: Dr. William H. McGehee. Pp. 119.

The present study examined the effects of rating scale format on merit ratings secured with Blanz and Ghiselli's (1972) mixed standard rating system. The study had a dual purpose: 1) to discover whether rater training and controls for faulty item construction could eliminate halo and leniency errors from ratings secured with mixed standard rating scales, and 2) to compare the amount of halo and leniency error in merit ratings obtained with rating scales which were arranged in mixed vs. unmixed formats.

Two rating forms were developed and used to measure the job proficiency of first-line production supervisors in a large textile corporation. Both rating forms were comprised of the same set of behaviorally based statements which pertained to all important aspects of a first-line supervisor's job. On one rating form (Form A), rating scales were arranged in Blanz and Ghiselli's (1972) mixed standard format. On the other rating form (Form B), rating scales were arranged in a more conventional format.

Following a special training session in the use of the mixed standard rating system, department heads ($N = 24$) and upper-level managers ($N = 13$) in eight textile mills rated the job performance of subordinate first-line

supervisors on Rating Form A, Rating Form B, or on both rating forms. Department heads rated a total of 88 first-line supervisors on Form A and upper-level managers used the same rating form to give reliability ratings on 44 of the 88 supervisors. One month later, department heads ($N = 19$) rerated the job performance of 51 first-line supervisors on Rating Form B.

Data analysis showed that 39% of the raters' responses on Form A were logically inaccurate and, therefore, in error according to the mixed standard rating system. Fourteen percent (14%) of the department heads' responses on Form B were logically inaccurate. Because of the high percentage of logical errors in the Form A ratings, no comparison of the amount of halo and leniency error in the Form A vs. Form B ratings was possible. Analysis of the Form B ratings indicated that they contained leniency but not halo error. It was hypothesized that the unexpectedly high incidence of logical errors in the Form A and Form B ratings was primarily attributable to item arrangement and raters' carelessness. The practicality of using mixed standard scales in industrial settings was discussed in the light of the results of this study.

A COMPARISON OF JOB PERFORMANCE RATINGS
OBTAINED WITH MIXED AND UNMIXED
STANDARD SCALES

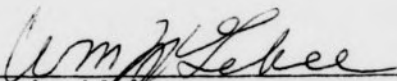
by

Stephen M. Moyer

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
1974

Approved by


Thesis Adviser

APPROVAL PAGE

This thesis has been approved by the following
committee of the Faculty of the Graduate School at The
University of North Carolina at Greensboro.

Thesis Adviser

Wm. M. Lake

Committee Members

Walter T. Enck

Kenneth Smith

March 26, 1973

Date of Acceptance by Committee

ACKNOWLEDGEMENTS

The author gratefully acknowledges the guidance and assistance rendered by Dr. William H. McGehee, adviser, throughout all phases of this research. Sincere thanks are also extended to Drs. Marilyn T. Erickson and Kendon Smith, members of the oral examination committee, for their interest and assistance.

The author wishes to express his appreciation to Dr. William A. Powers, University Statistical Consultant, for his invaluable advice and assistance in analyzing the data obtained in this study.

Special thanks are extended to the management staff of Cone Mills Corporation, without whose complete cooperation the present research would have been impossible.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
INTRODUCTION	1
METHOD	22
Subjects	22
Development of the Rating Forms	23
Development of a Training Guide for Raters	30
Collection of Ratings of Job Performance	31
Administration of Rating Form A	31
Administration of Rating Form B	33
Scoring of Job Performance Ratings	34
RESULTS	36
Analyses of Logical Errors in the Form A and Form B Ratings	38
Reliability of Form A Ratings	44
Comparison of Form A and Form B Rating Scores	44
Regression and Factor Analyses of Form B Ratings	52
DISCUSSION	58
REFERENCES	66
APPENDIX A: Supervisory Interview Form	69
APPENDIX B: Cone Mills Corporation First-line Supervisor Job Performance Rating Form (Form A)	71
APPENDIX C: Cone Mills Corporation First-line Supervisor Job Performance Rating Form (Form B)	85
APPENDIX D: Scoring Sheets for Rating Forms A and B	99

	Page
APPENDIX E. Training Guide for Cone Mills	
First-line Supervisor Job	
Performance Rating Form	102
APPENDIX F: Supplementary Tables	109
Table 1A: Distribution of Age;	
First-line Supervisors.	110
Table 2A: Distribution of Experience;	
First-line Supervisors.	111
Table 3A: Distribution of Education;	
First-line Supervisors.	112
Table 4A: Distribution of Raters'	
Age; Department Heads	113
Table 5A: Distribution of Raters'	
Experience; Department Heads	114
Table 6A: Distribution of Raters'	
Education; Department Heads	115
Table 7A: Scoring Key for Cone	
Mills' First-line Supervisor Job	
Performance Rating Form	116
Table 8A: List of Corresponding	
Statements on Rating Forms A and B.	117

LIST OF TABLES

Table	Page
1. Index of Rating Form Items	26
2. Average Number of Errors Made by Each Rater in Rating First-line Supervisors with Rating Forms A and B	40
3. Average Number of Errors Made by Nineteen Department Heads in Rating Fifty-one First-line Supervisors on Rating Forms A and B	41
4. Percentage of First-line Supervisors Rated Inaccurately on Individual Items of Rating Forms A and B	42
5. Illustration of Two Rating Symbol Patterns Which Accounted for Eighty-six Percent of All Rating Errors	45
6. Means and Standard Deviations of Ratings Obtained on Individual Items for Fifty-one First-line Supervisors Who Were Rated with Rating Forms A and B	46
7. Summary Statistics for the Distributions of Overall Rating Scores Obtained with Rating Forms A and B	51
8. Step-wise Regression Analysis of the Form B Ratings of Fifty-one First-line Supervisors	54
9. Factor Analysis of Twenty-four Rating Form Items Shown to Account for Ninety-nine Percent of the Variance in Overall Rating Scores on Form B	57

LIST OF FIGURES

Figure	Page
1. Distributions of the Form A and Form B Ratings of Fifty-one First-line Supervisors	50

INTRODUCTION

Performance appraisal of management personnel as well as of rank and file workers has been a standard practice in industry for decades (cf. McGregor, 1957). Employees' performance is reviewed periodically in most companies as the basis for decisions about wage and salary administration, promotions, transfers, dismissals, training needs, the predictive or concurrent validity of employee selection tests, the need for employee counseling, etc. Practically all personnel decisions made in industry are based directly or indirectly on the evaluation of job performance.

Measuring job performance is and always has been a major problem in industry (Campbell, Dunnette, Lawler, & Weick, 1970). When jobs are simple and require little if any special skills (e.g., routine production jobs), companies often rely heavily on so-called "objective" or "empirical" measures, like the number of units assembled or the volume of goods produced, to index job performance. However, there are so many variables beyond the control of the individual worker which could affect his production (e.g., the quality and state of repair of the machinery he must work with) that only a limited amount of confidence can be placed in such measures. As jobs become increasingly

complex and require an employee to be adept at a variety of mental as well as physical skills (as in managerial positions), the measurement problem is magnified even further. Empirical indices of performance, such as units of output, are either very hard to obtain or inapplicable for such jobs. The tangible work outputs which can be obtained are usually of questionable value, because they are unlikely to reflect the direct efforts of a single individual. In addition, quantity of work output is often less important than quality in gauging the job effectiveness of employees in managerial positions.

In view of the problems of contamination of most "objective" or "empirical" indices of job performance, there has been a long effort by businessmen and industrial psychologists to devise ways of obtaining subjective measures of job performance which are relevant and reliable. As a result, a number of standardized methods of securing job performance appraisals in the form of personal estimates and subjective opinions (usually those of experts or persons in authority) have been developed and widely used (e.g., Habbe, 1951; Spicer, 1951; Guion, 1970; Gordon, 1970). Evaluations of this type are commonly termed ratings (Ghiselli & Brown, 1955).

Most, if not all, rating methods can be classified into three main categories: 1) ranking procedures, 2) checklists, and 3) rating scales. There are certain

technical advantages and disadvantages connected with each of the three basic rating methods which professionals working in industrial settings must weigh very carefully before implementing one of them in a practical situation. For instance, ranking procedures can be easily designed and administered in an industrial setting, and they provide a very adequate general indication of where one employee stands in relation to others on overall job effectiveness. Unfortunately, rankings do not provide information about specific differences in job performance within and between employees and are, consequently, of little use for diagnostic purposes. Previous authors (e.g., Guilford, 1954; Ghiselli & Brown, 1955; Tiffin & McCormick, 1965; Nunnally, 1967) have discussed at length the relative merits of ranking methods, checklists, and rating scales from both theoretical and utilitarian standpoints. No attempt will be made to review their analyses here, but it can be safely stated that no rating method developed thus far has been found to provide a measure of job performance which is completely satisfactory (Obradovic, 1970). Apart from the technical problems involved in drawing up and administering specific rating procedures in industrial settings, all ratings share a common vulnerability to rater bias, a vulnerability which has caused many psychologists and business professionals to become very cautious about using them as an index of job performance.

Ideally, a high degree of correspondence should exist between an employee's job behavior and the proficiency rating he receives from one or more qualified¹ raters of that behavior. In reality, large rather than small discrepancies are too often found between an employee's job behavior and the job performance rating he receives from independent raters. Some of the inter-observer differences in performance ratings can be ascribed to the facts that ratings necessarily depend on the raters' memory of behaviors observed in the past, that raters differ in their perception of the same stimulus events, and that no two raters usually know a ratee equally well (Ghiselli & Brown, 1955; Guion, 1970; Ghiselli & Ghiselli, 1972). However, a myriad of studies has shown that ratings are frequently influenced by factors which have nothing to do with a ratee's personal qualities or behavior (Guilford, 1954). Two of the most well-documented artifacts affecting job performance ratings in industry are the rating errors of halo and leniency.

Halo refers to a rater's tendency to rate an individual high or low on many aspects of his job performance because he knows or believes that the ratee does one part of his job very well or very poorly. Halo errors occur because people tend to form a general impression of other individuals after meeting and interacting with them

¹A qualified rater is one who has both expert job knowledge and considerable "on-the-job" contact with an employee.

for a short period of time, and this general impression strongly influences the way the behavior of those individuals is subsequently perceived and judged. Thus, if an employer or job supervisor has formed a favorable impression of an employee, he will very likely give the employee a high rating on all aspects of job performance, even if his performance is unsatisfactory or only marginally acceptable in at least some aspects (Ghiselli & Brown, 1955).

Leniency error refers to a rater's tendency to rate individuals higher than they should be rated from an "objective" standpoint (Tiffin & McCormick, 1965). Leniency errors are caused by variables which affect a rater's ability or willingness to report unfavorable impressions of an employee (e.g., the possible consequences that a low rating would have on the rater or the ratee).

Bass (1956) suggests that a lenient rater could be motivated by several considerations:

- 1) He may feel that anyone under his jurisdiction who is rated unfavorably will reflect poorly on his own worthiness.
- 2) He may feel that anyone who could have been rated unfavorably had already been discharged from the organization.
- 3) He may feel that a derogatory rating will be revealed to the ratee to the detriment of relations between rater and ratee.
- 4) He may rate leniently in order to win promotions for his men and, therefore, indirectly increase his future control of his subordinates by earning a reputation as a superior with "influence upstairs."
- 5) He may be projecting.
- 6) He may feel it necessary to always approve others in order to gain approval for himself.
- 7) He may be operating on the basis "whoever associates with me is meritorious, therefore, I am meritorious."
- 8) He may rate leniently because there exists in our culture a response set to approve rather than disapprove [p. 360].

Some of these same ideas had been expressed earlier by Thorndike (1949) and Glickman (1955).

The traditional indices of halo and leniency errors in merit ratings are high² intercorrelations among ratings given to an individual or group of individuals on different traits or job behaviors (e.g., Stevens & Wonderlick, 1934; Ewart, Seashore, & Tiffin, 1941; Roach, 1956; Taylor & Hastman, 1956; Grant, 1955), and negative skewness in the distribution of ratings (e.g., Guilford, 1954; Barrett, Taylor, Parker, & Martens, 1958).

Several authors (Bingham, 1939; Bass, 1956; Ghiselli & Ghiselli, 1972) have pointed out that ratings which are characterized by the statistical properties commonly associated with halo and leniency should not automatically be considered biased. For example, a group of experienced, highly competent employees would be expected to receive high ratings on their performance of specific job duties as well as on overall performance. However, unless there is evidence to indicate that special circumstances prevail in a particular rating situation, psychologists ordinarily expect job performance ratings to show some degree of

²To the writer's knowledge, no specific correlational index has been universally agreed upon by industrial psychologists as the critical value determining whether halo error is or is not prevalent in a set of ratings. The experimental literature seems to suggest, however, that a median correlation of .50 or higher in an intercorrelation matrix is the value at which investigators tend to speak of ratings being marked by halo error.

variability within and between ratees (i.e., reflecting employees' relative strengths and weaknesses), and to be normally or almost normally distributed with respect to ratings given on overall performance (Kelley, 1919; Freyed, 1923; Paterson, 1923). The latter expectation has been challenged by some authors (e.g., Stevens & Wonderlic, 1938; Bass, 1956; Ghiselli & Ghiselli, 1972), who contend that selection practices and early dismissal of unsuitable employees insure that workers who remain on a company's payroll for any appreciable amount of time are at least "average performers." Nevertheless, the trend is for industrial psychologists to continue to consider ratings biased if they appear unreasonably complimentary (i.e., the distribution of ratings shows a high degree of negative skewness) and/or contain high inter-correlations among ratings given on different traits or behaviors.

To date, most attempts at eliminating halo and leniency errors from merit ratings have involved improving previously established rating procedures or inventing entirely new rating procedures.

Training raters so that they are thoroughly familiar with various rating methods as well as with the common rating errors has been claimed to be effective in reducing halo error. Tiffin & McCormick (1965) cite an unpublished study conducted by R. S. Driver at the Atlantic Refining Company of Philadelphia in which raters were given seven

hours of intensive training in rating methods before they were asked to do performance appraisals. The rating results were reportedly uncontaminated by halo error. Regrettably, Tiffin and McCormick (1965) do not report the experimental design or the exact data obtained in the Driver study, so the significance of its findings is difficult to evaluate.

Rater training is also mentioned by Ghiselli and Brown (1955) and later by Ghiselli and Ghiselli (1972) as a possible means of controlling halo, leniency, and other forms of rater bias, but these authors point out that no thoroughly systematic studies have been done with rater training procedures.

Stevens and Wonderlic (1934) demonstrated an effective means of reducing halo which involved a somewhat unconventional method of obtaining ratings with graphic rating scales. Rating instruments were constructed so that raters were compelled to rate all ratees on one job factor before rating them on other job factors. The rating instruments contained graphic rating scales for each of six job factors. After rating data were collected, statistical analysis showed that intercorrelations among the ratings given on the six job factors ranged from $-.14$ to $.39$. The results were interpreted to mean that the ratings had not been affected by halo.

Taylor and Hastman (1956) compared job performance ratings which had been obtained with the Stevens-Wonderlic

(1934) rating procedure and with the more conventional procedure in which individuals are rated separately on all rating scales. The experimenters found no significant differences in ratings collected according to the two procedures. Taylor and Hastman (1956) speculated that the failure to find significant differences could possibly be explained by the facts that the raters were experienced and well-trained, and that the graphic rating scales used were more highly structured than usual (i.e., they contained well-described scale anchors indicating high, average, and low levels of performance for the six job factors).

In the same study, Taylor and Hastman (1956) tested the unverified hypothesis of a number of authors (e.g., Tiffin, 1950; Taylor & Manson, 1951; Guilford, 1954) that merit ratings would be less biased if raters were supervised while they were doing employee evaluations than if they were not supervised. Job performance ratings were obtained under experimental conditions in which the investigators either were or were not present when work supervisors rated their subordinates on six job factors with graphic rating scales. No significant differences were found in the ratings obtained under the two rating conditions. Again, the experience of the raters or the design of the rating scales may have accounted for the failure to find significant differences in the two sets of ratings.

Barrett, Taylor, Parker, and Martens (1958) found that they could influence the degree of halo and leniency in merit ratings by manipulating rating scale content. In their study, four rating instruments were constructed and used to measure the job performance of U. S. Government employees. The instruments were comprised of rating scales arranged in one of four formats which differed in the degree of exactness with which a common set of crucial job traits was described. Format I contained only trait names which were followed by straight lines marked off in segments representing different trait levels. Format II was the same as Format I except that a detailed trait description was provided with each of the trait names. Format III included trait names and behavioral descriptions of the progressive trait levels indicated on each of the rating scales. Format IV contained detailed trait descriptions and behavioral descriptions of scale steps. Job performance ratings were obtained on 507 clerical employees who were randomly assigned to four groups and rated with instruments arranged in one of the four rating scale formats. The results showed lower inter-trait correlations and less negative skewness in the distributions of ratings obtained with Formats III and IV than in those obtained with Formats I and II. It was noted, however, that average intercorrelations were above .50 regardless of the rating scale format with which a set of ratings was collected; and the

distributions of ratings were not significantly skewed for any of the four groups of ratees. Barrett, Taylor, Parker, and Martens' (1958) findings could be taken to mean that providing detailed trait descriptions and behavioral descriptions of scale steps reduces halo by enabling the rater to understand more clearly exactly what he is to rate and in what ways scale positions differ in terms of observables. However, it is difficult to explain why leniency should not be as much enhanced as reduced by trait and scale step descriptions.

A few experimenters (Guilford, 1954; Bass, 1956) have introduced procedures for "post hoc" neutralizing of leniency errors in merit ratings. Guilford (1954) developed a rationale for using a three-way factorial analysis of variance (rater x ratee x trait) to isolate the amount of rating variance in a set of ratings which is attributable to leniency, and then adjust ratings so that leniency is extracted. According to Bass (1956), Guilford's procedure can be used very effectively as long as enough ratees have been rated by the same raters to permit the analysis of variance of rating data to be carried out. However, the latter requirement is sometimes hard to meet in practical situations, and competent statisticians are not always available to do the necessary data analysis.

In 1956, Bass presented a newly developed procedure for scoring Likert-type (i.e., graphic) rating scales (Likert, 1932) which turned out to be only marginally

successful in neutralizing leniency errors in merit ratings. The scoring system was developed by having sales managers from the grocery industry rate their best and their worst salesmen on rating scales dealing with the frequency with which certain critical job behaviors were emitted. Rating scale positions at which significantly more superior than inferior salesmen were rated were subsequently assigned a weight of 1. Other scale positions were weighted with a zero. After discriminant binary weights were established for all scales, 350 grocery salesmen were rated on a Sales Behavior Inventory comprised of the rating scales for which scoring weights had been determined. Rating scores were derived by summing the number of times a salesman was rated in a scale position having a binary weight of 1. Alternate rating scores were derived on the same set of ratings by using an older method of scoring graphic scales which was originally devised by Likert (1932). The Likert scoring procedure involves the simple addition of the arbitrary weights corresponding to the positions at which a rater places checks on 1 to N-point rating scales. The results showed that the degree of skewness and kurtosis in the ratings were less when they were scored according to Bass's (1956) scoring system rather than Likert's (1932), but the correlation between the rating scores derived with each of the two scoring methods exceeded .90. Furthermore, negative skewness was still found in both distributions of rating scores.

Bass (1956) concluded that in terms of the purpose for which his scoring system was developed, it was little better than the less-sophisticated Likert (1932) system.

Disguising the order-of-compliment of statements describing personal traits or ways that employees go about doing their jobs is a technique which has been used with varying amounts of success in minimizing (but not totally preventing) halo and leniency errors in job performance appraisals. Perhaps the most familiar rating procedure utilizing this technique is the forced-choice rating method introduced by Sisson (1948). In forced-choice ratings, the rater is usually shown a series of pairs or triads of trait names and job-behavior descriptions which appear equally favorable or unfavorable. The rater is instructed to check the trait name or behavior description in each pair or triad which is most applicable to the ratee. Each alternative has a scale value predetermined by means of pilot research which is known only to the person(s) administering and scoring the ratings. The final rating score is obtained by summing the scale values of the trait names and behavior statements checked as most closely describing the ratee.

Sisson (1948) administered forced-choice rating instruments to over 5,000 U. S. Army officers and reported that, in comparison to the performance appraisals obtained at an earlier time with graphic rating scales, the forced-choice ratings were "relatively free of the usual pile up at the top end of the scale."

In a later study, Taylor and Wherry (1951) demonstrated that forced-choice ratings were less subject to change in terms of leniency than graphic scale ratings when job performance appraisals were obtained first under "experimental" and then "for real" rating conditions. The results indicated that forced-choice ratings were more resistant to intentional rater bias than ratings obtained with graphic scales.

Travers (1951) strongly questioned Sisson's original (1948) statements about the superiority of the forced-choice technique and suggested that Sisson's data do not support his argument in favor of the forced-choice method. Travers (1951) further stated that a rater could steer forced-choice ratings in a favorable direction if he so desired by simply rating an ideal subordinate rather than the one he was assigned to rate. In addition, he claimed that most distributions of forced-choice ratings still contain a high degree of negative skewness.

Travers' (1951) contention that forced-choice ratings could be intentionally biased was clearly supported in a study conducted by Berkshire and Highland (1953). These experimenters used a forced-choice rating instrument to obtain two sets of job proficiency ratings on a group of U. S. Air Force technical school instructors. For the first set of ratings, instructor-supervisors were given normal instructions and asked to rate their subordinates as fairly as possible with the rating instrument provided. For the

second set of ratings, instructor-supervisors were asked to rate their subordinates as though each was the rater's best friend and he wanted to assure him the highest rating possible. The results showed that the proficiency ratings were much higher under the "requested bias" condition. In general, it can be safely stated that the forced-choice rating method has not been shown to be consistently more effective than any other rating procedure in preventing halo and leniency errors in job performance ratings.

In a fairly recent study, Blanz and Ghiselli (1972) introduced a new system of merit rating, the Mixed Standard Scale, which was designed to reduce the common rating errors of halo and leniency. The experimental manipulations included in the system were not new either conceptually or from the standpoint of previously published research on the reduction of halo and leniency errors (e.g., providing behavioral descriptions of scale steps [Barrett, Taylor, Parker, & Martens, 1958]; disguising the order-of-compliment in the rating scales [Sisson, 1948]), but the means of accomplishing the manipulations were unique and afforded the system certain advantages over former methods of merit rating.

The Mixed Standard Scale consists of an array of randomly arranged statements, each belonging to one of several triads of statements which describe progressive levels of proficiency on factors considered crucial to successful job performance. The rater is asked to respond

to each statement as it occurs on the rating form by putting a plus (+) by the statement if the ratee's standing or performance is superior to that described by the statement, a zero (0) by the statement if the ratee's performance is aptly described by the statement, or a minus (-) by the statement if the ratee's performance or standing is worse than or inferior to that described by the statement. In this way, statements describing different job factors are judged separately and not in the context of other statements describing alternate levels of proficiency of the same job factors. The order-of-merit in related statements is not readily apparent and the likelihood of intentional rater bias is thereby reduced.

After a rater has completed a rating form, its format is rearranged for scoring with the various triads of related statements being grouped together according to order-of-merit. Ratings on each job factor can then be examined in terms of response patterns which are logically consistent with one another. Blanz and Ghiselli (1972) explain that:

If a rater utilizes the procedure accurately, then whenever he checks one statement in a scale as "fits the ratee" (0), all statements in that scale which describe superior behavior will be checked as "the ratee is poorer than the statement" (-), and all those which describe inferior behavior will be checked as "the ratee is better than the statement" (+). If all three statements in a scale are checked +, it means that in the rater's opinion the ratee is very good in the trait, for his performance is superior even to the very best of the three descriptions. Similarly, if all three statements in a scale are checked -, it means that in the rater's opinion the ratee is very poor, for his performance is inferior even to the very poorest of the three descriptions.

With the three graded statements used in this manner, there is actually a 7-point scale on each trait, which also is an improvement on ordinary rating scales. Pursuant to the logic of the system, the various combinations of faultless responses to the items can be arranged as follows, and can be assigned the number of points indicated.

<u>Statements</u>			<u>Points</u>
<u>I</u>	<u>II</u>	<u>III</u>	
+	+	+	7
0	+	+	6
-	+	+	5
-	0	+	4
-	-	+	3
-	-	0	2
-	-	-	1

The foregoing combinations are faultless because there are no reversals in the order with which the three graded descriptions are checked. That is, whenever a statement is checked 0, no statement which describes better performance is checked either 0 or +, and no statement which describes inferior performance is checked either 0 or -. Furthermore, 0, which means the statement fits the ratee, is not employed for two or more statements which describe degrees of the trait. All combinations of responses to the three statements other than the seven given above are illogical and inconsistent, and therefore in error. Nevertheless, the logic of the system permits such scales to be scored.

<u>Combinations</u>			<u>Points</u>
<u>I</u>	<u>II</u>	<u>III</u>	
+	+	0	7
+	+	-	7
0	+	0	6
0	+	-	6
-	+	0	5
-	+	-	5
0	-	+	5
0	0	-	4
+	0	+	4
+	0	-	4
0	0	0	4
-	0	-	3
+	-	+	3
+	0	0	3
0	-	0	2
+	-	0	2
+	-	-	1
0	-	-	1

The special advantages of Blanz and Ghiselli's system of merit rating lie in the adaptability of the Mixed Standard Scale to jobs of different types in a variety of settings, and in the wealth of information which can be gained from the systematic analysis of logical errors in merit ratings obtained with the scale.

Errors in logical consistency can be used as separate indices of the reliability of measurement for each rater, for each ratee, and for each factor on which ratings are obtained. Ordinarily, a single reliability measure, which includes rater, ratee, and scale, is determined for the entire rating process. However, if desired standards are not met, the specific source of unreliability is not as easily detected from a single reliability measure as it is from the multiple reliability index provided by the Mixed Standard Scale.

Blanz and Ghiselli's (1972) study was concerned with the applicability of the mixed standard scale to managerial jobs. A rating form was developed which measured performance on 18 job factors, and performance ratings were subsequently obtained on 100 middle-level managers. A factor analysis of the rating results yielded four distinguishable factors, indicating that the high inter-scale correlations characteristic of halo error were not prevalent in the ratings. Further statistical analyses of the rating data indicated that leniency error, although diminished, had not been

eliminated by the mixed standard scale. If there had been no error of leniency, the mean rating for separate job factors and for all factors combined would have fallen at about the midpoint of Blanz and Ghiselli's seven-point scale (i.e., at 4.0), and individual ratings would have been symmetrically distributed about the mean. Instead, mean ratings for separate job factors ranged from 4.3 to 5.3, and the mean rating for all factors combined was 4.7. In addition, a graphic representation of the distribution of ratings showed a slight skewness toward the lower end of the scale. The authors suggested that the slightly inflated ratings were more likely a case of realistic description than rater bias, inasmuch as the ratees were managers of considerable experience; but variables other than ratees' job performance could have accounted for, or at least contributed to, the same rating results.

It is possible that some rating scales contained confusing statements which caused raters to make logically inconsistent responses which were scored with high rather than low scale values. For example, if a rater responded to a triad of statements describing increasing levels of job proficiency with the symbols 0, +, +, respectively, the ratee would be given a rating score of 7, even though the pattern of responses is internally inconsistent. If the pattern of responses to the same triad of statements was 00+, which is also logically inconsistent, the ratee

would be given a rating of 4. The two rating patterns cited above could occur if the statements describing the lower and intermediate levels of performance in a triad were unclear or not easily distinguishable from one another. Blanz and Ghiselli's (1972) data showed that statements describing certain job factors were rated inconsistently 20 to 30 percent of the time, and that job factors on which a large number of logical errors occurred also received some of the highest mean ratings. The positive relationship between logical errors and high mean ratings existed in only a few instances, however.

Raters' relative inexperience in applying the rating symbols to a set of descriptive statements is another variable which could have caused raters to make logically inconsistent responses in Blanz and Ghiselli's (1972) study. A rater training procedure designed to minimize rating errors due to raters' unfamiliarity with the rating system was evidently not included in the study.

The extent to which logical errors contributed to the negative skewness in the distribution of performance ratings obtained by Blanz and Ghiselli (1972) cannot be determined from the data reported in their study. However, to the extent that sources of spuriously high rating scores were not controlled as strictly as possible in the study, it may be questioned whether the full potential of the mixed standard scale for reducing halo and leniency errors has yet

been demonstrated. A further question left open to speculation by Blanz and Ghiselli's (1972) study is "To what degree is the mixed standard scale either more or less effective than some other scale in reducing halo and leniency errors in merit ratings?" Thus far, no direct comparisons have been performed experimentally.

The present study attempted to answer the questions raised in the foregoing paragraph. Specifically, the study was designed to: 1) discover whether introducing methodological improvements in the development and administration of a mixed standard scale would reduce halo and leniency errors in merit ratings even more effectively than they had been reduced in Blanz and Ghiselli's (1972) study; and 2) compare the amount of halo and leniency error in job performance ratings obtained with rating scales arranged in the mixed standard format, and rating scales arranged in a format in which descriptive statements were not mixed as prescribed by Blanz and Ghiselli (1972).

METHOD

The study was conducted through the Industrial Relations Department of Cone Mills Corporation, Greensboro, North Carolina. The rating instruments used in the study were developed as performance criteria for use in the company's research in test validation.

Subjects

The raters used in the study were 24 department heads from the Spinning, Weaving, and Dyeing Departments of eight Cone Mills' plants, as well as 13 higher level managers from the same plants who provided reliability ratings for one of two rating conditions.

The ratees were 88 first-line production supervisors who represent the very first level of company management in the plants. Each first-line supervisor was rated on a mixed standard scale (see Appendix B) by his immediate superior (i.e., department head) and, where possible, by a higher-level manager to provide necessary data for a company test-validation study. "Reliability" ratings were obtained on 44 of the 88 first-line supervisors who were rated on the mixed standard scale.

At a later date, 51 of the 88 first-line supervisors were rerated by their department heads ($N = 19$) with an

unmixed standard scale (see Appendix C) to provide data for the present study.

The length of time that department heads and reliability raters had supervised ratees ranged from one month to 18 years (the mean was 5 years, 6 months) and from two months to 22 years (the mean was 5 years, 2 months), respectively. Because it would be difficult to say how long a rater should supervise a subordinate before he is capable of giving him a fair rating, the experimenter considered all department heads and higher level managers qualified to do ratings unless they verbally specified otherwise. The distributions of age, experience, and education for both raters and ratees appear in Appendix F.

Development of the Rating Forms

As the first step in developing the rating instruments, department heads, plant superintendents, plant managers and others deemed sufficiently knowledgeable of the job of first-line production supervisor in each of eight plants were contacted and scheduled for individual interviews. The interviews were conducted by the experimenter and two other members of the company's personnel research staff. Each interview lasted approximately one hour, and a total of 46 individuals was interviewed.

During the interviews, department heads and knowledgeable persons from other levels of management were asked to specify verbally the behaviors which they felt could be used

to discriminate good from poor first-line supervisors. The responses of each interviewee were written down on a specially designed data collection form (see Appendix A). Individual interviews were used for purposes of collecting items instead of a group meeting (Smith & Kendall, 1963; Landy & Guion, 1970; Folgi, Hulin, & Blood, 1971) because of the scheduling problems which would be involved in the latter technique and because it was not considered advisable to remove all upper-level supervisors from their work stations simultaneously.

After all interviews were completed, a cumulative list of 85 different (or apparently different) job behaviors was compiled from the data supplied by individual contributors. This cumulative list was then resubmitted to each of the 46 interviewees by mail for evaluation. Participants were asked to review the list and assign from 0 to 10 points to each job behavior according to how important they felt the behavior was to supervisory success. They were told that assigning zero points (0) to a behavior would indicate that it was not considered relevant to job success; whereas assigning 10 points to a behavior would indicate that it was considered crucial to job success. All judges were asked to weight items and return the cumulative list to the Industrial Relations Department within ten days of its receipt. The weighting procedure was carried out to find out which items, if any, should be deleted from the list.

Inspection of the results after all copies of the cumulative list were returned showed that, for the most part, judges' weightings fell within a restricted range of from 7 to 10 points for each job behavior. The mean weightings on the individual behaviors ranged from 6.7 to 9.8, with a median of 9.1. Thus, all behaviors were generally given high weightings by most of the 46 judges. The zero weightings assigned to specific behaviors were widely scattered and no consistent pattern among them could be determined. It was finally decided that the item pool could be reduced to a reasonable number for the construction of a mixed standard scale if all job behaviors which had received a mean rating of less than 9.0 were dropped from the list. On the basis of this arbitrarily selected cut-off value, 41 behaviors were removed. Of the remaining 44 critical job behaviors, six were judged to be duplications of others on the list and were also removed. The final item pool thus consisted of 38 job behaviors, each showing a very strong relationship to successful job performance according to the judges' consensus, plus an added statement referring to overall job performance. The 39 items are listed sequentially in Table 1 according to a code number which was assigned to each item.

Next, statements reflecting three levels of performance for each critical job behavior were written and two rating forms were constructed. Each of the rating

TABLE 1
Index of Rating Form Items

Item Number	Description of Critical Job Behavior
1	Sets job goals.
2	Cooperates with fellow supervisors.
3	Responds positively to superiors instructions.
4	Keeps boss informed.
5	Helps employees with job problems.
6	Expresses honest opinions to boss; communicates totally.
7	Keeps relief supervisor informed.
8	Reprimands employees in private, not in public.
9	Prevents waste; corrects wasteful job practices.
10	Possesses and demonstrates technical knowledge of machinery in his department.
11	Discusses production problems with employees in a non-accusing manner.
12	Listens to employees' personal problems; counsels or refers when possible.
13	Refrains from spreading gossip.
14	Listens attentively to superiors' instructions.
15	Patrols and follows up on jobs in his work area.
16	Complies with personnel policies.
17	Makes decisions he is responsible for making.
18	Follows up on defective materials returned to his department. Corrects defects.
19	Follows up on defective materials entering his department. Contacts supervisor of previous department.
20	Corrects employees' unsafe work habits.
21	Makes sure preventive maintenance is performed on department machinery.
22	Wears required protective equipment.
23	Has safety hazards corrected.
24	Follows up on new employees to insure proper training.
25	Keeps work area clean.
26	Makes sure employees wear required protective equipment.
27	Gives employees regular feedback on job performance.
28	Reports to work properly dressed and groomed.
29	Maintains production quota expected of his section.
30	Admits and takes responsibility for personal errors.
31	Knows and demonstrates thorough knowledge of inter-departmental production network.
32	Checks machine settings; makes sure they comply with production standards.

TABLE 1 (Continued)

Item Number	Description of Critical Job Behavior
33	Follows up on absenteeism in his department.
34	Plans work thoroughly and systematically.
35	Relies on his own authority when assigning work.
36	Talks positively about company in front of employees.
37	Assigns work in a courteous manner; not gruff or commanding.
38	Encourages constructive comments and suggestions from employees.
39	General performance appraisal.

forms was comprised of the same 39 triads of behaviorally based statements (i.e., 117 separate statements), but the arrangement of statements on the forms differed. One rating form, Form A, was patterned after Blanz and Ghiselli's (1972) Mixed Standard Scale. Statements were distributed on a 12-page form in a semi-random order so that components of the same triad were spaced at least one or two pages apart from each other, with 9 to 12 statements in all appearing on each page. The other rating form, Form B, was arranged in a standard rating scale format with related statements appearing together in clearly distinguishable triads. Statements describing different levels of job performance were randomly arranged within triads on Form B.

A nearly identical set of rating instructions was attached to both rating forms. Raters were instructed to use Blanz and Ghiselli's "plus-zero-minus" rating system. The only difference in the instructions for completing the rating forms was that the Form B instructions contained an additional reminder to the rater to place one of the three rating symbols in front of each and every statement on the form. The additional reminder was intended to eliminate the foreseeable danger that raters would begin putting a checkmark or zero next to only one statement in each triad of statements instead of responding to all statements on the rating form.

Before the rating instruments were actually used to collect performance ratings, their statements were

subjected to some final checks by 12 experienced, highly regarded department heads and upper level managers from plants not participating in the test validation project of which the rating-form development was a subtask. Each individual was asked, first of all, to read over each statement on Rating Form B (on which job behavior statements appeared together in their respective triads) and point out in writing which statements or parts of statements were ambiguous or unclear and to explain why they were unclear. Secondly, he was instructed to read the triads of statements pertaining to each critical job behavior and to place a 1 in front of the statement which seemed to describe the lowest level of performance, a 2 in front of the statement which seemed to describe the next highest level of performance, and a 3 in front of the statement which seemed to describe the highest level of performance. Inasmuch as statements were randomly arranged within triads, it was necessary for reviewers to read carefully and to pay close attention to the content of each statement. Finally, they were asked to comment on the rating form generally (i.e., as to its relevance, etc.)

After feedback had been obtained from each of the 12 outside reviewers, slight alterations were made in some of the behavioral statements to make them more clear. The distinctiveness of the three levels of performance in the triad of statements written for each critical job behavior

was confirmed by the fact that all 12 reviewers interpreted the performance-level statements to mean exactly what the experimenter had intended them to mean. In addition, the 12 reviewers concurred that the content of the rating form included all important aspects of the job and should give a valid measure of job performance.

Development of a Training Guide for Raters

Since raters' unfamiliarity with, and inexperience in applying Blanz and Ghiselli's (1972) rating system was recognized as a potential source of error in the rating of job performance, a training booklet for raters was developed to prevent such errors from occurring (see Appendix E).

The training guide contained a series of five anecdotes which described the way that some fictitious supervisors went about doing their jobs. Following each anecdote, there appeared three to five behavior statements which were relevant to the content of the anecdote and similar to those on the actual rating forms. In addition, the job behavior statements in the training guide described only one or two levels of performance for any job behavior so that the internal logic of the rating system would not be inadvertently disclosed during the training of raters.

Raters were instructed to read each anecdote and then respond to the subsequent job behavior statements using the plus-zero-minus rating system devised by Blanz and Ghiselli (1972). Since only one rating symbol could be

"correct" according to the information given in the anecdote, it was possible for both the rater and the experimenter to evaluate the rater's understanding of the rating system and his ability to use it accurately.

Collection of Ratings of Job Performance

In order to compare the amount of halo and leniency error in the job performance ratings of individuals who were rated with both a mixed and an unmixed standard scale, it was necessary to have first-line supervisors rated with Form A before they were rated with Form B. There were two reasons for this: 1) administering Form B prior to Form A would disclose the internal logic of the rating system to the raters and enhance their ability to rate leniently if they so desired; and 2) splitting the raters into two groups and counterbalancing the order in which rating forms were administered would have been unwise, because the Form A ratings obtained were to be used as a criterion in the company's test-validation project. For purposes of validation research, it was desirable to have all first-line supervisors rated under the same conditions.

Administration of Rating Form A. Before a rater was asked to complete any of the job-performance rating forms, he was requested to meet with the experimenter for a training session which lasted approximately one half hour. During that time, the department head or reliability rater

was given a Rater Training Guide to read and fill out as directed. When the training guide had been completed, the rater's responses were checked and discrepancies between the rater's responses and a Rater Training Guide Answer Key were thoroughly discussed so that the rater would fully understand why his incorrect responses were in error. It was discovered that the "correct" answers to one or two of the items in the Rater Training Guide were debatable because of the wording which appeared in some of the anecdotes. Consequently, alternate responses were accepted as correct on some items if the rater could justify his answer. Finally, the rater was given the rating forms for the subordinates whose job performance he was to evaluate. He was asked to read the directions on one of the forms silently while the experimenter read them aloud. When the instructions had been read, the rater was encouraged to ask questions about anything that might be unclear to him. If the rater had no questions, he was reminded of the importance both of taking enough time to do the ratings carefully and of the confidentiality with which the ratings would be treated. He was then asked to complete the rating forms within a ten-day period without conferring with others in any way and to return them to the company's Industrial Relations Department. Raters were given the option of not rating a subordinate if they felt that they did not know him well enough to give a fair and accurate rating.

was given a Rater Training Guide to read and fill out as directed. When the training guide had been completed, the rater's responses were checked and discrepancies between the rater's responses and a Rater Training Guide Answer Key were thoroughly discussed so that the rater would fully understand why his incorrect responses were in error. It was discovered that the "correct" answers to one or two of the items in the Rater Training Guide were debatable because of the wording which appeared in some of the anecdotes. Consequently, alternate responses were accepted as correct on some items if the rater could justify his answer. Finally, the rater was given the rating forms for the subordinates whose job performance he was to evaluate. He was asked to read the directions on one of the forms silently while the experimenter read them aloud. When the instructions had been read, the rater was encouraged to ask questions about anything that might be unclear to him. If the rater had no questions, he was reminded of the importance both of taking enough time to do the ratings carefully and of the confidentiality with which the ratings would be treated. He was then asked to complete the rating forms within a ten-day period without conferring with others in any way and to return them to the company's Industrial Relations Department. Raters were given the option of not rating a subordinate if they felt that they did not know him well enough to give a fair and accurate rating.

Reliability ratings were not obtained on two supervisors because the raters exercised this option.

Administration of Rating Form B. Approximately one month after the initial performance rating of 88 first-line supervisors on Rating Form A, 51 of the supervisors were re-rated by their department heads ($N = 19$) on Rating Form B. The number of ratees and raters was reduced for the second rating because the experimenter did not want to risk overburdening department heads with additional paperwork and because a project deadline had to be met. Consequently, fewer Form B ratings were secured from a smaller number of department heads so that the rating data could be collected and analyzed within a relatively short period of time. No attempt was made to obtain reliability ratings on Form B because of the additional time which would have been involved in collecting, scoring, and analyzing additional ratings.

The administration of Form B ratings differed from that of Form A ratings. The experimenter contacted each of 19 department heads by phone and explained that he had heard from some department heads in other plants that Rating Form A had been a difficult instrument to use. Consequently, the Industrial Relations Department was seeking the cooperation of a small group of department heads to help to determine whether a revised edition of the rating form would be both easier to use and provide equally acceptable measures

of job performance. The department head was told that it would be necessary to rerate three of his subordinates (or fewer if fewer had been rated initially) to provide the data needed to compare results. If the department head agreed to participate, as each did, he was told that he would receive the revised rating forms (i.e., Form B) in the mail within the next few days. He was instructed to read the rating directions carefully, complete the forms, and then return them to the Industrial Relations Department within ten days of their receipt.

The three first-line supervisors that each department head rerated on Rating Form B were randomly selected by the experimenter. If a department head had rated fewer than three supervisors on Rating Form A, he merely rerated the same supervisors on Rating Form B.

Scoring of Job Performance Ratings

Both rating forms (i.e., Form A and Form B) were scored according to the system devised by Blanz and Ghiselli (1972) which was described earlier. It should be noted, however, that the tables of rating scores for individual patterns of rating symbols which Blanz and Ghiselli (1972) presented in their study did not include two of the possible permutations of three rating symbols (i.e., +00, -00). Consequently, a score of 3 was assigned to both patterns for scoring purposes in the present study. Since the patterns indicate that a rater feels that statements describing low

and average levels of performance most aptly describe a ratee, a score was assigned which was halfway between the scores of logically consistent responses containing a zero in the "low" or "average" performance positions.

After the rating forms were scored according to Blanz and Ghiselli's (1972) scoring system, an overall performance score was derived for each rating form by summing the rating scores a ratee received on individual items. Thus, the maximum overall performance score a first-line supervisor could receive was 273, and the minimum score he could receive was 39.

RESULTS

When the Form A and Form B ratings were scored, an unexpectedly large number of logically inaccurate responses to the rating form statements were discovered. In order to score the forms, the rating symbols which raters had placed in front of individual rating statements were transferred to specially designed score sheets (see Appendix D). On the score sheets, rating symbols were regrouped into their respective triads and the internal logic of a rater's responses to the separate triads was revealed. A triad of three rating symbols was considered one response in the data analysis and, hereafter, the terms "logically inaccurate response" or "logical error" will be used to refer to rating symbol patterns in which the ordering of symbols is internally inconsistent (and, therefore, in error) according to Blanz and Ghiselli's (1972) rating system.

On Form A, 39% of the department heads' responses to the rating form statements were logically inaccurate. The Form A ratings given by reliability raters contained the same percentage of logically inaccurate responses.

On Form B, 14% of the raters' responses were logically inaccurate. The percentage of logical errors in the Form B ratings is remarkably high considering that the arrangement

of statements on this form lends itself to perfect accuracy of rating.

Because of high percentages of logical error contained in the Form A and Form B ratings, no sound conclusions about the relative amounts of halo and leniency in mixed (Form A) and unmixed (Form B) standard scale ratings can be drawn from the data obtained in this study. The extent to which the statistical indices of halo and leniency in the ratings would be attributable to artifacts associated with Blanz and Ghiselli's (1972) system of scoring logical errors would be very difficult to determine.

Despite the fact that the planned comparisons of the relative amount of halo and leniency errors in the Form A and Form B ratings were no longer viable, the rating data were analyzed and compared in several ways which provided much useful information. First of all, the ratings obtained with both forms were subjected to various error analyses which were aimed at uncovering as much information as possible about the ways in which rating form format and other variables (e.g., faulty wording in rating form statements) affected the logical accuracy of the ratings. The type and outcome of each error analysis will be described later in this section. Secondly, a measure of "inter-rater agreement" was calculated for the Form A ratings given to 44 first-line supervisors by both department heads and reliability raters, even though the practical significance

of such a measure was questionable under the circumstances. The index of inter-rater agreement was calculated for purposes of inspection since both the department heads' and the reliability raters' Form A ratings contained 39% logical error. Next, indices of halo and leniency were computed for the Form B ratings since it is conceivable that 14% rating inaccuracy could be considered a tolerable margin of error by some experimenters. Finally, for purposes of inspection and comparison, a complete analysis of the distributions of rating scores obtained when first-line supervisors were rated with Form A as opposed to Form B was carried out.

Analyses of Logical Errors in the Form A and Form B Ratings

A Z-test of single proportions³ (Wyatt & Bridges, 1967) was performed on the relative frequency of rating inaccuracies in the department heads' Form A and Form B ratings of first-line supervisors. In each case, the percentage of rating inaccuracies was found to be highly significant ($P < .001$). Similarly, a Z-test was performed on the difference in the percentages (i.e., 41% vs. 14%) of rating inaccuracy contained in the ratings of 51 first-line supervisors who were rated by their department

³Technically, the use of the Z-test of proportions presumes that each sample observation (i.e., each item rating) is independent. Individual item ratings were not totally independent in the present data since many ratings were given by a single individual. However, the Z-test was employed here because it seemed the most nearly appropriate test for the rating data.

heads on Rating Forms A and B. The difference in the two percentages was also statistically significant ($P < .002$, $Z = 3.07$).

There were substantial differences among the raters in terms of the degree of accuracy with which they rated subordinates on one or both of the two rating forms. Table 2 shows the average number of errors per ratee made by each rater (i.e., department heads and reliability raters) in rating first-line supervisors with Form A, Form B, or both rating forms. On Form A, the most accurate rater averaged seven errors per ratee and the least accurate averaged 25.7 errors per ratee. On Form B, three raters made no errors and two raters averaged as many as 12 errors.

A further indication that raters were much more accurate when they did ratings with Form B rather than Form A is presented in Table 3. The table shows the average number of errors per ratee made by each of 19 department heads who rated two or three of their subordinates on both rating forms. All department heads except one (No. 11) showed large differences in the average number of rating inaccuracies contained in their Form A vs. Form B ratings.

The percentage of first-line supervisors who were rated inaccurately on each of the 39 items on Rating Forms A and B is presented in Table 4. The data are reported in four separate columns which represent the department heads' Form A ratings of 88 first-line supervisors, the

TABLE 2

Average Number of Errors Made by Each Rater in Rating
First-line Supervisors with Rating Forms A and B

Rater ID No.	Number of Supervisors Rated		Mean Number of Errors per Ratee	
	Form A	Form B	Form A	Form B
1	3	3	17.0	12.6
5	5	2	11.4	0.5
9	7	3	11.3	2.3
11	3	2	13.0	12.5
13	7	3	8.4	0
17	6	3	17.1	11.0
19	4	3	17.5	7.0
23	4	3	19.7	2.3
25	2	2	15.0	4.0
27	3	3	15.0	14.0
29	5	3	16.8	1.3
31	2	2	16.5	6.0
37	3	3	19.0	7.6
39	3	2	17.0	10.5
43	2	2	7.0	0
45	3	3	23.6	2.6
49	4	3	21.5	5.6
51	3	3	25.7	5.0
65	5	3	11.8	0
3	1	-	22.0	-
7	1	-	20.0	-
42	4	-	9.8	-
15	5	-	14.4	-
2	2	-	13.0	-
10	3	-	13.7	-
14	3	-	16.3	-
16	7	-	22.3	-
20	3	-	21.7	-
26	6	-	15.6	-
28	4	-	10.7	-
46	1	-	22.0	-
48	2	-	11.5	-
50	2	-	17.0	-
52	3	-	15.3	-
54	1	-	12.0	-
62	7	-	11.6	-

Note.--All even numbered raters except #42 were reliability raters.

TABLE 3

Average Number of Errors Made by Nineteen Department Heads
in Rating Fifty-one First-line Supervisors on Rating
Forms A and B

Rater Number	Number of Supervisors Rated	Average Number of Errors per Ratee	
		Form A	Form B
1	3	17.0	12.6
5	2	16.0	0.5
9	3	5.3	2.3
11	2	12.5	12.5
13	3	10.0	0.0
17	3	14.6	11.0
19	3	17.6	7.0
23	3	20.6	2.3
25	2	15.0	4.0
27	3	15.0	14.0
29	3	18.3	1.3
31	2	16.5	6.0
37	3	19.0	7.6
39	2	16.5	10.5
43	3	7.0	2.6
45	3	21.5	5.6
49	2	20.3	0.0
51	3	25.6	5.0
65	3	11.6	0.0

TABLE 4

Percentage of First-line Supervisors Rated Inaccurately
on Individual Items of Rating Forms A and B

Item Number	Form A (N=88) ^a	Form A _b (N=44) ^b	Form A _c (N=51) ^c	Form B _c (N=51) ^c
1	58	68	61	39
2	50	55	47	18
3	52	66	57	18
4	39	55	43	18
5	43	75	39	29
6	30	36	25	20
7	31	39	31	14
8	45	44	45	18
9	56	57	59	10
10	28	32	25	0
11	26	21	22	12
12	63	62	61	22
13	23	27	23	10
14	13	30	12	8
15	20	19	22	8
16	8	16	6	6
17	52	50	55	16
18	51	55	61	20
19	26	33	25	2
20	69	62	76	39
21	49	35	61	14
22	20	23	23	2
23	60	57	71	14
24	47	48	47	0
25	50	46	57	10
26	38	48	43	16
27	72	71	71	43
28	4	5	6	0
29	31	32	31	10
30	38	28	15	8
31	67	60	29	23
32	47	37	72	4
33	66	48	45	39
34	18	38	71	4
35	28	37	27	12
36	5	7	6	6
37	32	14	35	4
38	49	44	55	14
39	24	33	35	8
Median % Error	39	44	43	14

^aAll first-line supervisors rated on Form A.

^bFirst-line supervisors rated by reliability raters on Form A.

^cFirst-line supervisors rated by department heads on
Forms A and B.

upper-level managers' (reliability raters') Form A ratings of 44 first-line supervisors, and the department heads' ratings of 51 first-line supervisors on Forms A and B, respectively. In all Form A ratings given by department heads, the percentage of first-line supervisors who were rated inaccurately on individual items ranged from .13 to .72 (median, .39). In the department heads' Form B ratings, the percentage of ratees who were rated inaccurately on individual items ranged from .00 to .43 (median, .14). The range in the percentage of first-line supervisors who were rated inaccurately by reliability raters on each item of Form A was from .19 to .75 (median, .44). Again, the data confirm that department heads and reliability raters were about equal in the degree of accuracy with which they rated first-line supervisors on Form A, and that department heads rated more accurately on Form B than on Form A. However, Table 4 also provides an indication of which items on the rating forms were ambiguous or poorly written and caused raters to give inaccurate responses. By examining the logical error percentage on individual items across groups of raters and across rating forms, faulty items can be readily identified. For example, items 1, 5, 12, 20, and 33 were consistently rated with a high degree of inaccuracy on both rating forms and probably need to be rewritten to make them less confusing.

A careful inspection of the different types of logically inconsistent response patterns contained in the Form A

and Form B ratings revealed that 86% of all rating errors were attributable to two rating symbol patterns, -00 and 00+. The percentage of rating errors accounted for by pattern -00 was 23% and the percentage accounted for by pattern 00+ was 63%. Table 5 illustrates the logical inconsistency of these two rating symbol patterns more fully.

"Reliability" of Form A Ratings

A Pearson product-moment correlation was calculated on the overall rating scores received by the 44 first-line supervisors who were rated by department heads and upper-level managers on Form A. The overall rating scores correlated .58, which is a significant correlation ($P < .001$) but which does not denote a high degree of agreement between the department heads and upper-level managers. As mentioned earlier, the high percentage of logical errors in the job performance ratings rendered by both groups of raters (i.e., 39% for both groups) makes this measure of inter-rater agreement, at best, very difficult to interpret and, at worst, meaningless.

Comparison of Form A and Form B Rating Scores

Table 6 presents the means and standard deviations of the ratings obtained by 51 first-line supervisors who were rated by their department heads on each of the 39 items on Rating Forms A and B. The average rating score on all items combined was 4.8 for the Form A ratings and 5.0

TABLE 5

Illustration of Two Rating-symbol Patterns Which Accounted
for Eighty-six Percent of All Rating Errors

Rating Symbol Pattern		Job Behavior Statements (Example Triad)	Statement Number on Form A
a	b		
<u>0</u>	<u>+</u>	Unsafe acts committed by employees are too often allowed to go uncorrected.	32
<u>0</u>	<u>0</u>	Speaks to and cautions employees in many cases when he sees them perform- ing unsafe acts, but he could be more consistent.	83
-	<u>0</u>	Makes it a point, most of the time, to speak to employees and caution them if they are seen performing unsafe acts.	60

Note.--Raters used the rating symbol zero (0) to indicate that a ratee's performance was aptly described by a job behavior statement, plus (+) to indicate that a ratee's performance was better than described by a job behavior statement, and minus (-) to indicate a ratee's performance was worse than described by a job behavior statement.

^aPattern -00 accounted for 23% of all rating errors.

^bPattern 00+ accounted for 63% of all rating errors.

TABLE 6

Means and Standard Deviations of Ratings on
Individual Items for Fifty-one First-line
Supervisors Who Were Rated with Rating
Forms A and B

Item Number	Form A		Form B	
	Mean	SD	Mean	SD
1	3.9	.8	4.0	1.1
2	4.6	1.2	5.1	1.2
3	4.7	1.1	5.2	1.0
4	4.8	1.3	4.7	1.3
5	5.2	1.2	5.1	1.1
6	5.1	1.3	5.0	1.4
7	4.9	1.2	4.7	1.1
8	4.8	1.3	5.0	1.2
9	4.3	1.0	4.6	1.1
10	4.9	1.2	5.1	1.2
11	5.4	1.1	5.4	1.1
12	4.6	1.2	5.2	1.2
13	5.4	1.1	5.5	1.0
14	5.9	.9	5.4	1.0
15	5.5	1.1	5.5	.9
16	6.0	.8	5.7	.7
17	4.6	1.3	4.7	1.3
18	4.3	1.0	4.6	1.2
19	5.4	1.0	5.6	.9
20	4.3	1.1	4.4	1.1
21	4.3	1.2	4.7	1.1
22	5.5	1.2	5.5	.9
23	4.1	1.1	4.6	1.0
24	4.7	1.2	5.3	1.0
25	3.8	1.0	3.9	1.0

TABLE 6 (Continued)

Item Number	Form A		Form B	
	Mean	SD	Mean	SD
26	4.2	1.1	4.5	1.2
27	3.5	1.2	3.9	1.4
28	6.2	.7	6.2	.5
29	3.6	1.1	3.8	1.0
30	5.3	1.1	5.6	.8
31	4.5	1.0	5.1	1.0
32	4.7	1.2	4.7	1.4
33	3.9	1.3	4.2	1.4
34	5.6	1.0	5.6	.9
35	5.2	1.3	5.0	1.2
36	5.8	.7	5.7	.7
37	5.0	1.1	5.4	1.0
38	4.4	.5	4.5	1.0
39	4.8	1.2	4.8	1.1
Mean Rating per Item	4.8		5.0	
SD	1.2		1.2	

Note.--The data reported are based on ratings in which 41% of the raters' responses were logically inconsistent on Form A, and 14% of the raters' responses were logically inconsistent on Form B.

for the Form B ratings. The two means are not significantly different from each other, but each is significantly different ($P < .002$, $t = 3.36$, $df, 100$) from a mean rating of 4.0 which Blanz and Ghiselli (1972) designate as the number which would show that ratings had been approximately evenly distributed about the midpoint of the seven-point scale and that the ratings were free of leniency error.

Once again, interpretation of the results reported in Table 6 is difficult because of the extent to which the scores obtained were artifacts of Blanz and Ghiselli's (1972) method of scoring logical errors is uncertain. According to Blanz and Ghiselli's (1972) scoring system, double-zero rating patterns of types -00 and 00+ (which accounted for 86% of the logical errors in all ratings) are given a score of four points, whereas internally consistent response patterns containing the rating symbol zero (0) are given scores of 2, 4, or 6 points. Thus, the rating inaccuracies contained in the Form A and Form B ratings could have increased, decreased, or not affected the scores which ratees would have received if the raters had performed accurately. Since the Form B ratings were done more accurately than Form A ratings, it is less likely that the rating scores would change significantly if the department heads had rated with perfect accuracy. Consequently, it could be said that the Form B ratings obtained in this

study are marked by leniency error according to Blanz and Ghiselli's (1972) criterion (i.e., the mean rating exceeded 4.0).

The distributions of the ratings given to the 51 first-line supervisors who were rated on Rating Forms A and B are shown in Figure 1. The rating data are presented in two histograms which indicate the relative frequency with which rating scores ranging from one to seven were obtained by ratees on individual items in the Form A and Form B ratings, respectively. It is evident that both rating distributions are negatively skewed, with the degree of skewness being slightly larger for the Form B ratings. In general, first-line supervisors rarely obtained a scale score of 1, 2, 3, 5, or 7 on the 39 rating form items, whereas a score of 4 or 6 was obtained on practically all items.

The summary statistics for the distributions of overall rating scores obtained with the two rating forms appear in Table 7. On Rating Form A, the mean overall rating score given by department heads was 189.45, and the mean score given by reliability raters was 180.34. The difference in the two mean scores is statistically significant at the .05 level. On Rating Form B, the mean overall rating score given by department heads was 193.64 and this was not significantly different from the mean score given by the same raters on Form A (i.e., 189.45). As was the

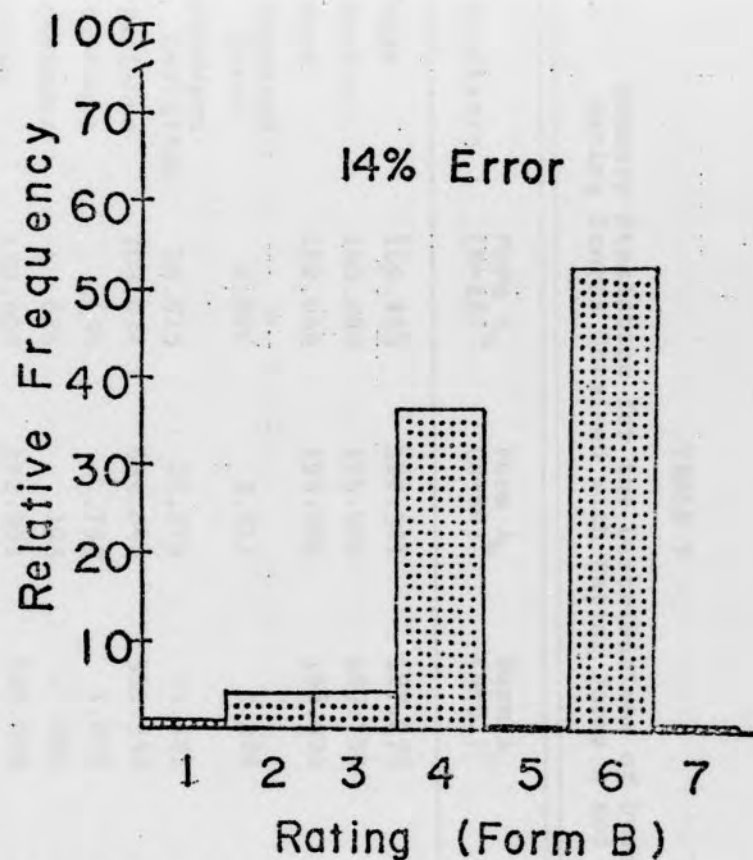
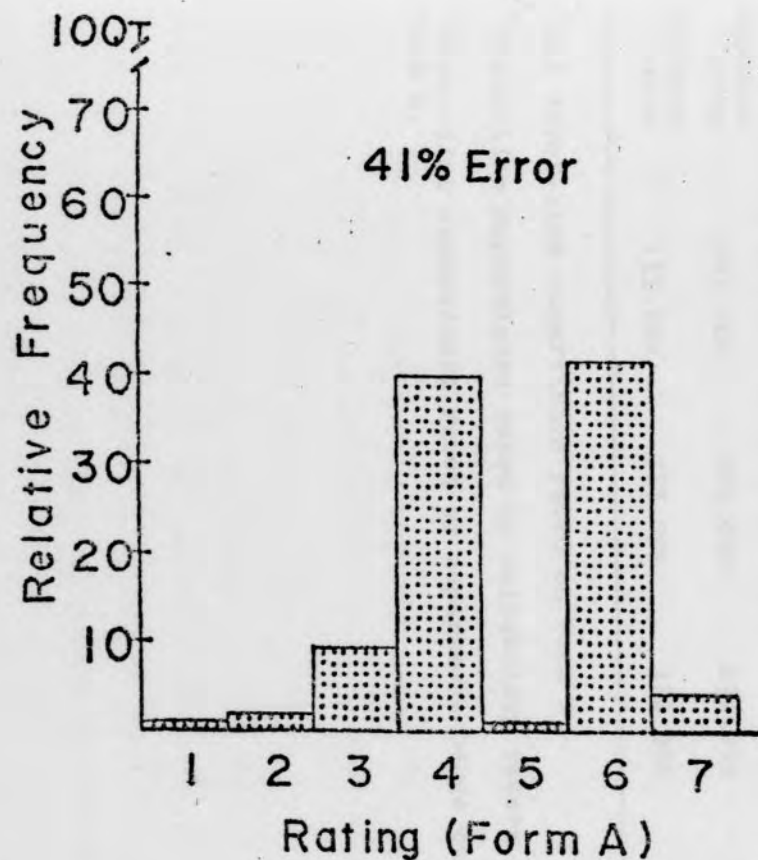


Fig. 1. Distributions of the Form A and Form B ratings of 51 first-line supervisors.

TABLE 7

Summary Statistics for the Distributions of Overall
Rating Scores Obtained with Rating Forms A and B

Statistic	Form A _a (N=88) ^a	Form A _b (N=44) ^b	Form A _c (N=51) ^c	Form B _c (N=51) ^c
Mean	189.455	180.341	187.275	193.647
Median	190.000	175.000	189.750	199.750
Mode	182.000	198.000	196.000	212.000
Standard Error	2.826	3.811	3.288	3.179
Standard Deviation	26.515	25.279	23.481	22.700
Variance	703.034	639.046	551.340	515.300
Kurtosis	.635	-.793	1.050	.938
Skewness	-.626	.193	-.781	-.913
Range	130.000	102.000	115.000	112.000
Maximum Score	241.000	229.000	231.000	230.000
Minimum Score	111.000	127.000	116.000	118.000

^aAll first-line supervisors rated on Form A.

^bFirst-line supervisors rated by reliability raters on Form A.

^cFirst-line supervisors rated by department heads on Form A and Form B.

case in the distributions of individual item ratings which were presented earlier, the separate distributions of overall rating scores given by department heads on Rating Forms A and B, respectively, were both negatively skewed. Again, the distribution of Form B rating scores was more negatively skewed than the distribution of Form A rating scores (i.e., $-.913$ vs. $-.781$).

Regression and Factor Analysis of Form B Ratings

On the assumption that 14% rating inaccuracy could be considered a tolerable amount of error in the Form B ratings, the data were subjected to a regression analysis and then a factor analysis to secure an estimate of the amount of halo error which was contained in the ratings.

The regression analysis was performed for two reasons:

1) to determine whether any items on the rating form could be dropped if a shortened version were developed for future use, and 2) to obtain the tables of inter-item correlations which were provided by the computer in performing a regression analysis so that an accurate estimate of halo error could be derived. The factor analysis was performed because there is an inverse relationship between the number of factors which can be obtained from a set of ratings and the degree of halo in the ratings (cf. Blanz & Ghiselli, 1972). Thus, if the number of factors obtained from the ratings was very small (e.g., one or two), there would be reason to suspect that the degree of halo in the ratings was fairly high.

The results of the regression analysis revealed that there was a fair amount of redundancy in the 39 rating form items. Ninety-nine percent (99%) of the total rating variance was attributable to 24 items, and 95% was accounted for by ten items alone. Correlations between individual item scores and the total rating scores ranged from .06 to .71, with a median correlation of .60.

The outcome of the regression analysis suggests that the rating form could be shortened in length by at least 15 items without any decrement in its validity as an instrument measuring performance in all important aspects of the first-line supervisor's job. The complete regression analysis of the Form B ratings appears in Table 8.

An inspection of the inter-item correlations in the Form B ratings revealed that they contained very little halo error. Seventy-four percent (74%) of the inter-correlations were below .40. Only 8% of the intercorrelations were above .50 and this small percentage can be easily attributed to the fact that there was redundancy in the focus and content of some of the items on the rating form as indicated by the regression analysis. The percentage of inter-item correlations falling between .30 and .50 was 38%. Although some of these correlations are statistically significant, they are certainly much lower than would be expected if raters had given first-line supervisors the same rating on all or most of the items on the rating form.

TABLE 8

Step-wise Regression Analysis of the Form B Ratings
of Fifty-one First-line Supervisors

Item Number	Multiple Correlation (R)	R ²	R ² Change	Simple Correlation with Overall Rating Scores
15	.7126	.5078	.5078	.7126
32	.8407	.7067	.1989	.6672
19	.8977	.8059	.0991	.7110
7	.9242	.8543	.0483	.5683
22	.9393	.8823	.0279	.6029
11	.9501	.9027	.0204	.4144
13	.9591	.9200	.0173	.3744
39	.9667	.9346	.0145	.6513
14	.9710	.9429	.0082	.6944
6	.9751	.9508	.0079	.6550
21	.9803	.9610	.0102	.6399
24	.9830	.9664	.0053	.6764
26	.9843	.9689	.0024	.5949
10	.9858	.9719	.0030	.5119
5	.9872	.9746	.0027	.0663
3	.9885	.9772	.0025	.4858
29	.9898	.9798	.0026	.5348
31	.9912	.9825	.0026	.4143
37	.9922	.9845	.0020	.3277
17	.9935	.9871	.0026	.5838
27	.9937	.9876	.0004	.5937
34	.9942	.9885	.0009	.6497
4	.9945	.9892	.0006	.5030
12 ^a	.9952	.9904	.0012	.2748
18	.9955	.9911	.0007	.6934
35	.9960	.9921	.0009	.6414

TABLE 8 (Continued)

Item Number	Multiple Correlation (R)	R ²	R ² Change	Simple Correlation with Overall Rating Scores
23	.9963	.9927	.0006	.6620
9	.9969	.9939	.0011	.6449
30	.9971	.9943	.0003	.5424
2	.9972	.9945	.0002	.3799
25	.9973	.9946	.0001	.4697
36	.9973	.9947	.0000	.5868
16	.9973	.9947	.0000	.3042
20	.9974	.9948	.0001	.4729
8	.9974	.9949	.0000	.4292
33	.9974	.9949	.0000	.4796
38 ^b	-	-	-	.4576
1	-	-	-	.4697
28	-	-	-	.3251

^aArbitrarily selected cut-off point at which 99 percent of the variance of overall rating scores was accounted for. Standard error of estimate = 3.07.

^bComputer discontinued further calculation of R values because R² Change had stabilized at zero.

Finally, the group of items shown by the regression analysis to account for 99% of the total rating variance was factor analyzed. The factor analysis yielded eight distinct (Eigen value ≥ 1.0 , varimax rotation) factors. The percentage of the rating variance accounted for by factors was 76%. No attempt was made to name the factors. Table 9 summarizes the factor analysis of the Form B ratings and shows the item numbers of the job behaviors included in each of the factors, as well as the loading of each item on its respective factor.

TABLE 9

Factor Analysis of Twenty-four Rating Form Items
Shown to Account for Ninety-nine Percent of the
Variance in Overall Rating Scores
on Form B

Factor	Item Number	Factor Loading
1	39	.7118
	34	.6376
	4	.6938
2	22	.7526
3	11	.7526
	12	.5784
4	15	.6244
	14	.7178
	5	.6028
5	32	.5625
	26	.5056
	29	.8602
	22	.6089
6	10	.7813
	17	.6571
7	3	.8802
8	24	.6068
	37	.5083

Note.--Factors accounted for 76% of the variance
in overall rating scores.

DISCUSSION

An unexpectedly high percentage of logical rating errors occurred in the ratings collected with Rating Forms A and B, despite the fact that elaborate precautions were taken at the time that the rating instruments were developed to prevent such errors. It is difficult to understand how double-zero rating patterns of types -00 and 00+ or of any other configuration could have occurred on Rating Form B except through raters' carelessness or through faulty wording of some rating form statements. The results of the error analyses suggested that the latter may have been a factor, but the former alternative seems the more likely since it was previously shown that a sample of 12 individuals were able to perfectly discriminate the meaning of rating form statements when they were arranged in the unmixed format.

The fact that a large number of double-zero rating errors occurred in the Form A ratings is amenable to one reasonable explanation. Raters were specifically instructed to respond to job behavior statements in order of occurrence on Form A and not to change their responses once they had made a final decision. Due to the semi-random order of arrangement of statements on Form A, it is conceivable that a conscientious rater could place a zero in front of two statements belonging to the same triad if he did not see the

statement which most aptly described a ratee until after he had placed a zero in front of a related statement which had occurred earlier on the form. If a rater followed directions and did not go back and change his responses, even if he realized the logical incongruity in some of them, the double-zero response pattern remained on the completed rating form when it was returned to the experimenter for scoring.

Comments that three or four department heads made to the experimenter after returning completed rating forms (i.e., Form A) to the Industrial Relations Department for scoring suggest that logical errors may have occurred in the manner hypothesized. One rater remarked, "It seemed like I was using an awful lot of zeros when I was filling out those forms and it didn't seem right to me, but the directions said not to change my answers so I didn't." The extent to which rating inaccuracies in the Form A ratings were attributable to a combination of item arrangement and the rating instructions could be tested by rearranging the individual statements on the rating form or by changing the instructions to allow raters to change responses at any time.

A variable which very likely had an effect on the accuracy of ratings obtained with both rating forms is the sheer length of the forms. A number of department heads commented that the forms were simply too long and that completing them was a tedious task. To complete a single

rating form correctly, it was necessary for a rater to spend from fifteen to twenty minutes of his time in sustained attention and careful thought. Many raters chose to do the ratings at home during off-duty hours rather than to interrupt their regular job duties to fill out rating forms. Considering that department heads were asked to rate anywhere from one to nine first-line supervisors on Form A (average number was four), and from two to three supervisors on Form B, it is not unlikely that rater fatigue caused some rating errors to occur in both sets of ratings. The results of the regression analysis performed on the Form B ratings suggest, however, that the number of items on the rating form can be reduced considerably. Thus, rating errors which may have been associated with the length of the rating forms in the present study could be easily eliminated if the same instruments were used for future research.

The facts that the raters used in this study did not have a special interest in test validation or rating research, were not specially compensated for participating in the project, and were possibly irritated by what may well have been considered "extra paperwork from the front office" could have very easily reduced their motivation to complete rating forms carefully and accurately. To maintain that these factors contributed to the number of rating errors obtained in the Form A and Form B ratings

would be pure speculation, however, and even if they did affect the ratings it would be difficult to determine how much effect they had.

There is a remote possibility that some degree of error in the Form A ratings was attributable to imprecise or cumbersome wording of specific rating form statements. An attempt was made when the rating forms were developed to control such errors, but there is a chance that performance level statements which were clearly distinguishable when viewed together in their respective triads were not clearly distinguishable when viewed separately. If, for instance, two related statements were perceived to be saying the same thing when raters viewed them separately in the mixed standard scale, one would expect that a double-zero rating error would occur. Some department heads expressed the belief that Form A contained items designed to "catch" a rater by asking the same question more than once. This type of comment led the experimenter to entertain the possibility that some statements belonging to the same triad looked closely enough alike when viewed apart from each other on Form A, to be taken to mean the same thing. On the other hand, other department heads made the same comment and followed it by saying that they found their suspicion to be incorrect when they checked the separate statements.

While the random or semi-random arrangement of job performance descriptions on the mixed standard scale (Form A)

probably helped raters to respond to individual statements with a greater independence of judgment, it is questionable whether it disguised the order of merit described by separate statements. As mentioned earlier, a strong effort was made when the rating instruments were developed to prevent rating inaccuracies from occurring which were due to ambiguous statements. Consequently, job behavior statements were written so that their meaning was as clear as possible. In order to make statements which belonged to the same triad of statements easily distinguishable from one another, it was sometimes necessary to use cue words which left no doubt as to the order of merit of a statement (e.g., almost always, often, never, acceptable but could be better, etc.). Statements not containing such cue words were still very obvious with respect to the level of job proficiency which each described.

To the extent that the order of merit of rating form statements was straightforwardly evident in most instances, it is questionable whether the mixed standard scale format could have reduced raters' ability to rate leniently if they so desired. As early as 1955, Ghiselli and Brown pointed out that:

The fact that a rater may not know the actual scale value of the items may lead to the false presumption that rater biases cannot operate. Even the most obtuse rater would know he is giving a low rating if he checks "Does as little as possible"[p. 113].

This statement is certainly applicable to the rating scales

used in this study and is at least partially applicable to the scales used in Blanz and Ghiselli's (1972) study.

From the feedback given to the experimenter by the raters after completing rating Form A, there is little doubt that most raters realized that the statements appearing on the rating form belonged to separate groups of three. As mentioned earlier, some department heads became suspicious when they came across similar sounding statements on the rating form and checked to be sure that the statements were not the same and that they were not being tricked in some way. It would be legitimate to ask why there wasn't a large number of perfect (i.e., free of errors) ratings on Form A if the raters were aware that statements belonged to separate triads. The most probable answer to this question is that it would have taken a rater an inordinate amount of time to locate and cross-check all related statements on Form A to give a single perfect rating.

The results of the present study provide valuable information about the mixed standard scale which could not have been predicted from Blanz and Ghiselli's (1972). The main finding was that when rating scale content is held constant, the mixed standard scale format caused raters to make about twice as many logically inconsistent responses to rating form statements as they made when the rating scales were arranged in an unmixed format. While one might have predicted a priori that there would be a difference

in the number of logical errors occurring in ratings obtained with mixed and unmixed standard scales, and may have predicted that the number of logical errors would be greater in mixed standard scale ratings, there was no reason to suspect that the difference in the number of logical errors which would occur in mixed and unmixed standard scale ratings would be as dramatic as it was found to be in this study.

A second important finding of the present study was that raters are apt to be confused by or suspicious of a mixed standard scale. This confusion and/or suspicion may well affect raters' motivation or ability to do ratings correctly, despite rater training and despite the fact that the rating scale content is familiar and reasonably clear to them.

The practical implications of the results of this study would appear to be that while, in some instances, a mixed standard may provide merit ratings which are relatively free of halo and leniency errors, it is also possible for a mixed standard scale to yield rating results which are so contaminated by rating inaccuracies that no valid inferences about ratees' job performance can be drawn from the rating data. In view of the great deal of time required to develop a mixed standard scale, to train raters to apply the rating system, to administer the scale, and to score the results, it would be wise to be very cautious about using a mixed standard scale to obtain job performance ratings in industrial settings.

A final comment which should be entered here is that Blanz and Ghiselli's (1972) rating system appears to have a great deal of merit apart from the mixed standard scale per se. If the system is used in conjunction with a rating scale arranged in an unmixed format, it provides a means of determining whether ratings have been done carelessly, and also forces raters to attend more closely to the content of individual rating scale statements than is usually the case if raters are required simply to check a statement which most closely describes the ratee.

REFERENCES

- Barrett, R. S., Taylor, E. K., Parker, J. W., & Martens, L. Rating scale content: 1. Scale information and supervisory ratings. Personnel Psychology, 1958, 11, 333-346.
- Bass, B. M. Reducing leniency in merit ratings. Personnel Psychology, 1956, 9, 359-369.
- Berkshire, J. R., & Highland, R. W. Forced-choice performance rating--a methodological study. Personnel Psychology, 1953, 6, 355-378.
- Bingham, W. V. Halo, invalid and valid. Journal of Applied Psychology, 1939, 23, 221-228.
- Blanz, F., & Ghiselli, E. E. The mixed standard scale: A new rating system. Personnel Psychology, 1972, 25, 185-199.
- Campbell, J. P., Dunette, M. D., Lawler, E. E., & Weick, K. E., Jr. Managerial behavior, performance, and effectiveness. New York: McGraw Hill, 1970.
- Ewart, A., Seashore, S., & Tiffin, J. A. A factor analysis of an industrial merit rating scale. Journal of Applied Psychology, 1941, 25, 481-486.
- Freyed, M. The graphic rating scale. Journal of Educational Psychology, 1923, 14, 83-102.
- Folgi, L., Hulin, C. L., & Blood, M. R. Development of a first-level behavioral job criterion. Psychological Bulletin, 1971, 55, 3-8.
- Ghiselli, E. E., & Brown, C. W. Personnel and industrial psychology. New York: McGraw Hill, 1955.
- Ghiselli, E. E., & Ghiselli, W. B. Ratings--"kungabe" or "beschreibung." Journal of Psychology, 1972, 80, 263-271.
- Glickman, A. S. Effects of negatively skewed ratings on motivations of the rated. Personnel Psychology, 1955, 8, 39-47.

- Gordon, M. E. The effect of the correctness of the behavior observed on the accuracy of ratings. Organizational Behavior and Human Performance, 1970, 5, 366-377.
- Grant, D. L. A factor analysis of managers' ratings. Journal of Applied Psychology, 1955, 39, 283-286.
- Guilford, J. P. Psychometric methods. New York: McGraw Hill, 1954.
- Guion, R. M. Personnel testing. New York: McGraw Hill, 1965.
- Habbe, S. The appraisal of job performance. The National Industrial Conference Board Studies in Personnel Policy, 1951, Report No. 121, 3-52.
- Kelley, T. L. Principles underlying the classification of men. Journal of Applied Psychology, 1919, 3, 50-67.
- Landy, F. J., & Guion, R. M. Development of scales for the measurement of work motivation. Organizational Behavior and Human Performance, 1970, 5, 93-103.
- Likert, R. A. A technique for the measurement of attitudes. Archives of Psychology, 1932, 22, 1-15 (No. 140).
- McGregor, D. An uneasy look at performance appraisal. Harvard Business Review, 1957, 35, 89-92.
- Nunnally, J. C. Psychometric theory. New York: McGraw Hill, 1967.
- Obradovic, J. Modification of the forced-choice method as a criterion of job proficiency. Journal of Applied Psychology, 1970, 54, 228-233.
- Paterson, D. G. Methods of rating human qualities. Annals of the American Academy of Political and Social Science, 1923, 110, 81-93.
- Roach, D. E. Factor analysis of rated supervisory behavior. Personnel Psychology, 1956, 9, 487-498.
- 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.

- Spicer, L. G. A survey of merit rating in industry. Personnel Journal, 1951, 27, 515-518.
- Stevens, S. N., & Wonderlic, E. F. An effective revision of the graphic rating technique. Personnel Journal, 1938, 13, 125-134.
- Taylor, E. K., & Hastman, R. Relation of format administration to the characteristics of graphic rating scales. Personnel Psychology, 1951, 4, 39-47.
- Taylor, E. K., & Manson, G. E. Supervisory ratings--making graphic scales work. Personnel Journal, 1951, 27, 504-514.
- Taylor, E. K., & Wherry, R. W. A study of leniency in two rating systems. Personnel Psychology, 1951, 4, 39-41.
- Tiffin, J. Merit rating: Its validity and technique. In M. J. Dooker and V. Marques (Eds.), Rating employee and supervisory performance; a manual of rating techniques. New York: American Management Association, 1950.
- Tiffin, J., & McCormick, E. J. Industrial psychology (5th ed.). Englewood Cliffs, N. J.: Prentice Hall, 1965. P. 242.
- Travers, R. M. W. A critical review of the validity and rationale of the forced choice technique. Psychological Bulletin, 1951, 48, 62-70.
- Van de Geer, J. P. Introduction to multivariate analysis for the social sciences. San Francisco: W. H. Freeman, 1971.
- Wyatt, W. W., & Bridges, C. M. Statistics for the behavioral sciences. Boston: D. C. Heath and Company, 1966.

APPENDIX A
SUPERVISORY INTERVIEW FORM

NAME: _____ PLANT: _____
POSITION: _____ DEPARTMENT: _____
NUMBER OF SUPERVISORS UNDER THIS MAN'S DIRECTION: _____
JOB TITLES OF SUBORDINATES SPECIAL FUNCTIONS (IF ANY)

- 1) What does a first-line supervisor do in your department?
(number, list, and provide a short description of each
supervisory duty).

- 2) For each of the job functions mentioned, what are some
observable behaviors which could be used to distinguish
good from poor performance?

No.

Good

Poor

(Use reverse side of this form if necessary)

APPENDIX A (Continued)
SUPERVISORY INTERVIEW FORM
(For Interviewer Only)

I. Job Areas of Interest:

- 1) Production
- 2) Quality Standards
- 3) Employee Relations
- 4) Maintenance of Equipment
- 5) Orientation and Training
- 6) Safety
- 7) Selection and Placement
- 8) Housekeeping
- 9) Operating and Waste Cost
- 10) Communications

II. Prompts:

- 1) How can you tell a good supervisor from a poor one (i.e., what sort of things do they do differently when they are on the job)?
- 2) On what basis do you evaluate supervisors at present? What sort of things do you write down on annual performance review?
- 3) If I asked you to rank order your supervisors from best to worst, could you do it? If you can, do it mentally right now and tell me how the man at the top of your list differs from the one at the bottom.
- 4) What are some of the things that supervisors have done over the years which have irritated you and caused you to think that they could be doing a better job?
- 5) In your experience, you have probably known supervisors who you have considered good supervisors (i.e., they do an acceptable job) and some you have considered really outstanding. What kinds of things did the really outstanding ones do that set them apart from all the rest in your opinion?

Plant _____

Rater _____

Ratee _____

CONE MILLS CORPORATION

FIRST-LINE SUPERVISOR

JOB PERFORMANCE RATING FORM

Form A

INSTRUCTIONS FOR RATING FIRST-LINE SUPERVISORS

This rating form contains 117 different statements which describe ways in which a first-line supervisor may perform those job behaviors which are important to supervisory success. You are to read each statement very carefully and, with an ink or ballpoint pen, respond to the statement EXACTLY AS IT IS WRITTEN by placing ONE of the three following symbols in the blank space in front of the statement:

- + = This supervisor's performance is better than described in this statement.
- 0 = This supervisor's performance is pinpointed fairly well by this statement. His performance is no better or worse than described.
- = This supervisor's performance is worse than described in this statement.

No statement contained in this rating form is exactly the same in make-up or meaning as any other statement on the form. Therefore, each statement requires separate attention and careful consideration before a response is made to it. Some statements may be similar to others on the form because they refer to the same specific job behavior, but no two statements are identical. Individuals differ in the way that they perform any given job behavior and statements have been constructed to reflect or indicate these differences. You should, therefore, consider each statement, as if it were the only statement on the rating form.

Read and respond to each and every statement IN ORDER of occurrence. Items have been written and arranged to maximize the raters ability to judge the supervisor on each statement independently. In other companies, different procedures for completing this type of rating form have been tried, but this one is the simplest and gives the most accurate results.

As you finish responding to each page of items, stop and make sure that you have used the symbol you intended to use in front of each statement. If you find an error, put a insgle slash through the "incorrect" symbol and place the "correct" symbol above the incorrect one (for example, + - 0, etc.) Then, NOTE AT THE BOTTOM OF THE PAGE why the change was necessary (for example "I had the right response in mind but put down the wrong symbol"). When you have finished responding to each page of items, rechecked your responses, and corrected any errors you may have made, PLACE THE PAGE FACE DOWN. DONOT GO BACK TO IT. If you have given each statement the necessary careful consideration, your initial responses will constitute the most accurate rating possible. Again be sure that you understand the rating procedures. Do not hesitate to ask questions. It is important for you to understand the directions and to follow them exactly.

1. — Is entirely too lax and is not conscientious enough in making sure that employees wear recommended and/or required safety equipment. Employees are frequently seen not wearing required safety equipment.
2. — This supervisor usually assigns work by asking employees in a pleasant, personal manner but he is sometimes gruff, impersonal, and/or overly commanding when assigning tasks.
3. — Frequently complains and makes derogatory remarks about the company and company policies in the presence of employees or in public which is likely to present a bad company image and/or weaken morale.
4. — Is irresponsible in the way that he gossips and spreads rumors which may downgrade superiors, peers or workers. The supervisor's conduct in this respect is serious and has a bad effect on employee morale.
5. — Will usually not deny mistakes or try to blame others for his inefficiencies when they are pointed out to him, but tends to accept responsibility for errors in a begrudging, half-hearted sort of manner.
6. — Instructions from superiors are passed on to workers by this supervisor as though they were his own instructions in many cases, but at times (when employees may not like the instructions, for example), he tends to act like a mere "messenger boy" of higher management and does not show support and commitment.
7. — Responds in an enthusiastic, positive ("I'll certainly try") manner when given instructions by superiors.
8. — Follows up on new employees frequently during and after training to make sure that they are trained properly.
9. — When employees have job-related problems, this supervisor, if asked for help, offers some suggestions to the employee as to what he might do to solve a problem, but too often is not available or neglects to stay with an employee and give the individual advice and instructions which may be required to solve the difficulty.
10. — When defective or inferior materials come into this supervisor's department from the preceding department, he rarely takes the necessary action to correct the cause of defects (for example, notifying the supervisor of the other department as well as his own supervisor).
11. — Seems reluctant to make decisions in many cases where he clearly has the authority and responsibility to make the decisions. His boss often ends up making decisions which he himself should make.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

12. ___ Has very little technical knowledge of the machine processes and equipment in his department. For example, he probably couldn't explain accurately how materials coming into the department are changed into the finished product of the department and why each step in the process is necessary.
13. ___ Is one of the best supervisors in the department. It would be difficult to find a man that does as good a job as he does.
14. ___ Gives employees constant and immediate feedback as to how they are doing on the job by telling them frankly, accurately and tactfully what they do well when they do a good job or what they are doing poorly when they do a poor job.
15. ___ Is about like "the average supervisor" in the amount of cooperation he gives supervisors in other departments when problems or emergency situations occur.
16. ___ Conscientiously and consistently follows up on most, if not all, materials returned to him from other departments. Finds the source of defects and has them corrected.
17. ___ Does not see that adequate preventive maintenance is being performed on machinery he is responsible for. Much improvement is needed to keep machinery running at minimally acceptable levels.
18. ___ Is not lax, but could be more conscientious in making sure that employees wear safety equipment. Employees are occasionally seen not wearing proper protective equipment in areas requiring its use.
19. ___ Instructions from superiors are passed on to workers by this supervisor as though they were his own instructions (does not act as a mere "messenger boy" of higher management.)
20. ___ Nearly always reports to work well-groomed, neat and clean in personal appearance (if male: beard and hair are neatly clipped, clothes are clean, face and hands are clean, etc.; if female: Make-up is neat, hair is combed, etc.)
21. ___ Doesn't seem to be on the lookout for safety hazards; accidents or near accidents usually have to occur before safety hazards are corrected.
22. ___ When employees have job-related problems, this supervisor volunteers assistance and makes himself easily available for advice and consultation and to give whatever work instructions are necessary to help the employee solve the difficulty.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

23. ___ Personal appearance is acceptable in most cases when this supervisor reports to work, but more attention should be given to personal neatness, grooming and cleanliness. (If male: his beard, hair, clothes, face and hands, etc., are not always clean and appropriately maintained. If female: hair is not always properly combed, makeup is not always neat, etc.)
24. ___ He undercuts, criticizes and does not comply with parts of personnel policy and procedures which he may not agree with.
25. ___ Encourages employees to suggest ways to improve work, work methods, etc. by asking for suggestions, commending suggestions given, and by acting on those suggestions which are practical and workable.
26. ___ The section he supervises usually falls below the quality and quantity of production expected of his section.
27. ___ When reprimanding an employee is necessary, this supervisor usually does it in private and not in the presence of fellow workers. However, he occasionally slips up and has been observed disciplining employees publicly.
28. ___ Consistently takes the necessary time to inform the supervisor relieving him about the status of the job (that is, lets him know about problems being worked on, difficulties which need attention, etc.) so that operations can progress smoothly through shift change.
29. ___ Routinely and systematically checks machine settings (such as speed settings, temperature and pressure settings, etc.) in his section to make sure machinery is running according to established standards for the materials being processed.
30. ___ Tends to begin work without planning it thoroughly. For example, operations are often held up by such things as not having the right amount of materials scheduled and on hand, and other avoidable difficulties resulting from poor planning and lack of systematic procedures.
31. ___ When this supervisor is given instructions by superiors, he listens attentively and almost always gets the job done without needing to have instructions repeated.
32. ___ Unsafe acts committed by employees are too often allowed to go uncorrected.
33. ___ His boss occasionally ends up making decisions for him even though he clearly had the authority and responsibility to make the decision himself. However, such decisions are not passed up to the boss unnecessarily on a regular basis.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

34. ___ Does not keep his boss adequately informed of problems he should know about; fails to leave notes, sketches, etc., or to communicate orally when it would be appropriate to do so.
35. ___ Does very little, if any, follow-up on new employees during and after training to make sure that they are trained properly.
36. ___ Is not one of the top men in technical knowledge of the machine processes and equipment in the department, but he possesses adequate technical knowledge. For example, he could probably explain with a fair amount of accuracy (though not complete accuracy) how materials coming into his department are changed into the finished product of the department and why each step in the process is necessary.
37. ___ The work area for which this supervisor is responsible is fairly clean most of the time (a few paper cups, old parts, etc. will probably be found in the aisles or near machinery). Occasionally the work area becomes too cluttered and the supervisor has to be reminded to get it cleaned up.
38. ___ This supervisor seems very cautious and close-mouthed when communicating with superiors. He seldom volunteers honest opinions about things and tends to hold back information (particularly unpleasant things) from superiors unless specifically asked.
39. ___ Has not been known to spread gossip or rumors which may downgrade superiors, peers or workers.
40. ___ Seems reluctant to give full cooperation to supervisors in other departments when problems or emergency situations occur. Is worse than "the average supervisor" in this respect.
41. ___ Is very conscientious in making sure that employees wear recommended and/or required safety equipment. Employees are rarely, if ever, seen without necessary protective equipment in areas requiring its use.
42. ___ Is a good supervisor, but not one of the very best in the department. It would not be easy to find a replacement that would do as good a job, but it could be done without an excessive amount of difficulty.
43. ___ If he can blame someone else for his mistakes and inefficiencies he will do it. He never willingly admits to errors and acts very resentful and defensive if someone points his mistakes out to him.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

44. ____ Accepts and usually follows-up on employees' suggestions about ways to improve work, work methods, etc. when the suggestions are practical and workable; but does not go out of his way to encourage suggestions by asking for them or commending employees for suggestions given.
45. ____ Lacks necessary knowledge of the way that the functions of his department are related to those of departments in the rest of the plant.
46. ____ Does not check machine settings (such as speed settings temperature and pressure settings, etc.) in his section often enough to insure that machines are running according to established standards for the materials being processed.
47. ____ Talks positively about the company and company policies in the presence of employees or in public. He does not let personal grievances or specific problems within the company cause him to complain or make derogatory remarks which could hurt morale or present a bad company image.
48. ____ While this supervisor is not the type who sets up long-term goals for himself, short term goals are usually communicated to workers and superiors on a day-to-day basis. Upper management does not have to map things out all of the time for this supervisor.
49. ____ This supervisor shows an unnecessary amount of authority when assigning work to employees and is often gruff, impersonal, and even rude when assigning tasks.
50. ____ When an employee is absent from work, this supervisor makes note of the absence on the employee's attendance record and issues warnings according to absentee policy. In addition, he usually takes time to review the absentee policy with the employee and to let him know where he stands with regard to disciplinary action, but proper counseling and investigation of the absence in order to determine the cause and prevent future absences are not carried out.
51. ____ Is considered among the best in technical knowledge of the machine processes and equipment in his department. For example, he can probably explain exactly how material coming into his department is changed into the finished product of the department and can explain accurately and correctly why each step in the process is necessary.
52. ____ Sees that enough preventive maintenance is performed on machinery he is responsible for to keep them operating at minimally acceptable levels, but more attention to preventive maintenance would help to keep production at a maximum.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

53. ___ Spends little time patrolling and following up on jobs in his work area compared to the time spent in his office or on activities outside his work area.
54. ___ Does not hesitate to make decisions in cases where he has authority and responsibility to do so. His boss rarely, if ever, ends up making decisions which he himself should make.
55. ___ Generally does not gossip or spread rumors which may downgrade superiors, peers or workers, but has been known to say things which could cause an individual in the plant to be looked on in a bad light by others.
56. ___ When this supervisor is given instructions by superiors he usually fails to listen attentively and rarely, if ever, gets the job done without needing to have instructions repeated.
57. ___ Shows an interest in employees as persons by taking time to listen to personal as well as job-related problems. This supervisor attempts to help the employee solve personal problems and refers the employee (if possible) to a qualified person who might be of assistance if specialized help is necessary.
58. ___ Does an acceptable amount of follow-up on defective materials returned to him from other departments. This supervisor generally follows defects back to their source and has problems corrected, but could do a more consistent and thorough job in this area.
59. ___ Makes it a point, most of the time, if not all of the time, to speak to employees and caution them if they are seen performing unsafe acts.
60. ___ Has a thorough knowledge of the way that the functions of his department are related to those of other departments in the plant.
61. ___ Is a mediocre supervisor and it would probably be very easy to find someone who does at least as well as he does, if not better.
62. ___ Occasionally voices complaints and makes derogatory remarks about the company and company policies in the presence of employees or in public which may cause bad morale or present a bad company image.
63. ___ Does a fair amount of follow-up (but not enough) on new employees during and after training to make sure that they are trained properly.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

64. ___ Discusses production problems with employees in an accusing, judgemental manner rather than in a constructive manner. This supervisor is more likely to say, "What are you doing wrong?", rather than, "What is the problem?".
65. ___ Does a fair (but improvable) job of looking for safety hazards, and has them corrected before they cause injuries for the most part; but, occasionally, trouble areas are missed or neglected and the supervisor must be told to have hazards corrected before they cause injuries.
66. ___ Sets a good example by wearing proper protective equipment almost without fail in work areas where safety equipment is needed and/or recommended.
67. ___ Lets employees know how they are doing on the job by telling them immediately and regularly when they are doing poor work and occasionally complimenting them when they are doing a good job.
68. ___ Discourages employees from making suggestions about ways to improve work, work methods, etc. because he fails to commend or to act upon suggestions which are practical and workable.
69. ___ Does a very poor job of observing employees at work and consistently fails to correct poor work habits which result in unnecessary cost and waste (for example, says nothing to a fixer who is working on a loom without covering the cloth).
70. ___ Has an adequate, though not expert, knowledge of the way that the functions of his department are related to departments in the rest of the plant.
71. ___ This supervisor assigns work by asking employees in a pleasant personal manner. He is never gruff, impersonal, or rude when assigning tasks.
72. ___ Checks, but not routinely and systematically, machine settings (such as speed settings, temperature and pressure settings, etc.) in his section to make sure machines are running according to established standards for the materials being processed.
73. ___ Seems unconcerned about personal cleanliness, grooming and neatness and reports to work looking disorderly too much of the time (if male: hair, beard, clothes and face and hands are rarely clean and appropriately maintained. If female: not made-up neatly, hair not combed, etc.)

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

74. — The section he supervises usually surpasses the quality and quantity of production expected of his section.
75. — This supervisor tends to neglect housekeeping (paper, lint, old parts, etc., are frequently found cluttering aisles and machines) and the supervisor must often be reminded to clean up the work area for which he is responsible.
76. — Almost always gives full cooperation to supervisors in other departments when problems or emergency situations occur. Is better than "the average supervisor" in this respect.
77. — Can be relied on to keep his boss fully informed of problems which he should know about by communicating either orally or in writing whenever it is appropriate to do so.
78. — Sometimes fails (but not so often that it has become a major problem) to take the time necessary to inform the supervisor who is relieving him about the status of the job (that is, problems being worked on, etc.), so that operations can progress smoothly through shift change.
79. — When employees have job-related problems, this supervisor helps in any way he can if asked (that is, makes himself available for advice and consultation and to give whatever work instructions are necessary to solve the difficulty) but is not one to volunteer assistance.
80. — This supervisor could not be considered one who sets job goals for himself. Superiors must usually draw up goals and instruct the supervisor to carry them out.
81. — When defective or inferior materials come into this supervisor's department from the preceding department, his handling of the problem is fairly inconsistent. At one time necessary steps are taken to correct the cause of defects (for example, notifying supervision of the preceding department as well as his own supervision) and at other times, little or no action is taken.
82. — Makes sure that prescribed or recommended preventive maintenance is performed on all (or practically all) machinery he is responsible for. He does a better job of insuring preventive maintenance than most supervisors.
83. — Speaks to and cautions employees in many cases when he sees them performing unsafe acts, but he could be more consistent.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

84. — Frequently responds in a negative and resentful manner to instructions given to him by superiors (for example, may question instructions in a nasty or irritating manner--"Why do we have to do that?").
85. — When reprimanding an employee is necessary, this supervisor too often does it in the presence of other workers rather than in private.
86. — Gives full support to personnel policies and procedures even though he may not agree with some parts of them.
87. — Accepts responsibility for his own mistakes or inefficiencies (rather than blaming someone else) and encourages the people he works with to bring errors to his attention.
88. — Is occasionally negligent and could do a better job of setting a good example by wearing protective equipment more consistently in work areas where safety equipment is needed and/or recommended.
89. — Rarely follows up properly on defective materials returned to him from other departments. He acts as if materials are no longer his responsibility when they have left his department.
90. — Spends a large portion of his time patrolling and following up on jobs in his work area compared to time spent in the office or on activities outside his work area.
91. — Conscientiously observes employees at work and consistently corrects poor work habits which would result in unnecessary cost and waste (for example, speaks to loom fixers when he sees them working on a loom without covering the cloth).
92. — The work area for which this supervisor is responsible is clean at all times (paper, lint, old parts, etc. are rarely found cluttering aisles and machines). Does not have to be reminded to clean work area.
93. — Lets employees know how they are doing on the job by telling them immediately and regularly when they are doing a poor job, but usually does not comment on their work when they are doing a good job.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

94. — When an employee is absent from work, this supervisor not only makes note of the absence on the employee's attendance record and issues warnings according to absentee policy, he also takes time to do whatever is necessary to prevent future absences. For example, he reviews the absentee policy with his employee, lets him know where he stands with regard to disciplinary action and reminds the employee of the need for his regular services. He also attempts to determine the exact reason for the absence, and by proper counseling, attempts to help the employee solve problems which may cause future absences.
95. — Is a supervisor who plans work thoroughly and systematically so job emergencies and/or unnecessary hold-ups are avoided. For example, operations are seldom held up due to material shortages resulting from lack of planning or systematic procedures.
96. — Too often fails to take the necessary time to inform the supervisor who is relieving him about the status of the job (that is, problems being worked on, difficulties which need attention, etc.) Unnecessary problems have arisen because of this supervisor's irresponsibility in this area.
97. — The section he supervises usually equals the quality and quantity of production expected of his section.
98. — This supervisor is very willing to listen to employees' job difficulties, but tends to "not want to get involved with personal problems". He seems reluctant to listen to personal problems and does not spend time trying to help an employee solve a personal problem.
99. — Looks for safety hazards and has them corrected before they cause injuries.
100. — Discusses production problems with workers in a constructive rather than accusing, judgemental manner. This supervisor tends to approach an employee by saying, for example, "What is the problem?", rather than, "What are you doing wrong?".
101. — Keeps his boss fairly well-informed of problems he should know about by communicating either orally or in writing at the appropriate times, but sometimes the boss learns of problems from employees or some other source because the supervisor has failed to inform him.
102. — This supervisor is fairly open and honest when communicating with superiors, but sometimes does not express personal opinions about things and/or neglects to give information (pleasant and unpleasant) which superiors should know about.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

103. — When an employee needs to be reprimanded for his actions, this supervisor takes care of the matter in private and not in the presence of other workers.
104. — Demonstrates initiative and interest in the job by routinely setting short and long-term job goals for himself (safety, production, absenteeism, etc.) and communicating goals to workers and superiors.
105. — Is fairly observant of employees at work and does a reasonably good (but not the best possible) job of correcting poor work habits which would result in unnecessary cost and waste (for example, usually speaks to loom fixers when he sees them working on a loom without covering the cloth).
106. — Instructions from superiors are passed on to workers by this supervisor in a manner which makes him appear to be a mere "messenger boy" of higher management (does not present instructions to workers as though they were his own).
107. — Sets a bad example by failing too often to wear protective equipment in work areas where safety equipment is needed and/or recommended.
108. — Spends a moderate portion of his time patrolling and following up on jobs in his work area and a moderate (but more than necessary) amount of time sitting in the office or on activities outside his work area.
109. — This supervisor plans work before starting it, but some degree disorganization shows up in most jobs he does. Hold-ups in operations sometimes occur because of a lack of systematic procedures and sometimes because of faulty planning.
110. — When an employee is absent from work, this supervisor makes note of the absence on the employee's attendance record and issues warnings to absentee policy. If the employee has not given an excuse for his absence, the supervisor may ask for one, but the matter is not followed up to determine the cause and insure proper absentee control.
111. — When this supervisor is given instructions by superiors, he, occasionally fails to listen attentively and is apt to need to have instructions repeated before the job gets done.
112. — This supervisor's reaction to instructions given to him is usually not enthusiastic, but is not negative. He usually responds in an acceptable manner.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

113. — Shows an interest in employees as persons by taking time to listen to personal as well as job-related problems, but is not likely to spend much time trying to help an employee solve a personal problem (for example, not apt to counsel with the employee and work out a possible solution, refer the employee to someone who might be able to help, etc.).
114. — This supervisor communicates with superiors openly and honestly and does not withhold personal opinions or other information (pleasant and unpleasant) which superiors should know about.
115. — Is just as likely to discuss production problems with workers in an accusing and judgemental manner as he is to be constructive. This supervisor is as likely to say to an employee, "What are you doing wrong?" as he is to say, "What is the problem?".
116. — Complies with personnel policies and procedures, but is openly critical (in a non-constructive sort of way) of any parts of the policies which he may not agree with.
117. — When defective or inferior materials come into this supervisor's department from the preceding department, he can be relied on to take steps to correct the cause of the defects (for example, notifying supervision of the preceding department as well as his own supervisor).

RATING CORRECTIONS AND EXPLANATIONS:

Plant _____

Rater _____

Ratee _____

Number of Years Rater Has Supervised This
Employee _____

CONE MILLS CORPORATION

FIRST-LINE SUPERVISOR

JOB PERFORMANCE RATING FORM

Form B

INSTRUCTIONS FOR RATING FIRST-LINE SUPERVISORS

This rating form contains 117 different statements which describe ways in which a first-line supervisor may perform those job behaviors which are important to supervisory success. You are to read each statement very carefully and, with an ink or ballpoint pen, respond to the statement EXACTLY AS IT IS WRITTEN by placing ONE of the three following symbols in the blank space in front of the statement:

- + = This supervisor's performance is better than described in this statement.
- 0 = This supervisor's performance is pinpointed fairly well by this statement. His performance is no better or worse than described.
- = This supervisor's performance is worse than described in this statement.

No statement contained in this rating form is exactly the same in makeup or meaning as any other statement on the form. Therefore, each statement requires separate attention and careful consideration before a response is made to it. Some statements may be similar to others on the form because they refer to the same specific job behavior, but no two statements are identical. Individuals differ in the way that they perform any given job behavior and statements have been constructed to reflect or indicate these differences. You should, therefore, consider each statement, as if it were the only statement on the rating form.

Read and respond to EACH AND EVERY statement IN ORDER of occurrence. Items have been written and arranged to maximize the raters ability to judge the supervisor on each statement independently. In other companies, different procedures for completing this type of rating form have been tried, but this one is the simplest and gives the most accurate results.

As you finish responding to each page of items, stop and make sure that you have used the symbol you intended to use in front of each statement. If you find an error, put a single slash through the "incorrect" symbol and place the "correct" symbol above the incorrect one (for example, + - 0 0; /; /, etc.) Then, NOTE AT THE BOTTOM OF THE PAGE why the change was necessary (for example "I had the right response in mind but put down the wrong symbol"). When you have finished responding to each page of items, recheck your responses, and correct any errors you may have made, PLACE THE PAGE FACE DOWN. DO NOT GO BACK TO IT. If you have given each statement the necessary careful consideration, your initial responses will constitute the most accurate rating possible. Again be sure that you understand the rating procedures. Do not hesitate to ask questions. It is important for you to understand the directions and to follow them exactly.

NOTE: Be sure to respond to each and every statement. Respond with your best guess if you are unsure about the way the ratee performs a specific aspect of the job.

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

1. _____ While this supervisor is not the type who sets up long-term goals for himself, short term goals are usually communicated to workers and superiors on a day-to-day basis. Upper management does not have to map things out all of the time for this supervisor.
 2. _____ This supervisor could not be considered one who sets job goals for himself. His superior must usually draw up goals and instruct the supervisor to carry them out.
 3. _____ Demonstrates initiative and interest in the job by routinely setting short and long-term job goals for himself (safety, production, absenteeism, etc.) and communicating goals to workers and superiors.
-
4. _____ Almost always gives full cooperation to supervisors in other departments when problems or emergency situations occur. Is better than "the average supervisor" in this respect.
 5. _____ Seems reluctant to give full cooperation to supervisors in other departments when problems or emergency situations occur. Is worse than the "the average supervisor" in this respect.
 6. _____ Is about like "the average supervisor" in the amount of cooperation he gives supervisors in other departments when problems or emergency situations occur.
-
7. _____ Responds in an enthusiastic, positive ("I'll certainly try") manner when given instructions by superiors.
 8. _____ This supervisor's reaction to instructions given to him is usually not enthusiastic, but is not negative. He usually responds in an acceptable manner.
 9. _____ Frequently responds in a negative and resentful manner to instructions given to him by superiors (for example, may question instruction in a nasty or irritating manner -- "Why do we have to do that?").
-
10. _____ Keeps his boss fairly well-informed of problems he should know about by communicating either orally or in writing at the appropriate times, but sometimes the boss learns of problems from employees or some other source because the supervisor has failed to inform him.
 11. _____ Can be relied on to keep his boss fully informed of problems which he should know about by communicating either orally or in writing whenever it is appropriate to do so.
 12. _____ Does not keep his boss adequately informed of problems he should know about; fails to leave notes, sketches, etc., or to communicate orally when it would be appropriate to do so.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

13. _____ When employees have job-related problems, this supervisor volunteers assistance and makes himself easily available for advice and consultation and to give whatever work instructions are necessary to help the employee solve the difficulty.
14. _____ When employees have job-related problems, this supervisor helps in any way he can if asked (that is, makes himself available for advice and consultation and to give whatever work instructions are necessary to solve the difficulty) but is not one to volunteer assistance.
15. _____ When employees have job-related problems, this supervisor, if asked for help, offers some suggestions to the employee as to what he might do to solve a problem, but too often is not available or neglects to stay with an employee and give the individual advice and instructions which may be required to solve the difficulty.
-
16. _____ This supervisor seems very cautious and close-mouthed when communicating with superiors. He seldom volunteers honest opinions about things and tends to hold back information (particularly unpleasant things) from superiors unless specifically asked.
17. _____ This supervisor communicates with superiors openly and honestly and does not withhold personal opinions or other information (pleasant and unpleasant) which superiors should know about.
18. _____ This supervisor is fairly open and honest when communicating with superiors, but sometimes does not express personal opinions about things and/or neglects to give information (pleasant and unpleasant) which superiors should know about.
-
19. _____ Sometimes fails (but not so often that it has become a major problem) to take the time necessary to inform the supervisor who is relieving him about the status of the job (that is, problems being worked on, etc.), so that operations can progress smoothly through shift change.
20. _____ Consistently takes the necessary time to inform the supervisor relieving him about the status of the job (that is, lets him know about problems being worked on, difficulties which need attention, etc.) so that operations can progress smoothly through shift change.
21. _____ Too often fails to take the necessary time to inform the supervisor who is relieving him about the status of the job (that is, problems being worked on, difficulties which need attention, etc.). Unnecessary problems have arisen because of this supervisor's irresponsibility in this area.
-

EATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

22. _____ When reprimanding an employee is necessary, this supervisor usually does it in private and not in the presence of fellow workers. However, he occasionally slips up and has been observed disciplining employees publicly.
23. _____ When an employee needs to be reprimanded for his actions, this supervisor takes care of the matter in private and not in the presence of other workers.
24. _____ When reprimanding an employee is necessary, this supervisor too often does it in the presence of other workers rather than in private.
-
25. _____ Does a very poor job of observing employees at work and consistently fails to correct poor work habits which result in unnecessary cost and waste (for example, says nothing to a fixer who is working on a loom without covering the cloth).
26. _____ Is fairly observant of employees at work and does a reasonably good (but not the best possible) job of correcting poor work habits which would result in unnecessary cost and waste (for example, usually speaks to loom fixers when he sees them working on a loom without covering the cloth).
27. _____ Conscientiously observes employees at work and consistently corrects poor work habits which would result in unnecessary cost and waste (for example, speaks to loom fixers when he sees them working on a loom without covering the cloth).
-
28. _____ Is considered among the best in technical knowledge of the machine processes and equipment in his department. For example, he can probably explain exactly how material coming into his department is changed into the finished product of the department and can explain accurately and correctly why each step in the process is necessary.
29. _____ Is not one of the top men in technical knowledge of the machine processes and equipment in the department, but he possesses adequate technical knowledge. For example, he could probably explain with a fair amount of accuracy (though not complete accuracy) how materials coming into his department are changed into the finished product of the department and why each step in the process is necessary.
30. _____ Has very little technical knowledge of the machine processes and equipment in his department. For example, he probably couldn't explain accurately how materials coming into the department are changed into the finished product of the department and why each step in the process is necessary.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

31. _____ Is just as likely to discuss production problems with workers in an accusing and judgemental manner as he is to be constructive. This supervisor is as likely to say to an employee, "What are you doing wrong?" as he is to say, "What is the problem?".
32. _____ Discusses production problems with workers in a constructive rather than accusing, judgemental manner. This supervisor tends to approach an employee by saying, for example, "What is the problem?", rather than, "What are you doing wrong?".
33. _____ Discusses production problems with employees in an accusing, judgemental manner rather than in a constructive manner. This supervisor is more likely to say, "What are you doing wrong?", rather than, "What is the problem?".
-
34. _____ Shows an interest in employees as persons by taking time to listen to personal as well as job-related problems. This supervisor attempts to help the employee solve personal problems and refers the employee (if possible) to a qualified person who might be of assistance if specialized help is necessary.
35. _____ This supervisor is very willing to listen to employees' job difficulties, but tends to "not want to get involved with personal problems". He seems reluctant to listen to personal problems and does not spend time trying to help an employee solve a personal problem.
36. _____ Shows an interest in employees as persons by taking time to listen to personal as well as job-related problems, but is not likely to spend much time trying to help an employee solve a personal problem (for example, not apt to counsel with the employee and work out a possible solution, refer the employee to someone who might be able to help, etc.).
-
37. _____ Generally does not gossip or spread rumors which may downgrade superiors, peers or workers, but has been known to say things which could cause an individual in the plant to be looked on in a bad light by others.
38. _____ Has not been known to spread gossip or rumors which may downgrade superiors, peers or workers.
39. _____ Is irresponsible in the way that he gossips and spreads rumors which may downgrade superiors, peers or workers. The supervisor's conduct in this respect is serious and has a bad effect on employee morale.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

40. _____ When this supervisor is given instructions by superiors he usually fails to listen attentively and rarely, if ever, gets the job done without needing to have instructions repeated.
41. _____ When this supervisor is given instructions by superiors, he listens attentively and almost always gets the job done without needing to have instructions repeated.
42. _____ When this supervisor is given instructions by superiors, he, occasionally fails to listen attentively and is apt to need to have instructions repeated before the job gets done.
-
43. _____ Spends a large portion of his time patrolling and following up on jobs in his work area compared to time spent in the office or on activities outside his work area.
44. _____ Spends little time patrolling and following up on jobs in his work area compared to the time spent in his office or on activities outside his work area.
45. _____ Spends a moderate portion of his time patrolling and following up on jobs in his work area and a moderate (but more than necessary) amount of time sitting in the office or on activities outside his work area.
-
46. _____ Gives full support to personnel policies and procedures even though he may not agree with some parts of them.
47. _____ Compiles with personnel policies and procedures, but is openly critical (in a non-constructive sort of way) of any parts of the policies which he may not agree with.
48. _____ He undercuts, criticizes and does not comply with parts of personnel policy and procedures which he may not agree with.
-
49. _____ Does not hesitate to make decisions in cases where he has authority and responsibility to do so. His boss rarely, if ever, ends up making decisions which he himself should make.
50. _____ His boss occasionally ends up making decisions for him even though he clearly had the authority and responsibility to make the decision himself. However, such decisions are not passed up to the boss unnecessarily on a regular basis.
51. _____ Seems reluctant to make decisions in many cases where he clearly has the authority and responsibility to make the decisions. His boss often ends up making decisions which he himself should make.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

52. _____ Does an acceptable amount of follow-up on defective materials returned to him from other departments. This supervisor generally follows defects back to their source and has problems corrected, but could do a more consistent and thorough job in this area.
53. _____ Conscientiously and consistently follows up on most, if not all, materials returned to him from other departments. Finds the source of defects and has them corrected.
54. _____ Rarely follows up properly on defective materials returned to him from other departments. He acts as if materials are no longer his responsibility when they have left his department.
-
55. _____ When defective or inferior materials come into this supervisor's department from the preceding department, he can be relied on to take steps to correct the cause of the defects (for example, notifying supervision of the preceding department as well as his own supervisor).
56. _____ When defective or inferior materials come into this supervisor's department from the preceding department, he rarely takes the necessary action to correct the cause of defects (for example, notifying the supervisor of the other department as well as his own supervisor).
57. _____ When defective or inferior materials come into this supervisor's department from the preceding department, his handling of the problem is fairly inconsistent. At one time necessary steps are taken to correct the cause of defects (for example, notifying supervision of the preceding department as well as his own supervision) and at other times, little or no action is taken.
-
58. _____ Makes it a point, most of the time, if not all of the time, to speak to employees and caution them if they are seen performing unsafe acts.
59. _____ Speaks to and cautions employees in many cases when he sees them performing unsafe acts, but he could be more consistent.
60. _____ Unsafe acts committed by employees are too often allowed to go uncorrected.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

61. _____ Makes sure that prescribed or recommended preventive maintenance is performed on all (or practically all) machinery he is responsible for. He does a better job of insuring preventive maintenance than most supervisors.
62. _____ Sees that enough preventive maintenance is performed on machinery he is responsible for to keep them operating at minimally acceptable levels, but more attention to preventive maintenance would help to keep production at a maximum.
63. _____ Does not see that adequate preventive maintenance is being performed on machinery he is responsible for. Much improvement is needed to keep machinery running at minimally acceptable levels.
-
64. _____ Is occasionally negligent and could do a better job of setting a good example by wearing protective equipment more consistently in work areas where safety equipment is needed and/or recommended.
65. _____ Sets a bad example by failing too often to wear protective equipment in work areas where safety equipment is needed and/or recommended.
66. _____ Sets a good example by wearing proper protective equipment almost without fail in work areas where safety equipment is needed and/or recommended.
-
67. _____ Looks for safety hazards and has them corrected before they cause injuries.
68. _____ Does a fair (but improvable) job of looking for safety hazards, and has them corrected before they cause injuries for the most part; but, occasionally, trouble areas are missed or neglected and the supervisor must be told to have hazards corrected before they cause injuries.
69. _____ Doesn't seem to be on the lookout for safety hazards; accidents or near accidents usually have to occur before safety hazards are corrected.
-
70. _____ Does very little, if any, follow up on new employees during and after training to make sure that they are trained properly.
71. _____ Follows up on new employees frequently during and after training to make sure that they are trained properly.
72. _____ Does a fair amount of follow-up (but not enough) on new employees during and after training to make sure that they are trained properly.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

73. _____ The work area for which this supervisor is responsible is clean at all times (paper, lint, old parts, etc. are rarely found cluttering aisles and machines). Does not have to be reminded to clean work area.
74. _____ This supervisor tends to neglect housekeeping (paper, lint, old parts, etc., are frequently found cluttering aisles and machines) and the supervisor must often be reminded to clean up the work area for which he is responsible.
75. _____ The work area for which this supervisor is responsible is fairly clean most of the time (a few paper cups, old parts, etc., will probably be found in the aisles or near machinery). Occasionally the work area becomes too cluttered and the supervisor has to be reminded to get it cleaned up.
-
76. _____ Is very conscientious in making sure that employees wear recommended and/or required safety equipment. Employees are rarely, if ever, seen without necessary protective equipment in areas requiring its use.
77. _____ Is entirely too lax and is not conscientious enough in making sure that employees wear recommended and/or required safety equipment. Employees are frequently seen not wearing required safety equipment.
78. _____ Is not lax, but could be more conscientious in making sure that employees wear safety equipment. Employees are occasionally seen not wearing proper protective equipment in areas requiring its use.
-
79. _____ Gives employees constant and immediate feedback as to how they are doing on the job by telling them frankly, accurately and tactfully what they do well when they do a good job or what they are doing poorly when they do a poor job.
80. _____ Lets employees know how they are doing on the job by telling them immediately and regularly when they are doing a poor job, but usually does not comment on their work when they are doing a good job.
81. _____ Lets employees know how they are doing on the job by telling them immediately and regularly when they are doing poor work and occasionally complimenting them when they are doing a good job.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

82. _____ Nearly always reports to work well-groomed, neat and clean in personal appearance (if male: beard and hair are neatly clipped, clothes are clean, face and hands are clean, etc.,; if female: make-up is neat, hair is combed, etc.).
83. _____ Seems unconcerned about personal cleanliness, grooming and neatness and reports to work looking disorderly too much of the time (if male: hair, beard, clothes and face and hands are rarely clean and appropriately maintained. If female: not made-up neatly, hair not combed, etc.).
84. _____ Personal appearance is acceptable in most cases when this supervisor reports to work, but more attention should be given to personal neatness, grooming and cleanliness. (If male: his beard, hair, clothes, face and hands, etc., are not always clean and appropriately maintained. If female: hair is not always properly combed, makeup is not always neat, etc.).
-
85. _____ The section he supervises usually falls below the quality and quantity of production expected of his section.
86. _____ The section he supervises usually equals the quality and quantity of production expected of his section.
87. _____ The section he supervises usually surpasses the quality and quantity of production expected of his section.
-
88. _____ Accepts responsibility for his own mistakes or inefficiencies (rather than blaming someone else) and encourages the people he works with to bring errors to his attention.
89. _____ If he can blame someone else for his mistakes and inefficiencies he will do it. He never willingly admits to errors and acts very resentful and defensive if someone points his mistakes out to him.
90. _____ Will usually not deny mistakes or try to blame others for his inefficiencies when they are pointed out to him, but tends to accept responsibility for errors in a begrudging, half-hearted sort of manner.
-
91. _____ Has a thorough knowledge of the way that the functions of his department are related to those of other departments in the plant.
92. _____ Lacks necessary knowledge of the way that the functions of his department are related to those of departments in the rest of the plant.
93. _____ Has an adequate, though not expert, knowledge of the way that the functions of his department are related to departments in the rest of the plant.
-

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

94. _____ Does not check machine settings (such as speed settings, temperature and pressure settings, etc.) in his section often enough to insure that machines are running according to established standards for the materials being processed.
95. _____ Checks, but not routinely and systematically, machine settings (such as speed settings, temperature and pressure settings, etc.) in his section to make sure machines are running according to established standards for the materials being processed.
96. _____ Routinely and systematically checks machine settings (such as speed settings, temperature and pressure settings, etc.) in his section to make sure machinery is running according to established standards for the materials being processed.
-
97. _____ When an employee is absent from work, this supervisor makes note of the absence on the employee's attendance record and issues warnings according to absentee policy. If the employee has not given an excuse for his absence, the supervisor may ask for one, but the matter is not followed up to determine the cause and insure proper absentee control.
98. _____ When an employee is absent from work, this supervisor makes note of the absence on the employee's attendance record and issues warnings according to absentee policy. In addition, he usually takes time to review the absentee policy with the employee and to let him know where he stands with regard to disciplinary action, but proper counseling and investigation of the absence in order to determine the cause and prevent future absences are not carried out.
99. _____ When an employee is absent from work, this supervisor not only makes note of the absence on the employee's attendance record and issues warnings according to absentee policy, he also takes time to do whatever is necessary to prevent future absences. For example, he reviews the absentee policy with his employee, lets him know where he stands with regard to disciplinary action and reminds the employee of the need for his regular services. He also attempts to determine the exact reason for the absence, and by proper counseling, attempts to help the employee solve problems which may cause future absences.
-
100. _____ Is a supervisor who plans work thoroughly and systematically so job emergencies and/or unnecessary hold-ups are avoided. For example, operations are seldom held up due to material shortages resulting from lack of planning or systematic procedures.
101. _____ This supervisor plans work before starting it, but some degree disorganization shows up in most jobs he does. Hold-ups in operations sometimes occur because of a lack of systematic procedures and sometimes because of faulty planning.
102. _____ Tends to begin work without planning it thoroughly. For example, operations are often held up by such things as not having the right amount of materials scheduled and on hand, and other avoidable difficulties resulting from poor planning and lack of systematic procedures.

RATING CORRECTIONS AND EXPLANATIONS:

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

103. _____ Instructions from superiors are passed on to workers by this supervisor in a manner which makes him appear to be a mere "messenger boy" of higher management (does not present instructions to workers as though they were his own).
104. _____ Instructions from superiors are passed on to workers by this supervisor as though they were his own instructions (does not act as a mere "messenger boy" of higher management).
105. _____ Instructions from superiors are passed on to workers by this supervisor as though they were his own instructions in many cases, but at times (when employees may not like the instructions, for example), he tends to act like a mere "messenger boy" of higher management and does not show support and commitment.
-
106. _____ Occasionally voices complaints and makes derogatory remarks about the company and company policies in the presence of employees or in public which may cause bad morale or present a bad company image.
107. _____ Frequently complains and makes derogatory remarks about the company and company policies in the presence of employees or in public which is likely to present a bad company image and/or weaken morale.
108. _____ Talks positively about the company and company policies in the presence of employees or in public. He does not let personal grievances or specific problems within the company cause him to complain or make derogatory remarks which could hurt morale or present a bad company image.
-
109. _____ This supervisor assigns work by asking employees in a pleasant personal manner. He is never gruff, impersonal or rude when assigning tasks.
110. _____ This supervisor usually assigns work by asking employees in a pleasant, personal manner but he is sometimes gruff, impersonal and/or overly commanding when assigning tasks.
111. _____ This supervisor shows an unnecessary amount of authority when assigning work to employees and is often gruff, impersonal and even rude when assigning tasks.
-
-

DO NOT COMPLETE THIS FORM UNTIL YOU HAVE READ AND UNDERSTAND DIRECTIONS

112. _____ Encourages employees to suggest ways to improve work, work methods, etc. by asking for suggestions, commending suggestions given, and by acting on those suggestions which are practical and workable.
113. _____ Discourages employees from making suggestions about ways to improve work, work methods, etc. because he fails to commend or to act upon suggestions which are practical and workable.
114. _____ Accepts and usually follows-up on employees' suggestions about ways to improve work, work methods, etc. when the suggestions are practical and workable; but does not go out of his way to encourage suggestions by asking for them or commending employees for suggestions given.
-
115. _____ Is one of the best supervisors in the department. It would be difficult to find a man that does as good a job as he does.
116. _____ Is a good supervisor, but not one of the very best in the department. It would not be easy to find a replacement that would do as good a job, but it could be done without an excessive amount of difficulty.
117. _____ Is a mediocre supervisor and it would probably be very easy to find someone who does at least as well as he does, if not better.

RATING CORRECTIONS AND EXPLANATIONS:

FORM A SCORING SHEET

80	12	10	73	59
85	36	21	21	2
100	52	127	20	74
102	24	32	25	28
105	115	23	97	44
75	130	53	14	26
84	98	17	31	22
112	143	5	42	33
7	57			

APPENDIX D

Scoring Sheets for Rating Forms A and B

31	4	172	14
81	19	63	70
95	70	54	60
9	36	71	44
79	111	53	72
25	23	39	19
38	59	25	120
101	106	43	56
134	90	8	21
34	28	75	30
78	116	37	109
20	36	32	25
85	11	1	106
87	33	18	25
93	54	41	19
69	69	92	3
105	50	67	62
61	16	14	47

FORM A SCORING SHEET

80	12	10	73	49
48	36	81	23	2
104	51	117	20	71
40	64	32	26	68
15	115	83	97	44
76	100	59	74	25
84	98	17	43	61
112	113	52	5	42
7	57	82	87	13
34	4	107	45	
101	55	88	70	
77	39	66	60	
9	56	21	46	
79	111	65	72	
22	31	99	29	
38	53	35	110	
102	108	63	50	
114	90	8	94	
96	24	75	30	
78	116	37	109	
28	86	92	95	
85	11	1	106	
27	33	18	6	
103	54	41	19	
69	89	93	3	
105	58	67	62	
91	16	14	47	

FORM B SCORING SHEET

2	25	51	74	97
1	26	50	75	98
3	27	49	73	99
5	30	54	77	102
6	29	52	78	101
4	28	53	76	100
9	33	56	80	103
8	31	57	81	105
7	32	55	79	104
12	35	60	83	107
10	36	59	84	106
11	34	58	82	108
15	39	63	85	111
14	37	62	86	110
13	38	61	87	109
16	40	65	89	113
18	42	64	90	114
17	41	66	88	112
21	44	69	92	117
19	45	68	93	116
20	43	67	91	115
24	48	70	94	
22	47	72	95	
23	46	71	96	

APPENDIX E

TRAINING GUIDE FOR CONE MILLS'

FIRST-LINE SUPERVISOR JOB PERFORMANCE RATING FORM

Industrial Relations Department
May, 1973

Because of the importance of correct employee ratings, this training booklet was constructed to give you practice using the Cone Mills' First-Line Supervisor Job Performance Rating Form. The rating symbols you will use to respond to these practice items are the same as those you will use when you actually rate your supervisors.

Read carefully the following short paragraphs which give examples of how some fictitious supervisors go about doing their jobs. Then read the statements below the paragraph which describe the way that a supervisor may perform a job behavior which is important for supervisory success. Respond to each statement by putting a plus(+) in front of it if the fictitious supervisor's performance is better than described by the statement; a zero (0) in front of the statement if the fictitious supervisor's job performance is represented fairly well by the statement (that is: it is not much better or worse than described); and a minus (-) if the fictitious supervisors job performance is worse than described by the statement. In short, you are to place in front of each statement the rating symbol which best indicates how the fictitious supervisor's performance compares with that described in the job behavior statement.

Since only one symbol can be "correct" from the information given you about each fictitious supervisor, you will be able to see how well you understand the rating system from the number of "right" and "wrong" answers you have when each of the items is reviewed and discussed toward the end of this rater training session.

Start now and read the paragraphs about each fictitious supervisor and then "rate" the supervisor by responding to the job behavior statements below each paragraph with the appropriate rating symbol. For your convenience, the rating symbols and their meanings are summarized at the top of each page.

RATING SYMBOLS

- + = the supervisors performance is better than described in the job behavior statement.
- 0 = the supervisors performance is accurately represented by the job behavior statement (that is, it is not much better or worse than described).
- = the supervisors performance is worse than described in the job behavior statement.

A) George R is a supervisor in the weaving department. Since he became a supervisor two years ago, his section has consistently had the highest production of first quality cloth in the weave room. Employee turnover has really never been a problem for George -- he's only lost three workers in two years; one of the three retired. If you were to ask George's employees why he is such a successful supervisor, you would probably hear a variety of different comments, but you would hear one or two things over and over: "George is out here on the job with us just about all the time. He's always checking with us to make sure the job is running all right. If you're making mistakes or being careless, you will hear about it, but he always tells you when you are doing a good job too. I like working for a guy like that."

- 1) _____ Spends little time patrolling and following up in his work area.
- 2) _____ Meets or surpasses the quality and production standards expected of his department.
- 3) _____ Corrects employees when he sees them doing poor work but rarely, if ever, compliments them when they are doing a good job.

RATING SYMBOLS

- + = the supervisors performance is better than described in the job behavior statement.
- 0 = the supervisors performance is accurately represented by the job behavior statement (that is, it is not much better or worse than described).
- = the supervisors performance is worse than described in the job behavior statement.

B) Mack is a dyehouse supervisor. He is very skilled in getting his workers to do exactly what he wants them to do, and he can get the job done right without having to watch over his people every step of the way. One thing that he has successfully taught his workers is cost consciousness. He continually reminds them of how much it costs the company (in needless overhead) every time too much dye is mixed and not all used, every time color formulas are not strictly followed and dyes need to be thrown out, etc. Waste by his employees is minimal. On Mack's recent end-of-the-year performance appraisal his boss mentioned only one area which needs improvement -- being more conscientious in seeing to it that his workers wear required safety equipment. Two of his employees have received serious eye injuries in the past four months because they failed to use protective glasses around dye chemicals.

- 4) _____ Demonstrates he is cost conscious by stressing in his contacts with workers the dollar loss which can result from poor work habits.
- 5) _____ Tends to let employee's careless work habits go uncorrected.
- 6) _____ Makes sure that employees use required protective equipment -- earplugs, safety shoes, goggles, etc.

RATING SYMBOLS

- + = the supervisors performance is better than described in the job behavior statement.
- 0 = the supervisors performance is accurately represented by the job behavior statement (that is, it is not much better or worse than described).
- = the supervisors performance is worse than described in the job behavior statement.

C) John M. is a first-line supervisor who runs a section on the first shift in the Finishing Department. As a rule, John shows up on the job no earlier than 6:50 A.M. in the morning. His workers are experienced and dependable so the shift gets under way on time just about every day. But sometimes workers are absent and machines stand for twenty or thirty minutes while John tries to organize the section and reassign jobs. John's boss has spoken to him on two or three occasions about the dollar loss which accompanies unnecessary delays in production. John M. typically defends himself by saying: "I've warned my people about laying out for every little thing. We just don't have the reliable help we used to have anymore."

- 7) _____ Admits mistakes and does not blame others for his own inadequacies.
- 8) _____ Usually reports to work early enough that he can be ready to reassign manpower if necessary, continue trouble-shooting machinery, etc. as soon as his shift begins so that needless production loss is avoided.
- 9) _____ Accepts and acts upon critical feedback.

RATING SYMBOLS

107

- + = the supervisors performance is better than described in the job behavior statement.
- 0 = the supervisors performance is accurately represented by the job behavior statement (that is, it is not much better or worse than described).
- = the supervisors performance is worse than described in the job behavior statement.

D) Lenny B. is an experienced supervisor who just came to his present job from another company. A clean work area was one thing that was highly stressed at his old company and Lenny has emphasized it in his new position. You almost never find paper cups, candy wrappers, old parts or other litter around machines in Lenny's section. But a clean work area is not the only thing that reflects the good job that he does as a supervisor. His section produces more first quality cloth than any other in the weave room. His workers have never complained to upper management about the way they are treated, and his fellow supervisors speak of him as one of the most cooperative and capable men they have ever worked with. Generally and specifically, Lenny is one of the best supervisors in the plant.

- 10) _____ Rarely meets and never surpasses production and quality standards expected of his department.
- 11) _____ Sees that his work area is as clean as possible at all times.
- 12) _____ Cooperates with supervisors in his own department and in other departments that he must work closely with.
- 13) _____ Tends to neglect housekeeping in his work area (paper, lint, and old parts are frequently found cluttering aisles and machines).

+ = the supervisors performance is better than described in the job behavior statement.

0 = the supervisors performance is accurately represented by the job behavior statement (that is, it is not much better or worse than described).

- = the supervisors performance is worse than described in the job behavior statement.

E) Sam has been a supervisor in the weave room for about six months.

He was promoted to his current position from a Head Loom Fixer.

He gets along well with his employees and he has known most of them for at least five years.

As a supervisor, Sam is still a novice and is not able to answer a lot of his employees' questions about company policies, etc.; but he can be counted on to find out the correct answer and let his employees know right away.

There have been a few occasions on which Sam has had to speak to a new employee about wasting time on the job. As always, the matter was taken care of in private and without public embarrassment for the employee. Sam has spoken many times of a boss he once had who used to "Call a man down in front of everybody" and he has made sure that he doesn't act the same way.

If you had to name one of Sam's weak points it would be that he tends to "pitch in and help" workers a little too often and can often be found working on a machine himself instead of supervising.

14) _____ Rarely follows-up on an employees' questions which he is unable to answer.

15) _____ Sees to it that machine problems are fixed by those responsible for repair and does not use his time improperly by working a machine himself instead of supervising.

16) _____ Reprimands an employee in confidence and not in the presence of fellow workers.

TABLE 1A
Distribution of Age
First-line Supervisors

Age Range	Number of Supervisors
40-44	2
35-39	5
30-34	2
25-29	6
20-24	5
15-19	10
10-14	5
5-9	2
0-4	2
35-39	3
30-34	3
25-29	3
20-24	7
15-19	3
10-14	2

$N = 80$
Mean = 31.75
SD = 12.98

APPENDIX F

Supplementary Tables

TABLE 1A
Distribution of Age
First-line Supervisors

Age Range	Number of Supervisors
62-64	2
59-61	6
56-58	2
53-55	6
50-52	6
47-49	8
44-46	10
41-43	5
38-40	9
35-37	8
32-34	5
29-31	9
26-28	7
23-25	3
20-22	2

N = 88

Mean = 41.49

SD = 10.98

TABLE 2A

Distribution of Experience
First-line Supervisors

Years of Experience	Number of Supervisors
30-31	2
28-29	1
26-27	1
24-25	1
22-23	2
20-21	2
18-19	2
16-17	2
14-15	6
12-13	7
10-11	7
8-9	8
6-7	7
4-5	12
2-3	16
0-1	12

N = 88

Mean = 8.61

SD = 7.49

TABLE 3A

Distribution of Education
First-line Supervisors

Years of Education Completed	Number of Supervisors
16	2
15	0
14	5
13	4
12	32
11	8
10	13
9	8
8	9
7	3
6	3
5	1

N = 88

Mean = 10.72

SD = 2.20

TABLE 4A
Distribution of Raters' Age
Department Heads

Years	Number of Supervisors
60-62	1
57-59	1
54-56	5
51-53	4
48-50	4
45-47	0
42-44	2
39-41	1
36-38	1
33-35	2
30-32	1
27-29	0
24-26	2

N = 24
Mean = 46.25
SD = 38

TABLE 5A
Distribution of Raters' Experience
Department Heads

Years	Number of Supervisors
21	1
19	1
16	2
14	1
12	2
11	0
10	1
9	1
8	1
7	2
6	3
5	2
4	3
3	1
2	1
1	1
0	1

N = 24

Mean = 8.20

SD = 5.68

TABLE 6A

Distribution of Raters' Education
Department Heads

Years of Education Completed	Number of Supervisors
16	5
15	0
14	1
13	0
12	4
11	3
10	5
9	3
8	1
7	1
6	1

N = 24

Mean = 11.38

SD = 2.96

TABLE 7A

Scoring Key for Cone Mills' First-line Supervisor
Job Performance Rating Form

Rating Pattern				Points
L	M	H		
-	-	-	*	1
-	0	-	*	3
-	+	-	*	5
-	0	0	*	4
-	+	0	*	4
-	-	+	*	3
-	0	+	*	4
-	+	+	*	7
-	-	0	*	2
0	-	-	*	2
0	0	-	*	3
0	+	-	*	5
0	-	0	*	2
0	0	0	*	4
0	+	0	*	6
0	-	+	*	2
0	0	+	*	3
0	+	+	*	7
+	-	-		3
+	0	-		4
+	+	-		5
+	-	0	*	5
+	0	0	*	4
+	+	0		6
+	-	+	*	3
+	0	+	*	4
+	+	+		7

Legend: L = Statement for low level of performance
M = Statement for average level of performance.
H = Statement for high level of performance.
* = Logical inconsistency--rating error.

TABLE 8A

List of Corresponding Statements
on Rating Forms A and B

Form A to Form B		Form B to Form A	
Statement No.	Statement No.	Statement No.	Statement No.
1	77	1	48
2	110	2	80
3	107	3	104
4	39	4	76
5	90	5	40
6	105	6	15
7	7	7	7
8	71	8	112
9	15	9	84
10	55	10	101
11	51	11	77
12	30	12	34
13	115	13	22
14	79	14	79
15	6	15	9
16	53	16	38
17	63	17	114
18	78	18	102
19	104	19	78
20	82	20	28
21	69	21	96
22	13	22	27
23	84	23	103
24	48	24	85
25	112	25	69
26	85	26	105
27	22	27	91
28	20	28	51
29	96	29	36
30	102	30	12
31	41	31	115
32	60	32	100
33	50	33	64
34	12	34	57
35	70	35	98
36	29	36	113
37	75	37	55
38	16	38	39
39	38	39	4
40	5	40	56

TABLE 8A (Continued)

Form A to Form B		Form B to Form A	
Statement No.	Statement No.	Statement No.	Statement No.
41	76	41	31
42	116	42	111
43	89	43	90
44	114	44	53
45	92	45	108
46	94	46	86
47	108	47	116
48	1	48	24
49	111	49	54
50	98	50	33
51	28	51	11
52	62	52	58
53	44	53	16
54	49	54	89
55	37	55	10
56	40	56	117
57	34	57	81
58	52	58	59
59	58	59	83
60	91	60	32
61	117	61	82
62	106	62	52
63	72	63	17
64	33	64	88
65	68	65	107
66	66	66	66
67	80	67	99
68	113	68	65
69	25	69	21
70	93	70	35
71	109	71	8
72	95	72	63
73	83	73	92
74	87	74	75
75	74	75	37
76	4	76	41
77	11	77	1
78	19	78	18
79	14	79	14
80	2	80	67
81	57	81	93
82	61	82	20

TABLE 8A (Continued)

Form A to Form B		Form B to Form A	
Statement No.	Statement No.	Statement No.	Statement No.
83	59	83	73
84	9	84	23
85	24	85	26
86	46	86	26
87	88	87	74
88	64	88	87
89	54	89	43
90	43	90	5
91	27	91	60
92	73	92	45
93	81	93	70
94	99	94	46
95	100	95	72
96	21	96	29
97	86	97	110
98	35	98	50
99	67	99	94
100	32	100	95
101	10	101	109
102	18	102	30
103	23	103	106
104	3	104	19
105	26	105	6
106	103	106	62
107	65	107	3
108	45	108	47
109	101	109	71
110	97	110	2
111	42	111	49
112	8	112	25
113	36	113	68
114	17	114	44
115	31	115	13
116	47	116	42
117	56	117	61

MOYER, STEPHEN MICHAEL. A Comparison of Job Performance Ratings Obtained with Mixed and Unmixed Standard Scales. (1974)

Directed by: Dr. William H. McGehee. Pp. 119.

The present study examined the effects of rating scale format on merit ratings secured with Blanz and Ghiselli's (1972) mixed standard rating system. The study had a dual purpose: 1) to discover whether rater training and controls for faulty item construction could eliminate halo and leniency errors from ratings secured with mixed standard rating scales, and 2) to compare the amount of halo and leniency error in merit ratings obtained with rating scales which were arranged in mixed vs. unmixed formats.

Two rating forms were developed and used to measure the job proficiency of first-line production supervisors in a large textile corporation. Both rating forms were comprised of the same set of behaviorally based statements which pertained to all important aspects of a first-line supervisor's job. On one rating form (Form A), rating scales were arranged in Blanz and Ghiselli's (1972) mixed standard format. On the other rating form (Form B), rating scales were arranged in a more conventional format.

Following a special training session in the use of the mixed standard rating system, department heads ($N = 24$) and upper-level managers ($N = 13$) in eight textile mills rated the job performance of subordinate first-line

supervisors on Rating Form A, Rating Form B, or on both rating forms. Department heads rated a total of 88 first-line supervisors on Form A and upper-level managers used the same rating form to give reliability ratings on 44 of the 88 supervisors. One month later, department heads ($N = 19$) rerated the job performance of 51 first-line supervisors on Rating Form B.

Data analysis showed that 39% of the raters' responses on Form A were logically inaccurate and, therefore, in error according to the mixed standard rating system. Fourteen percent (14%) of the department heads' responses on Form B were logically inaccurate. Because of the high percentage of logical errors in the Form A ratings, no comparison of the amount of halo and leniency error in the Form A vs. Form B ratings was possible. Analysis of the Form B ratings indicated that they contained leniency but not halo error. It was hypothesized that the unexpectedly high incidence of logical errors in the Form A and Form B ratings was primarily attributable to item arrangement and raters' carelessness. The practicality of using mixed standard scales in industrial settings was discussed in the light of the results of this study.