Instructional Control of Leniency in Self-Appraisals

Nelson

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Recommended Citation

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
INSTRUCTIONAL CONTROL OF LENIENCY
IN SELF-APPRAISALS

by

Maria Barnum Nelson

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Master of Arts
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
December 1985
INSTRUCTIONAL CONTROL OF LENIENCY
IN SELF-APPRAISALS

Maria Barnum Nelson, M.A.
Western Michigan University

Self- and supervisor ratings were obtained on four classifications of non-teaching public school employees across two instructional conditions. Self-ratings showed more leniency in the condition where employees were instructed not to meet with the supervisor to discuss the ratings; self-ratings showed less leniency in the condition where employees were instructed to meet with the supervisor to discuss the ratings. Despite foreknowledge of the research question and instructional conditions, supervisors showed a tendency to rate employees differentially according to the instructional conditions. The results of the analyses demonstrate the potency of instructional control on rating behavior by employees and supervisors. Questions were raised as to the adequacy of using supervisor ratings as a comparison or criterion measure of leniency.
ACKNOWLEDGMENTS

I would like to express my appreciation and gratitude to the people who offered their assistance towards the completion of this research project.

I extend sincere thanks to Dennis Couch, Barbara Gstohl, and the classified employees and supervisors of the Stanwood School District who offered their time and cooperation despite hectic end-of-the-school-year schedules. Special thanks go to Norman Peterson who helped get this project off the ground and to Alyce Dickinson who helped bring it back down to earth and finished on time. Sincere appreciation goes to Matt Carman who was willing to provide advice on statistics "long-distance." I would also like to acknowledge Dale Brethower for serving as a committee member. And to Doug, my husband, special appreciation for providing the support and encouragement when I most needed it.

Maria Barnum Nelson
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)” If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................. ii

LIST OF TABLES ............................................................ iv

CHAPTER

I. INTRODUCTION ...................................................... 1

II. METHOD ................................................................. 17
   Subjects ............................................................... 17
   Scale Development ................................................ 18
   Procedure ........................................................... 20
   Rating Conditions ................................................ 21

III. RESULTS ............................................................. 22

IV. DISCUSSION ......................................................... 28

APPENDICES

A. Behavioral Observation Scales ................................. 37
B. Instructions For BOS ............................................. 46

BIBLIOGRAPHY ........................................................ 48
LIST OF TABLES

1. T Test for Performance Ratings Between Rating Sources
   Across Conditions .................................................. 23

2. Percentage of Raters in Each Category for Items on
   Behavioral Observation Scales ............................... 24

3. T Test for Performance Ratings Within Rating Sources
   Across Conditions .................................................. 26

4. T Test for Performance Ratings by Supervisors
   Across Conditions .................................................. 27
CHAPTER 1

INTRODUCTION

In 1957, McGregor described a procedure for performance appraisal in which the subordinate takes an active part in assessing past job performance and setting future job goals. Since that time, and consistent with the findings by Vroom (1959) on participative leadership, the benefits of allowing employees a greater level of participation in the process of appraising their job performance have been well-documented. In contrast to traditional methods of appraisal, which proved threatening to the subordinate and in some cases actually decreased the level of subsequent job performance (Kay, Meyer & French, 1965), increasing employee participation can result in less defensiveness shown by the employee during the appraisal (Bassett & Meyer, 1968), can increase the employee's satisfaction with the appraisal (Greller, 1975), improve relations between the employee and the supervisor (French, Kay & Meyer, 1966); increase motivation on the job (Wexley, Singh & Yukl, 1973); and can result in improved job performance (Meyer, Kay & French, 1965).

Despite the many advantages of incorporating self-appraisals into the regular appraisal process, most employers still use the more conventional approach which allows only the supervisor to evaluate the employee. Research has supported the fear by management that employees, when given the opportunity, will inflate their ratings (Prien & Liske, 1962). A study by Parker, Taylor, Barrett & Martens (1959) suggested that employees may inflate their ratings even when
admitting that their supervisor's rating will most likely be lower than their own. If organizations are to allow their employees the opportunity for increased participation in the appraisal process, they must be assured that the employee will give a more accurate reflection of his or her own job performance.

While the degree of discrepancy between an employee's self-appraisal and that of the supervisor's has been explored, research has not gone far in explaining why such discrepancies exist. In a review of the literature, Thornton (1980) concluded that "Existing data do not allow any conclusion whether the quality of self-appraisals is a function of scale format, amount of rater training, type of judgment, or purpose of the appraisal" (p. 268). This may be because the majority of the research has been conducted on issues such as leniency, reliability and validity of self-ratings (Levine, 1980) without explanations of how to overcome biased ratings. Although Bandura (1978) proposed a theory of the process of self-regulated behavior which could explain how self-evaluation may operate, it did not answer the main problem that "persons in applied settings face a strong inducement to distort such self-assessments or self-appraisals in a more favorable direction." (Levine, 1980, p. 261).

It is possible that employees, when told that future promotions or salary adjustments will be based upon the rating of past job performance, will rate their performance high. It is also possible that when self-ratings are conducted under conditions when a supervisor cannot see or dispute the rating, that ratings will be high. Some researchers have hypothesized that the conditions under which self-appraisals are made could influence the rating made by the employee.
Parker et al., (1959) alleged that self-appraisals which are made public should tend to be more modest and realistic than those given in confidence. Meyer (1980) stated that employees who underestimate their performance on a self-appraisal form would be likely to receive more positive statements from the supervisor regarding job performance than would an employee whose appraisal overestimates it. If this is true, the conditions under which self-ratings are made and the possible, or implied, consequences which follow the self-rating could influence whether the rating will be high, low, or roughly equal to the rating by the supervisor.

Skinner (1974) described the effects of reinforcement and punishment of behavior and how these events serve to alter the frequency of behavior. However, such an analysis has not yet been applied to the question of why, or under what circumstances, employees tend to overrate their job performance. In an effort to address this question, the purpose of the present study was to experimentally manipulate conditions under which self-ratings occur to determine if implied consequences can affect the occurrence of inflated (lenient) ratings in self-appraisal.

Before reviewing past research regarding leniency in self-appraisal, it is important to provide an operational definition of it as some researchers have found that different operational definitions result in different values of leniency error (Saal, Downey & Lahey, 1980). In general terminology, lenient ratings typically occur at the high end of the rating scale, while severe ratings are those that concentrate at the low end of the scale. Saal et al. offer three
operational definitions for assessing leniency error. The first, and most popular definition, involves comparing mean dimension ratings with the midpoint of the rating scale. Mean ratings which exceed the midpoint of the scale reflect leniency while those which are below the midpoint reflect severity. A second definition is based on a rater x rater x dimension analysis of variance approach. A significant rater main effect, especially one which can account for a sizable proportion of the total variance, reflects leniency. A last, and less used approach, is to assess leniency or severity by the degree of skewness in the rating distribution. Negative skewness to a significant degree is described as leniency while positive skewness to a significant degree is described as severity.

Sharon and Bartlett (1969) discussed two conditions which should be considered when proposing a definition of leniency. First, leniency should not be called an "error" unless an external criterion exists for comparison. Unless actual performance data can be obtained, no absolute degree of leniency can be assessed. Second, the definition should acknowledge that the rating situation may also affect the degree of leniency. Some raters who consistently rate high or low on the scale may exhibit this tendency in some situations more so than in others. A definition used by Sharon and Bartlett which meets these two conditions is the "significant shift in mean ratings in the favorable direction from one rating condition to another" (p. 252).

The occurrence of leniency error in ratings can pose problems for a supervisor and an employee. According to Latham and Wexley (1981) when a supervisor rates an employee too high (positive leniency) the employee may develop expectations for higher wages, a promotion, or
increased job responsibilities which are not warranted and will not be forthcoming. In the case when an employee receives a rating which is too low (negative leniency), the employee fails to be rewarded for a job well-done. In the first case, the employee fails to receive further rewards which seem justified based on the appraisal, and in the second, the employee fails to receive praise for good job performance. In both, the likely result will be decreased job performance.

In addition, leniency error poses a measurement problem for the researcher. Holzbach (1978) defined leniency error as occurring "when ratings from different rating sources on the same ratee group are significantly different" (p. 579). According to Holzbach, when this situation occurs, it causes restriction of range on the performance ratings which can then limit the magnitude of possible relationships between performance ratings and other variables.

In respect to leniency error as it occurs across rating sources, the indication is that self-ratings are most lenient (Holzbach, 1978; Klimoski and London, 1974; Parker et al., 1959; Thornton, 1968). The evidence does not point consistently in this direction, however, as self-ratings were found to be less lenient than supervisor ratings by Heneman (1974) and less lenient than supervisor or peer ratings (Bolt, 1984). Heneman (1974) attributed the absence of leniency to the fact that the self-ratings were collected under conditions where they were used for research and not for official evaluation purposes which could affect the salary or career of the subjects. In this way, the subjects faced no inducement to bias their ratings. Bolt (1984) also collected self-appraisal data which were not used for official evaluation.
purposes.

Other studies which involved the collection of self-ratings where the results were used only for research purposes did not replicate the findings of Heneman (1974). Holzbach (1978) collected self-ratings from management and professional employees as part of a larger study on managerial effectiveness. Each employee completed a self-rating and a peer-rating under conditions which specified that the data would be used for research purposes only and would remain confidential. The results indicated that mean self-ratings were higher than those by peers and supervisors. In addition, correlations between the three rating sources showed higher correlations between peer and supervisor ratings than with self-ratings.

Klimoski and London (1974) reported similar results from self-, peer, and supervisor ratings of registered nurses. Again, under "research only" conditions, self-ratings were significantly higher than peer and supervisor ratings. Parker et al., (1959) collected self- and supervisor ratings and also requested that employees give an estimate of how they believed they would be rated by their supervisor. Under "research only" conditions the self-ratings were highest, the estimates were somewhat lower, and the supervisor ratings were lowest. From the estimates, the subjects indicated that, although they gave themselves high ratings, they believed that their supervisors would rate lower, which they did, although the estimates did not correctly predict how low.

According to the results of the above study, subjects who give themselves high ratings do so with the full knowledge that their supervisors would not agree with the ratings. Under "research only"
conditions, employees do not justify their ratings to their supervisor and therefore do not undergo the uncomfortable situation of having to explain ratings for themselves which the supervisor believes are too high based on objective observations of the employee's job performance. Teel (1978) found that when self-ratings are discussed with the supervisor but do not in themselves determine salary adjustments or promotions, self-ratings are not lenient.

Thornton (1968) collected self-and supervisor ratings from management level employees under conditions where the appraisals were used to actually assess job performance. A "promotability index" served as a criterion measure. Again, the overall mean self-rating was higher than the overall mean supervisor rating. In addition, the employees who rated themselves highest were considered the least promotable according to the criterion comparison. Assuming that these employees had received some indication from their supervisors that future promotions were not forthcoming, they were the ones who faced the most inducement to overrate their own performance. In doing so, they could only hope to improve their chances by "making themselves look good."

In summary, the general conditions under which an employee rates his or her job performance, when examined more closely, can help to explain the occurrence of overrating. Despite this, no research has yet been published which compares the effect of differential rating conditions on leniency error in self-appraisals.

In studying the difference between ratings by employees and supervisors, some researchers have emphasized personality variables,
such as self-esteem, to explain the discrepancy. In an effort to determine how self-esteem affects rating behavior, Stotland, Thorley, Thomas, Cohen and Zander (1957) engineered failure and success conditions for subjects who were determined to have high or low self-esteem. They found that high self-esteem subjects rated their performance more favorably after failing a task than did low self-esteem subjects. Rasmussen and Zander (1954) suggested that relevant peer groups can affect an individual's self-esteem which could then influence self-evaluation responses. In this study, teachers who felt that they were not conforming to the standards of a desirable peer group rated themselves low on self-appraisals. Baird (1977) hypothesized that an individual's level of self-esteem could help to explain the differences in ratings by employees and supervisors. The results indicated that the largest difference in scores were between employees with high self-esteem who were rated low in performance by their supervisors.

Unfortunately, studies which emphasize personality variables as causes for inflated ratings do not point the way toward solutions to overcome biased ratings. According to the studies discussed above, in order to change the ratings, the self-esteem of the the subjects must be changed. Skinner (1974) offers an alternative explanation for feelings, or inner states of the individual. He explains that "the feelings are merely collateral products of the conditions responsible for the behavior" (p. 52). For example, if an individual receives frequent reinforcement, feelings of self-esteem will be elicited; frequent punishment can elicit feelings of low self-esteem. Similarly, consequences produced by respected colleagues for professional behavior
can result in various feelings of accomplishment or failure. Consequences, in the form of feedback, can have a powerful effect on behavior. Deutsch and Solomon (1959) reported that regardless of actual success or failure on a task, subjects rated themselves more favorably when given positive feedback than when given negative feedback. Under Skinner's analysis, to change rating behavior, the consequences for that behavior must be changed, rather than the inner state or feelings of the individual.

Other researchers have proposed that dissimilar ratings across rating sources may occur because employees may be rating themselves according to the effort they exhibit on the job rather than performance outcomes (Mitchell & Albright, 1972). The supervisor often cannot observe the employee's effort but can evaluate the outcome of the effort. If an employee bases self-ratings on effort and a supervisor bases ratings on performance, the ratings will probably have low agreement. In a study to test how employees and supervisors agree on ratings of effort and performance, Pritchard and Sanders (1973) reported low correlations between employee and supervisor ratings of effort. Mitchell and Albright reported similar findings and also found low correlations between self-rated effort and self-rated performance. Williams & Seiler (1973) reported low correlations of self- and superior ratings of effort but found higher correlations for self- and superior ratings of performance. Porter and Lawler (1968) concluded that although effort can be considered a part of performance, it is still not the same as job performance. However, the question of whether self-ratings are confounded by employee ratings of effort, rather than
job performance, has not been documented.

A lack of agreement between rating sources may not be unreasonable if employees and supervisors have different expectations for priorities and performance for the ratee's job (Zedeck, Imparato, Krausz & Oleno, 1974). According to Schneier and Beatty (1978) "Specifically, differences in job duties and proximity, causing differing frequencies and or duration of observation of ratee performance could account for divergent ratings" (p. 130). Borman (1974) proposed a similar hypothesis in a study assessing the agreement between two levels of raters using performance dimensions developed by each. The results indicated that rater agreement was higher on the rater's own dimensions than on the other level's scales.

However, if both levels of raters work together and agree on rating dimensions, agreement between the rating sources should improve. Unfortunately, the majority of the scales used in self-appraisal have been graphic scales listing various traits regarded by management to be important to the job in question. Traits, such as "honesty, ambition, patience, and cooperation" tend to be ambiguous and cause misinterpretation on the part of all raters despite their level in the organization. With graphic rating scales, there is no guarantee that different rating sources have the same perspective of what behaviors each trait represents. Cascio and Bernardin (1981) suggested that graphic rating scales should be avoided altogether and Miner (1968) recommended basing scale dimensions on behaviors characteristic to the job which influence performance. In a comparison of descriptive and subjective scales, Stockford and Bissell (1949) found more reliability, less bias and smaller degrees of leniency and severity in descriptive
scales rather than in subjective scales.

Two types of scales which are based on job analysis and include critical job behaviors are the Behaviorally Anchored Rating Scale (BARS) and the Behavioral Expectation Scale (BES) developed by Smith and Kendall (1963). Both scales are developed through the critical incident technique developed by Flanagan (1954) although they differ slightly in format. On the BES the behavioral statements are in the form of expectations; on the BARS the statements are written as neutral behaviors. For both scales, behavioral anchors are generated from the critical incidents, the anchors are allocated to specific dimensions, and seven of the anchors are used to form a Thurstone-type rating scale. Intuitively, one would expect that behaviorally based scales would prove psychometrically superior to graphic rating scales, however, this has not always been the case (see Schwab, Heneman and DeCotiis, 1975 for a review of the literature). In a study comparing three rating instruments which included the graphic (trait) scale and the BARS, DeCotiis (1977) found that all three were comparable in terms of their resistance to leniency error and central tendency.

In order to improve on the BARS and graphic rating scales, but also retain the advantages of a behaviorally based rating scale, Latham and Wexley (1977) developed the Behavioral Observation Scale (BOS). After generating critical incidents through job analysis, the incidents are translated into behavioral descriptions, and the descriptions are allocated to specific dimensions with a Likert-type scale attached to each description. The main differences between BARS/BES and BOS is that all relevant job behaviors are found on the BOS while only
behaviors believed to be "critical" are listed on the BARS/BES. On the BOS a rater simply rates the frequency of behavior observed for each behavioral incident. On the BARS/BES, the rater indicates which of the seven behaviors would be most likely to occur by the job incumbent.

To support their claims that the BOS is satisfactory in terms of both reliability and validity for assessing performance, Latham and Wexley (1981) wrote "In previous studies (Latham & Wexley, 1977; Latham, Wexley & Rand, 1975; Ronan & Latham, 1974) the test-retest and interobserver reliability, as well as the validity of the BOS in indicating employee attendance and productivity, were demonstrated" (p. 63). To further test the psychometric characteristics of the BOS, Bolt (1984) developed a BOS for psychiatric aides. Supervisors and psychiatric aides used the scale to rate job performance across three rating levels; supervisor, self, and peers. The results indicated no support for discriminant validity and showed evidence of substantial rater bias. Although the results supported convergent validity, they do not indicate vast improvements over those using BARS/BES. As DeCotiis (1977) concluded, "rather than continuing to pursue the elusive ultimate rating format, future research could perhaps more fruitfully be focused on explicating the rating process and the organizational and individual variables that define it" (p. 265).

One variable which may prove relevant in understanding the rating process is the purpose of the appraisal (Zedeck & Cascio, 1982). In studies where ratings were made for different purposes, or under different conditions, instructions were used to communicate the purpose of the appraisal to the rater. Berkshire and Highland (1953) used three instructional conditions to assess the degree of leniency in
ratings of Air Force technical instructors. One level of supervisors completed ratings on the technical instructors. The first condition was to "rate favorably" or to give the ratee the highest score possible to serve as a known biasing condition. The second condition involved an operational or "for keeps" condition where the ratings were used for administrative decision making purposes. The third condition was an experimental or "research only" condition where the ratings were used for research purposes. A comparison of the ratings in the three conditions revealed that the highest ratings were in the "rate favorably" condition. Of the two remaining conditions, the highest ratings occurred in the "for keeps" condition and the lowest occurred in the "research only" condition.

Taylor and Wherry (1951) compared the ratings from two of the above instructional conditions to assess differences in leniency in ratings of Army officers. Again, supervisors completed ratings on officers below them in rank. Under conditions of "research only" and "for keeps" mean ratings increased from the "research only" to the "for keeps" conditions using two scale formats (graphic and forced choice).

In these two studies, ratings made which could result in administrative action were the highest. Sharon and Bartlett (1969) produced similar results using four instructional conditions for introductory psychology students rating their graduate assistant instructors. The four conditions included identification and justification components in addition to the two conditions used above. The first was a control condition where the ratings were used for
"research only" and the rater did not attach a signature or any other form of identification to the rating. The second was an evaluation condition where the ratings were "for keeps" but again, raters did not identify themselves. The third was an identification condition where the ratings were for "research only" and raters identified themselves. The fourth was a justification condition which was "for keeps" and in addition to identifying themselves on the rating, raters were requested to meet with the ratee and explain (justify) the rating. The results indicated that the identification component did not affect the ratings. However, means in the evaluation and justification conditions were higher than in the conditions where the ratings were for "research only" using both graphic and forced choice scales. The mean ratings of the justification condition were highest overall.

In summary, with respect to supervisors evaluating their subordinates, and students evaluating their instructors, mean ratings under the "for keeps" conditions have been shown to be most lenient. In addition, in conditions where the rater must justify the rating to the ratee, mean scores climb even higher. The results from these studies clearly indicate that instructional conditions, which communicate the purpose of the appraisal to the subjects can affect how individuals rate others.

Although instructions do not serve the function of a consequence which could control behavior, they imply that certain consequences will follow which can affect behavior. Skinner (1974) made a distinction between contingency-shaped behavior where consequences that immediately follow a behavior affect the likelihood of that behavior occurring again, and rule-governed behavior where rules formulated from a past
reinforcement history affect the likelihood of certain behaviors occurring in the future. Because most individuals have sophisticated verbal repertoires which enable them to continually develop rules from contingencies experienced in the past, examples of pure contingency-shaped behavior are difficult to generate. However, the following examples of contingency-shaped and rule-governed behavior will serve as general illustrations for these concepts. An employee who disagrees with ratings by the supervisor and is presented with a lecture about incidences of bad job performance would most likely refrain from verbalizing disagreements in the future; this is an example of contingency-shaped behavior. If that employee is later instructed to appraise his or her job performance and then must justify that rating to the supervisor, the implied consequence of listening to another lecture could affect how high the rating will be, an example of rule-governed behavior.

Skinner (1974) asserted that rules can exert rapid control over behavior. Also, a person following rules may behave differently from a person exposed to the contingencies which are described by the rules. Skinner further pointed out that different types of rule-governed behavior, such as following instructions, heeding warnings and obeying laws can themselves be influenced by their consequences.

To test these assumptions, Galizio (1979) studied the behavior of instruction-following to determine the effect reinforcement and extinction would have when contingent on instruction-following behavior and if discrimination control could be established. Using several subjects in a long-term operant experiment he found evidence supporting
Skinner's assertions. The behavior evoked or controlled by the instructions were shown to be influenced by the consequences following them, instruction-following was brought under stimulus control of differential reinforcement, and that, like simple discriminative stimuli, instructions functioned as reinforcement. In conclusion, Galizio (1979) wrote "The potency of instructional control can be interpreted not as a limitation of reinforcement control of human behavior but rather as an instance of reinforcement history affecting rule-governed behavior" (p. 69).

Under this analysis, people follow instructions presumably because they have been reinforced for doing so in the past. When new instructions are given, the consequences implied by the instructions along with the individual's history of reinforcement affect the subsequent behavior. Instructional conditions have been shown to affect leniency in ratings of others (Berkshire & Highland, 1953; Sharon & Bartlett, 1969; Taylor & Wherry, 1951). The purpose of the present study was to determine if instructional control could be demonstrated for leniency in self-appraisals.
CHAPTER II

METHOD

Subjects

Subjects for hypothesis testing and development of the Behavioral Observation Scale (BOS) included 27 female and 10 male non-teaching (classified) employees and six male supervisors from a public school district in northwestern Washington State which enrolls 2100 full-time students annually. Subjects from the employee group had job classifications of secretary, teaching assistant, bus driver or custodial/maintenance worker.

Subjects from the employee group were randomly selected according to their job classification to insure that the proportion of employees selected from each classification was representative of each classification's proportion to the total group. For example, 38% of the total population available were bus drivers, therefore, 38% of the subjects selected were from the bus driver classification. The number of subjects participating from each classification was bus drivers, 14, custodians, 6, teacher aides, 11, and secretaries, 6.

The supervisors responsible for the yearly evaluation of each of the subjects from the employee group also participated. One supervisor was responsible for all of the custodians; similarly one supervisor was responsible for all of the bus drivers. The remaining four supervisors were building principals and supervised the secretaries and teacher aides working in each school.
Scale Development

The development of the Behavioral Observation Scale was modeled after the procedure used by Latham and Wexley (1977, 1981) and Latham, Fay & Saari (1979). The critical incident technique (Flanagan, 1954) was used to collect critical incidents from the subjects and the supervisor. Subjects met in groups ranging in size from one to five employees of the same classification. Supervisors were interviewed separately. Each subject generated 5 to 15 incidents describing effective and ineffective job behaviors for the job in question. Briefly, effective performance was defined as behavior which, when it occurred, could stand as a model for all employees in that classification. Ineffective performance was defined as behavior which may occur once or several times and leads to the conclusion that the employee may not be competent in that job.

A total of 437 incidents with a range of 85 to 139 incidents for each classification was collected. Duplicate and ambiguous items were deleted leaving a total of 284 items with a range of 57 to 79 items for each classification. The remaining incidents were then re-written into clear behavioral statements.

The experimenter grouped similar incidents together into behavioral criteria or job dimensions. Dimensions varied across job classifications depending on the job tasks and responsibilities listed. For example, dimensions for the secretary classification included "organizational abilities and technical skills" and "administration and supervision", while those for the bus driver classification included "safety procedures" and "bus operations and regulations". General items
for each job classification which did not fall into specific dimensions were classed together as "Work Habits."

The incidents for each job classification were then placed in random order and reassigned to the behavioral dimensions generated above by a labor representative of classified public school employees who was familiar with the duties and responsibilities of each of the four classifications. The calculation for interjudge agreement between this individual and the experimenter was the number of incidents that both agreed should be placed in a certain dimension divided by the combined number of incidents placed in that dimension. Dimensions with an interjudge agreement of .80 were retained. A total of 100 incidents were eliminated across all classifications.

Content validity was assessed by two methods. First, recently updated job descriptions for each classification were reviewed to insure that each item/statement was represented by at least one incident. This resulted in no changes to the number of incidents remaining. Second, each supervisor reviewed a completed BOS form for the job classification(s) which they were responsible for supervising with the instructions to add, delete or edit the items to insure that each item was job relevant. Twelve items were deleted and fourteen were combined into existing incidents.

After these corrections, a Behavioral Observation Scale was completed for each of the four job classifications with an item range of 35 to 51 items (See Appendix A). Each item represented an effective behavior with anchors ranging from (1) almost never, to (5) almost always. As detailed in the instructions given to each employee and supervisor (See Appendix B), the anchors represented the frequency of
occurrence for which they observed each respective behavioral item. A score of 1 indicated the rater observed that the behavior occurred 0-50 percent of the time, 2 for 50-65 percent of the time, 3 for 65-80 percent of the time, 4 for 80-90 percent of the time, and 5 for 90-100 percent of the time.

Procedure

During the interview process, all subjects received general information regarding the study. This information included the fact that the information gathered during the interview would be used to develop an appraisal form for each specific classification. It was explained that this form would not be part of the school district's formal evaluation program, but would be used to test their reactions to a process which allowed them to evaluate themselves in addition to being evaluated by their supervisors. After the appraisal forms were completed, employees participating in the study would receive a form specific to his or her job classification and would be asked to rate his or her own job performance during the past school year. The supervisors would also receive a form and evaluate the employees' performance. Some employees would receive instructions to meet with their supervisor to discuss the ratings. Ratings by the remainder of the employees would be kept confidential. All employees would then fill out a form evaluating the process and return it with the appraisal form to the author.
Rating Conditions

Half of the subjects were randomly selected by job classification to participate in the control condition. Approximately 10 days prior to the end of the school year, subjects and the supervisors received an appraisal packet including instructions, a copy of the relevant appraisal form and a general evaluation form to rate the procedure. Specifically, subjects received instructions that indicated they were not to schedule a meeting with their supervisor to discuss the self-rating. Therefore, for this condition, the designated employees evaluated themselves on the appropriate form and their supervisors also evaluated the employees on identical forms. The forms were then returned to the experimenter and no official communication occurred between the employee and the supervisor regarding the ratings.

Half of the subjects were randomly selected and assigned to the experimental condition. These employees also received an appraisal packet similar to those in the control condition; only the instructions differed. Subjects were instructed to set up a brief time period to review their completed appraisal with their supervisor, who had also rated the employee on an identical form. The purpose of the meeting, as stated in the instructions, was to generate a discussion on areas of agreement and disagreement found in the two ratings.

Supervisors received a list of employees to evaluate along with the appropriate forms. Names of subjects in the experimental condition were marked with an asterisk indicating that they had received instructions to meet with the supervisor to compare the ratings. Supervisors rated from 3 to 14 employees each.
CHAPTER III

RESULTS

To test for lenient ratings, two-way t tests were computed using the data from the four BOS's. Employee and supervisor ratings were analyzed by classification across both the control and experimental conditions. Leniency error was operationalized as a significant rater mean difference from one rating condition to another.

Table 1 presents the results of the t tests of the comparisons between self- and supervisor-ratings across the control and experimental conditions. The results indicate that mean self-ratings in three of the four classifications were higher than mean supervisor-ratings in the control condition. In this condition, employees completed self-ratings under instructions not to meet with their supervisor to compare self-ratings with the supervisor's rating. Although, in the control condition, only the employees in the teacher aide classification rated themselves significantly higher than their supervisors rated them, the combined means for all the classifications shows that, overall, employees rated themselves leniently when they were not required to justify their ratings to the supervisors.

In the experimental condition, where employees were instructed to meet and discuss the ratings with their supervisor, no lenient self-ratings were found. In an opposite trend from the control condition, employees in three of the four classifications rated themselves lower than they were rated by their supervisors. Custodians gave self-
Table 1

T Test for Performance Ratings
Between Rating Sources Across Conditions

<table>
<thead>
<tr>
<th>Classification</th>
<th>Control</th>
<th></th>
<th>Experimental</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Self Supervisor</td>
<td>Na</td>
<td>t</td>
<td>Self Supervisor</td>
</tr>
<tr>
<td>Bus Drivers</td>
<td>4.86</td>
<td>4.87</td>
<td>200</td>
<td>.09</td>
</tr>
<tr>
<td>Custodians</td>
<td>4.75</td>
<td>4.63</td>
<td>143</td>
<td>1.31</td>
</tr>
<tr>
<td>Secretaries</td>
<td>4.60</td>
<td>4.59</td>
<td>133</td>
<td>.11</td>
</tr>
<tr>
<td>Teacher Aides</td>
<td>4.74</td>
<td>4.48</td>
<td>258</td>
<td>3.61***</td>
</tr>
<tr>
<td>Combined</td>
<td>4.75</td>
<td>4.63</td>
<td>734</td>
<td>2.84**</td>
</tr>
</tbody>
</table>

Note. All scores reflect mean values.

aN = number of items rated.

*p < .05  **p < .01  ***p < .001
ratings significantly lower than supervisor ratings and the combined means for the four classifications in the experimental condition indicate, overall, that mean self-ratings were significantly lower than supervisor-ratings when employees were instructed to justify their ratings to their supervisor.

To further illustrate the different rating distributions exhibited by the employees and the supervisors across the experimental conditions, Table 2 shows the percentage of raters in each rating scale category.

Table 2
Percentage of Ratings in Each Category
for Items on Behavioral Observation Scales

<table>
<thead>
<tr>
<th>Scale Category</th>
<th>Control Self</th>
<th>Control Supervisor</th>
<th>Experimental Self</th>
<th>Experimental Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>.5%</td>
<td>.9%</td>
<td>.8%</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>.5%</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>1.1%</td>
<td>6.3%</td>
<td>3.0%</td>
<td>4.8%</td>
</tr>
<tr>
<td>4</td>
<td>22.2%</td>
<td>21.8%</td>
<td>26.6%</td>
<td>22.4%</td>
</tr>
<tr>
<td>5</td>
<td>76.7%</td>
<td>71.3%</td>
<td>68.8%</td>
<td>72.0%</td>
</tr>
</tbody>
</table>

From the table it can be seen that self-ratings in the control condition contain a higher percentage of ratings in the top category than do supervisor ratings. In the experimental condition, however, self-ratings contain a lower percentage of ratings in the top category than supervisor ratings.
To test if employee self-ratings were under control of the instructional conditions, and not simply determined lenient based on comparison with supervisor ratings, two-way t tests were computed to compare mean self-ratings across the control and experimental conditions without reference to supervisor ratings. Table 3 indicates that, in the control condition, mean self-ratings for all classifications were higher than mean self-ratings in the experimental condition. These results support the above findings that employees rated themselves more favorably under conditions where no justification of the self-rating was necessary.

Table 3 also shows mean supervisor ratings across the control and experimental conditions without reference to the employee self-ratings. The majority of past research on leniency in self-ratings used supervisor ratings as a criterion to determine if self-ratings were lenient (Heneman, 1974; Holzbach, 1978; Klimoski & London, 1974; Parker et al., 1959; Thornton, 1968; Waldman & Thornton, 1979). Unless supervisor ratings are shown to be stable across conditions and between supervisors, an employee's self-rating could be determined lenient simply from comparison to supervisor ratings which tend to vary across conditions. From Table 3 it can be seen that supervisor ratings for each classification were highly variable.

Mean supervisor ratings of employees in the bus driver and custodian classifications indicate ratings made by a single supervisor; one supervisor rated all bus drivers and one supervisor rated all custodians. Mean supervisor ratings for the remaining two classifications indicate averaged ratings of four supervisors for each
Table 3

T Test for Performance Ratings
Within Rating Sources Across Conditions

<table>
<thead>
<tr>
<th>Classification</th>
<th>Self</th>
<th></th>
<th>Supervisor</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cntrl.</td>
<td>Exp.</td>
<td>N&lt;sup&gt;a&lt;/sup&gt;</td>
<td>t</td>
</tr>
<tr>
<td>Bus Drivers</td>
<td>4.86</td>
<td>4.79</td>
<td>245</td>
<td>.9</td>
</tr>
<tr>
<td>Custodians</td>
<td>4.75</td>
<td>4.64</td>
<td>195</td>
<td>1.26</td>
</tr>
<tr>
<td>Secretaries</td>
<td>4.60</td>
<td>4.47</td>
<td>167</td>
<td>1.13</td>
</tr>
<tr>
<td>Teacher Aides</td>
<td>4.74</td>
<td>4.42</td>
<td>319</td>
<td>5.23**</td>
</tr>
<tr>
<td>Combined</td>
<td>4.75</td>
<td>4.59</td>
<td>926</td>
<td>4.0**</td>
</tr>
</tbody>
</table>

Note. All scores reflect mean values. Cntrl. = Control; Exp. = Experimental

<sup>a</sup>N = number of items rated.

*p < .01   **p < .001
classification. To further analyze variability of supervisor ratings, Table 4 shows the results of two-way t tests for individual mean supervisor ratings across both experimental conditions.

Table 4

T Test for Performance Ratings by Supervisors Across Conditions

<table>
<thead>
<tr>
<th>Supervisor</th>
<th>Control</th>
<th>Experimental</th>
<th>N</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.87</td>
<td>4.70</td>
<td>192</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>4.63</td>
<td>4.86</td>
<td>199</td>
<td>3.15*</td>
</tr>
<tr>
<td>3</td>
<td>4.11</td>
<td>4.03</td>
<td>58</td>
<td>.52</td>
</tr>
<tr>
<td>4</td>
<td>4.93</td>
<td>5.00</td>
<td>133</td>
<td>4.52**</td>
</tr>
<tr>
<td>5</td>
<td>4.35</td>
<td>4.74</td>
<td>195</td>
<td>7.99**</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>4.25</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*p < .01  **p < .001

Although the supervisors were aware of the research question and the instructional conditions of the present study, three of the supervisors seem to have responded to the instructional conditions by rating employees in the experimental condition significantly higher than employees in the control condition. In other words, three supervisors rated employees, who were expected to discuss the rating with the supervisor, significantly higher on the scale than employees who did not have the opportunity to see or discuss the rating made by the supervisor.
CHAPTER IV

DISCUSSION

The results illustrate the control that instructions can exert over the occurrence of leniency in self-ratings. Sharon and Bartlett (1969) demonstrated that instructional conditions have differential effects on leniency in ratings of others. The present study brought this analysis within the sphere of self-appraisal.

In the control condition where the instructions stated that no meeting should be scheduled to discuss the self-rating with the supervisor, three classifications of employees rated themselves approximately equal to or higher than their supervisors. Employees in the teacher aide classification, however, showed self-ratings significantly higher than the supervisor ratings. Due to the highly lenient self-ratings found in the teacher aide classification, and the trends shown by the remaining classifications, the overall mean for all employee groups indicate a tendency for employees to rate themselves leniently when the instructions do not include a requirement to justify the self-rating. These findings agree with most studies of self-ratings under "research only" conditions (Holzbach, 1978; Klimoski & London, 1974; Parker et al., 1959; Waldman & Thornton, 1979) in showing self-ratings to be more lenient than other rating sources, including supervisor-ratings.
Heneman (1974) reported opposite findings; self-ratings were found to be less lenient than supervisor ratings under conditions where the results were confidential and were used only for research. In a review of the literature, Thornton (1980) suggested that Heneman’s findings could be explained by an apparent sample bias for the supervisor ratings. Heneman collected self- and supervisor-ratings through a mail survey in which less than 50 percent of the supervisors responded. The supervisors who did respond rated their subordinates higher than the subordinates rated themselves. Given the low response rate by the supervisors, it is feasible that the supervisors who would have rated their subordinates low on the scale did not exert the effort to complete the rating or to return it to the experimenter.

Bolt (1984) also reported self-ratings to be less lenient than ratings by other sources when the ratings were used only for research purposes. Bolt developed a Behavioral Observation Scale (BOS) to use as the appraisal instrument; the BOS has not previously been used for self-appraisal purposes. Downs, Farr and Colbeck (1978) reported that realistic self-appraisals can be expected when the appraisal "is confined to essentially behavioral-type tests which take place within the visual field of the testee..., i.e., the testee is visually able to observe her own performance" (p. 276). Comparison of the BOS with other rating formats is needed before firm conclusions can be made regarding the effects of the BOS on leniency in self-ratings.

In the present study, the instructions given to subjects in the experimental condition where employees were instructed to meet with their supervisor to discuss and compare the different ratings, did serve to restrict lenient ratings in all classifications. More
importantly, the two classifications which showed the highest self-ratings in the control condition showed self-ratings lower than supervisor-ratings in the experimental condition. Key features of the procedures used by previous studies under "for keeps" conditions make comparison of the findings with the present study difficult. Thornton (1968) used a trait scale with executive level managers where the appraisals were used as part of the formal evaluation process of the company. Executives discussed the evaluation with the supervisor and mutually established goals for future performance. Unlike the present study, the results showed lenient self-ratings for some of the subjects. The subjects rated as least promotable showed the most leniency indicating that those with the most to gain by overrating themselves did so. Perhaps this resulted from use of the self-appraisal as part of the formal evaluation program which could affect promotional opportunities for the subjects. This feature was not included in the present study.

Teel (1978) also used self-appraisals in which the employee and supervisor compared ratings as part of a formal evaluation procedure. Although Teel did not conduct an experimental study, but only reported general findings from personal experience, he reported that self-ratings were not lenient using this procedure. He attributed this to two factors. First, that employees are more aware than the supervisor of their own poor performance. Second, he alluded to consequences provided by the supervisor for high self-ratings.
"By underrating their performance, subordinates structure the situation so that the superior is more likely to praise than to criticize them...overrating one's self invites criticism. Consequently, many subordinates underrate themselves to invite favorable comments from their supervisors." (p. 366 & 367).

A topic for future research could examine leniency of self-ratings under conditions where the ratings are used as tools for management decision making as compared to "research only" and "justification" conditions. Such research could help to resolve the discrepancy between the findings of the present study and those of Thornton (1968) and Teel (1978).

Galizio (1979) documented the effect of consequences on instruction-following behavior. In one experiment, Galizio demonstrated the effects of a past reinforcement history for instruction-following on subsequent behavior. When given accurate instructions, subjects showed stable responding under multiple free-operant avoidance schedules where monetary loss served as the aversive event. The consequence of monetary loss for inaccurate responding was then withdrawn without the subject's knowledge. Under this condition, subjects continued to respond according to the past instructions although no contact with the contingencies was made. This showed evidence of past contingencies controlling current instruction-following behavior. These findings provide an explanation for the results shown in the present study. In both studies, subjects made responses according to consequences which were implied based on exposure to previous contingencies. Presumably, in the present study, when employees responded differently according to the instructional condition in effect, past reinforcement histories along with the
implied consequences in each instructional condition helped to
determine their response.

The mean self-ratings for employees in the bus driver
classification did not follow the trends shown by the other
classifications. In the control condition, bus drivers rated
themselves slightly lower than the supervisor; in the experimental
condition, bus drivers rated themselves higher than the supervisor.
Two factors may be responsible for this. Bus drivers perform their job
away from the school site; the supervisor does not have the opportunity
to observe them closely on the job. According to Borman (1974),
agreement shouldn't be expected from different levels who have
different orientations towards ratees and observe them in different
settings. In addition, after the experiment had been completed, it was
revealed that the supervisor had been so pleased with the BOS form for
the drivers, that he used it in conjunction with the regular school
district form in completing the annual evaluation of some of the
drivers. This resulted in some of the drivers assigned to the control
condition discussing their self-appraisal with the supervisor, and some
of the drivers assigned to the experimental condition who had already
received their formal evaluation before the self-rating was made, not
getting the opportunity to discuss the self-rating with the supervisor.
It is possible that, when drivers who had not yet completed their self-
rating saw that the instructional conditions assigned to them did not
describe the conditions in effect, the instructions failed to control
the behavior.

In regard to inaccurate instructions, as with the bus drivers,
Galizio (1979) documented that instruction-following behavior decreases
in the presence of instructions shown to be incorrect. When accurate instructions and positive consequences are reinstated, instruction-following behavior resumes. In other words, accurate instructions serve as a discriminative stimulus for appropriate behavior. When instructions are incorrect, and the consequences for following the instructions are negative, the instructions no longer serve as a discriminative stimulus associated with reinforcement. Therefore, in the presence of inaccurate instructions, instruction-following behavior decreases. This could serve to explain why bus driver self-ratings did not show trends similar to the self-ratings in the other classifications.

An interesting phenomenon occurred with the supervisor ratings. Despite knowledge of the research question and instructional conditions, three supervisors rated employees significantly higher in the experimental condition than in the control condition. Similar results were reported by Berkshire and Highland (1953) and Taylor and Wherry (1951). In each of these studies, supervisors rated subordinates lower in conditions where the results were used for research purposes only. Sharon and Bartlett (1969) included a "justification" condition where subjects were instructed to justify the rating to the ratee. Under this condition, ratings were higher than in other conditions where "justification" was not included. These results support those found in the present study. Of the five supervisors who rated employees in both conditions, three supervisors rated employees who were instructed to meet and discuss ratings with them higher than those who were not instructed to meet with them.
Miner (1968) also described similar results in a study where managers filled out evaluations without showing them to the ratees. Later, the same managers filled out a second appraisal which involved a discussion with the ratee. The mean ratings rose considerably when the discussion was included. Miner observed that "when faced with the prospect of making face-to-face negative comments, many managers avoided the problem by inflating their ratings." (p 88). Apparently, in addition to the employees, the supervisors also responded differentially to the instructional conditions in the present study. However, the employees and the supervisors responded differently to the instructions to meet and justify their respective ratings. The employees gave lower self-ratings when justification was instructed whereas the supervisors gave higher ratings when justification was instructed.

These results indicate the potency of instructional conditions on human operant behavior. Considering that supervisors also showed some evidence of instructional control, caution seems indicated in using supervisor ratings as a criterion or comparison measure. Often, self-ratings are determined lenient only when compared to supervisor ratings, regardless of the conditions which may affect them. As an alternative, self-ratings in the present study were contrasted across instructional conditions without reference to supervisor ratings.

True leniency error was not assessed due to the absence of actual performance measures. Therefore, the degree to which instructional conditions affect leniency in self-appraisals as compared to each individual's performance on the job was not examined. Future research should concentrate on methods to measure actual job performance which

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
could then be compared with ratings by various rating sources within organizations to determine if lenient ratings are justified. After examining the psychometric characteristics of the BOS without access to actual performance measures, Bolt (1984) concluded that "Overcoming the hurdles to obtaining true performance measures may be a more productive avenue for research than attempting to overcome the hurdles of subjective rating scales." (p. 45).

Skinner (1974) asserted that following instructions is a form of rule-governed behavior which is controlled by its consequences. Implied consequences were shown to also affect behavior in the present study. However, in a self-appraisal program which is repeatedly administered in an organizational setting, actual consequences are delivered each time self-ratings are discussed with the supervisor. The present study did not examine the effects of actual consequences delivered by the supervisor. The effects of such consequences could be crucial in determining leniency in subsequent self-ratings. Teel (1978) developed a system where rating discrepancies of one point were assigned to the higher rating regardless of the source. Rating discrepancies of two points, however, were discussed by the employee and the supervisor until agreement was reached on a new rating. This system resulted in very few rating discrepancies of two points or more. Waldman and Thornton (1979) also suggested that repeated administrations of a self-appraisal program should help to improve communication between employees and supervisors and reduce rating discrepancies between rating sources over time. In the future, research designed to study the actual effects of experimentally
controlled consequences on leniency in self-ratings could provide additional information on the variables which control lenient ratings.

Rather than documenting that rating errors occur, future research should concentrate on studying the environmental variables which affect the rating process. The present research leaves many questions regarding self-appraisals unanswered, but shows the utility of using an experimental analysis. After controlling the conditions associated with lenient ratings, self-appraisals could be used successfully as a means of obtaining more accurate performance measures, and ultimately, to improve job performance.
APPENDIX A

BEHAVIORAL OBSERVATION SCALES
PERFORMANCE APPRAISAL FOR BUS DRIVER POSITIONS

EMPLOYEE NAME__________________________  RATER NAME ____________________________

<table>
<thead>
<tr>
<th>ALMOST NEVER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>

WORK HABITS

1. Maintains a good personal driving record.  
   1  2  3  4  5

2. Can be depended on to be at work each day and on time.  
   1  2  3  4  5

3. Maintains a regular time schedule on the route with minor day-to-day variations.  
   1  2  3  4  5

4. Generally displays a positive attitude.  
   1  2  3  4  5

5. Refers a problem to the supervisor when appropriate.  
   1  2  3  4  5

6. Functions appropriately in stressful situations.  
   1  2  3  4  5

7. Calls ahead if expects to be late or absent from work.  
   1  2  3  4  5

SAFETY PROCEDURES

8. Allows traffic to clear before putting out stop sign when unloading the bus.  
   1  2  3  4  5

9. Keeps noise volume on the bus low enough to hear emergency vehicles.  
   1  2  3  4  5

10. Students are not allowed to approach the bus until it has stopped and the door is opened.  
    1  2  3  4  5

11. Knows location and operation of safety equipment.  
    1  2  3  4  5

12. Drives within the legal speed limit.  
    1  2  3  4  5

13. Allows students to leave the bus only after all traffic has stopped.  
    1  2  3  4  5

14. Starts the bus only after all the students are seated.  
    1  2  3  4  5

15. Keeps eyes moving from the road, to the mirrors, to the gauges and to the students at all times.  
    1  2  3  4  5

16. Allows a student to get off at a different stop only if they have a bus pass.  
    1  2  3  4  5

17. Takes extra precautions when driving in hazardous weather.  
    1  2  3  4  5

18. Incorporates defensive driving techniques on all driving assignments.  
    1  2  3  4  5
### INTERPERSONAL RELATIONS

**19. Carries out requests by the supervisor in a cooperative manner.**

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

**20. Works well with principals and teachers in resolving problems with students.**

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

**21. Does not let personal differences with other drivers interfere with work.**

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

**22. Knows that his/her communication with the public influences the community's concept of the District.**

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

### STUDENT INTERACTION

**23. Refrains from threatening disciplinary action which he/she does not carry out.**

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

**24. Consistently enforces District standards for student behavior on the bus.**

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

**25. Shows politeness in dealing with students without signs of anger or impatience.**

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

**26. Does not become intimidated in difficult situations with the students.**

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

**27. Discusses student issues confidentially when possible.**

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

### BUS OPERATIONS AND REGULATIONS

**28. Successfully completes a route or extra trip when not familiar with the route.**

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

**29. Familiar with rules and regulations listed in the bus driver handbook.**

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

**30. Sweeps the bus daily and washes the bus at least once a month.**

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

**31. Is familiar with, and can operate, any bus in the fleet when requested.**

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

**32. Stays on the bus during bus warm-up.**

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

**33. Bus condition reports are filled out and submitted when mechanical problems exist on the bus.**

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

**34. Warms the bus up for 5 to 10 minutes, and idles it down for 3 minutes.**

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

**35. Allows other objects to contact the bus which may harm it.**

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

**36. Operates the bus in a mechanically correct manner.**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>
PERFORMANCE APPRAISAL FOR CUSTODIAN POSITIONS

EMPLOYEE NAME_________________________  RATER NAME ________________________________

ALMOST NEVER  1  2  3  4  5  ALMOST ALWAYS

WORK HABITS

1. Generally displays a positive attitude.  1  2  3  4  5
2. Can be depended on to be at work each day and on time.  1  2  3  4  5
3. Demonstrates a willingness to accomplish assigned tasks and special projects.  1  2  3  4  5
4. Does not spend too much time off-task during the work shift.  1  2  3  4  5
5. Calls ahead if he expects to be late or absent.  1  2  3  4  5
6. Makes up time when reporting late to work.  1  2  3  4  5
7. Observes time limits for breaks and lunches  1  2  3  4  5
8. Follows directions from head custodian and/or supervisor.  1  2  3  4  5

INDEPENDENCE AND SELF-MOTIVATION

9. Requires little or no supervision in completing daily work assignment.  1  2  3  4  5
10. Completes work which needs to be done but which has not been assigned.  1  2  3  4  5
11. Alters work schedule and job priorities to accommodate special requests.  1  2  3  4  5
12. Plans work schedule to most efficiently complete work assignments.  1  2  3  4  5
13. Volunteers to use special or individual skills.  1  2  3  4  5
14. Refers major problems to head custodian and/or supervisor.  1  2  3  4  5

PUBLIC RELATIONS

15. Is courteous with community groups using District facilities.  1  2  3  4  5
16. Knows that the condition of the schools reflects the District's image and works to improve it.  1  2  3  4  5

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
**INTERPERSONAL RELATIONSHIPS**

17. Responds positively to requests by teachers and principal and cooperates when possible.  
18. Brings suggestions to head custodian before trying them out.  
19. Is willing to discuss complaints about his work.  
20. Works cooperatively with other custodians when working in a team.  
21. Doesn't allow personal differences with other staff to interfere with work.  
22. Works patiently with and trains student helpers.

**JOB ASSIGNMENTS**

23. Assists in outside maintenance as needed.  
24. Cleans windows when needed.  
25. Re-stocks towels and tissues daily in bathroom.  
26. Makes an effort to keep sidewalks clean.  
27. Thoroughly cleans and organizes each assigned room.  
28. Checks lights and replaces tubes or exchanges ballasts when necessary.  
29. Reports and repairs damage to facilities.  
30. Assists in summer maintenance projects.

**SAFETY AND SECURITY**

31. Knows location and operation of safety equipment.  
32. Keeps supplies looked up at all times.  
33. Combines cleaning and sanitizing solutions correctly.  
34. Dilutes cleaning solutions according to directions.  
35. Checks that windows and doors are shut and locked when leaving a room and before leaving at the end of the shift.  
36. Opens locked rooms for students or the public when appropriate.  
37. Checks rooms for cleanliness and security after community groups have left them.  
38. Does not allow other people to use or make copies of keys.
## PERFORMANCE APPRAISAL FOR HEAD CUSTODIAN POSITIONS

**EMPLOYEE NAME______________________________ RATER NAME __________________**

<table>
<thead>
<tr>
<th>ALMOST NEVER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
</table>

### JOB RESPONSIBILITIES

1. Orders supplies when needed.  
   1 2 3 4 5

2. Anticipates and checks for areas that may need attention.  
   1 2 3 4 5

3. Checks and repairs furnace or boiler when necessary.  
   1 2 3 4 5

4. Calls outside contractors when needed for repairs.  
   1 2 3 4 5

5. Monitors and adjusts heat in all building areas.  
   1 2 3 4 5

6. Discusses and helps to resolve complaints from school staff and the community.  
   1 2 3 4 5

7. Refers problems to supervisor when appropriate.  
   1 2 3 4 5

8. Insures that all clocks are accurate and in good working order.  
   1 2 3 4 5

### SUPERVISION

9. Delivers complaints to night custodians and provides discipline when appropriate.  
   1 2 3 4 5

10. Delegates tasks to night custodians  
    1 2 3 4 5

11. Plans summer maintenance projects and assigns duties to night custodians.  
    1 2 3 4 5

12. Monitors night custodian's work.  
    1 2 3 4 5

13. Trains substitutes and night custodians well.  
    1 2 3 4 5

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
PERFORMANCE APPRAISAL FOR SECRETARIAL POSITIONS

<table>
<thead>
<tr>
<th>WORK HABITS</th>
<th>ALMOST NEVER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Completes projects or tasks by deadline.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2. Functions appropriately in stressful situations with constant interruptions.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3. Demonstrates a willingness to accomplish assigned tasks and special projects.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4. Calls ahead if she expects to be late or absent from work.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5. Reports to work each day and on time.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6. Finds tasks which need to be done without having to be told.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7. Proofreads and corrects all material before sending out.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>8. Takes and deliver messages accurately and promptly.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>9. Enforces school and District policy consistently.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10. Generally displays a positive attitude.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

ADMINISTRATION AND SUPERVISION

<table>
<thead>
<tr>
<th>ADMINISTRATION AND SUPERVISION</th>
<th>ALMOST NEVER</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Makes suggestions for improved policies or procedures.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>12. Brings potential problems to the principal's attention.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>13. Correctly determines when to refer a problem to the principal.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>14. Makes decisions consistent with the principal's judgement in his absence.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>15. Delegates tasks as needed.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>16. Follows through with little guidance on an assigned project.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>17. Keeps the principal's confidential information to herself.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>18. Supervises and trains office personnel and student helpers.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>19. Composes letters when given a brief outline of their content.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
### Interpersonal Relationships

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20. Doesn't allow personal differences with other staff to interfere with work.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>21. Shows politeness with others without signs of anger or impatience.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### Organizational Abilities and Technical Skills

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22. Locates materials or supplies quickly when needed.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>23. Keeps all bookkeeping and other records up-to-date and accurate.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>24. Maintains inventory and orders supplies when necessary.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>25. Prioritizes tasks to meet changing deadlines.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>26. Organizes and prepares reports accurately with minimal supervision.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>27. Proficient with business machines relevant on the job.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### Public Relations

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>28. Listens calmly and patiently to angry parents.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>29. Knows that her communication with the public influences the community's concept of the District.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

### Student Relations and Procedures

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30. Disciplines students firmly and consistently when necessary.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>31. Dispenses medication according to school policy.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>32. Attempts to contact a parent quickly when a student is seriously injured or ill.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>33. Discusses student issues confidentially when possible.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>34. Keeps student records accurate and up-to-date.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>35. Follows up on needed information when enrolling students.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
PERFORMANCE APPRAISAL FOR AIDE POSITIONS

EMPLOYEE NAME ___________________________ RATER NAME ________________________

WORK HABITS

1. Generally displays a positive attitude. 1 2 3 4 5
2. Can be depended on to be at work each day and on time. 1 2 3 4 5
3. Demonstrates a willingness to accomplish assigned tasks and special projects without procrastinating. 1 2 3 4 5
4. Stays calm in a crisis situation. 1 2 3 4 5
5. Completes tasks despite frequent interruptions. 1 2 3 4 5
6. Trustworthy and accurate with money. 1 2 3 4 5
7. Finds tasks which need to be done without having to be told. 1 2 3 4 5
8. Proficient with business machines relevant to the job. 1 2 3 4 5
9. Calls ahead if she expects to be late or absent from work. 1 2 3 4 5
10. Uses school policy, rather than own judgement, when making decisions. 1 2 3 4 5

STUDENT INTERACTION

11. Offers assistance to students whenever possible. 1 2 3 4 5
12. Praises students for accomplishments. 1 2 3 4 5
13. Disciplines students firmly and consistently according to school policy. 1 2 3 4 5
14. Calmly handles sick or injured students; dispenses medication according to policy. 1 2 3 4 5
15. Gets to know the students as individuals. 1 2 3 4 5
16. Instructs and/or supervises students with minimal supervision. 1 2 3 4 5
17. Maintains discipline of the students in the absence of the teacher. 1 2 3 4 5
18. Assists in setting educational goals for students within their range of ability. 1 2 3 4 5
**INTERPERSONAL RELATIONSHIPS**

19. Willingly accepts changes introduced by the teacher, supervisor, or co-workers.  
20. Tactfully makes suggestions for program or procedural changes.  
21. Takes criticism in a non-defensive manner.  
22. Works in concert with the teacher to achieve educational goals.  
23. Refrains from showing anger and impatience when dealing with others.  
24. Knows that her interactions with the public influences the community's concept of the District.  
25. Shows politeness in day-to-day relations.  
26. Shows that personal differences with other staff do not interfere with her work.  

**EDUCATIONAL MATERIALS AND LESSONS**

27. Files and retrieves needed materials quickly.  
28. Organizes and types materials accurately when requested.  
29. Is prepared with, and knowledgeable of, lesson materials when meeting with students.  
30. Follows guidelines and lesson plans of the teacher regardless of her own opinions of the procedure.  
31. Conducts, scores and records tests according to written and verbal instructions.  
32. Has curriculum and lesson plans approved by teacher.  

**COMMUNICATION SKILLS**

33. Demonstrates good empathic skills.  
34. Keeps student issues confidential when possible.  
35. Directs others to appropriate persons when additional information is requested.  
36. Asks questions when unsure of procedures.
APPENDIX B

INSTRUCTIONS FOR PERFORMANCE APPRAISAL
INSTRUCTIONS FOR PERFORMANCE APPRAISAL

PLEASE RATE YOURSELF ON THE ATTACHED FORM AS ACCURATELY AND AS HONESTLY AS POSSIBLE.

CIRCLE THE "1" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 0-50% OF THE TIME.

CIRCLE THE "2" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 50-65% OF THE TIME.

CIRCLE THE "3" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 65-80% OF THE TIME.

CIRCLE THE "4" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 80-90% OF THE TIME.

CIRCLE THE "5" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 90-100% OF THE TIME.

IF THE ITEM DOESN'T PERTAIN TO YOU, LEAVE IT BLANK.

YOU DO NOT NEED TO MEET WITH YOUR SUPERVISOR TO REVIEW THIS FORM.

PLEASE RETURN THIS TO EITHER YOUR SUPERVISOR OR TO THE MAIN OFFICE IN YOUR BUILDING NO LATER THAN FRIDAY, JUNE 7TH. IF YOU HAVE A PROBLEM WITH THIS DATE, AND HAVEN'T TURNED YOUR FORM IN BY THAT TIME, I WILL CONTACT YOU BY PHONE.

THANK YOU FOR YOUR TIME AND COOPERATION!!!
INSTRUCTIONS FOR PERFORMANCE APPRAISAL

PLEASE RATE YOURSELF ON THE ATTACHED FORM AS ACCURATELY AND AS HONESTLY AS POSSIBLE.

CIRCLE THE "1" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 0-50% OF THE TIME.

CIRCLE THE "2" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 50-65% OF THE TIME.

CIRCLE THE "3" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 65-80% OF THE TIME.

CIRCLE THE "4" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 80-90% OF THE TIME.

CIRCLE THE "5" IF YOU ENGAGE IN THE INDICATED BEHAVIOR 90-100% OF THE TIME.

IF THE ITEM DOESN'T PERTAIN TO YOU, LEAVE IT BLANK.

YOU SHOULD SCHEDULE AN APPOINTMENT WITH YOUR SUPERVISOR TO REVIEW THIS FORM. THIS SHOULD TAKE NO LONGER THAN 15 MINUTES. SIMPLY REVIEW THE AREAS WHERE YOU AGREE AND DISAGREE.

PLEASE RETURN THIS TO EITHER YOUR SUPERVISOR OR TO THE MAIN OFFICE IN YOUR BUILDING NO LATER THAN FRIDAY, JUNE 7TH. IF YOU HAVE A PROBLEM WITH THIS DATE, AND HAVEN'T TurnED YOUR FORM IN BY THAT TIME, I WILL CONTACT YOU BY PHONE.

THANK YOU FOR YOUR TIME AND COOPERATION!!!
BIBLIOGRAPHY


