Psychology Theses & Dissertations   Psychology

Summer 1982

Improving the Accuracy of Performance Evaluations: A Comparison of Three Methods of Performance Appraiser Training

Jerry Willard Hedge
Old Dominion University

Recommended Citation

IMPROVING THE ACCURACY OF PERFORMANCE EVALUATIONS:

A COMPARISON OF THREE METHODS OF

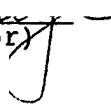PERFORMANCE APPRAISER TRAINING

by

Jerry Willard Hedge
M.S., May, 1980, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

INDUSTRIAL/ORGANIZATIONAL PSYCHOLOGY

Old Dominion University
August, 1982

Approved by:

Michael J. Kavanagh (Director)

ABSTRACT

IMPROVING THE ACCURACY OF PERFORMANCE EVALUATIONS:
A COMPARISON OF THREE METHODS OF
PERFORMANCE APPRAISER TRAINING

Jerry W. Hedge
Old Dominion University
Director: Dr. Michael J. Kavanagh

Researchers in the area of rater training have relied almost exclusively on rater error measures to assess training effectiveness. A reduction in rater tendency to commit these errors subsequent to training is viewed as evidence that these raters have become more accurate in rating their employees. This assumed relationship between rater errors and rating accuracy has recently been questioned. This uncertain relationship between psychometric errors and accuracy was the focus of the current research effort. Supervisory personnel were trained under one of three training programs (psychometric error training, observation training, or decision-making training). Halo, leniency, range restriction and accuracy measures were collected before, and after training from the three training groups, and a no-training control group. The results suggested that while psychometric error training reduced rater errors, it also detrimentally affected rating accuracy. However, observation and decision-making training had no effect on, or increased error rate, but caused performance rating

accuracy to increase after training. The need for a reconceptualization of rater training content and focus was discussed.

DEDICATION

To Carolyn, for her love, support, encouragement and belief in me.

ii

ACKNOWLEDGMENTS

As is always the case when a project of this scope is undertaken, there are many people instrumental in its successful completion. First, the contributions of the faculty, graduate students and staff of the Center for Applied Psychological Studies are gratefully acknowledged.

I would also like to thank the members of my dissertation and defense committees, Drs. Glynn Coates, Glenn DeBiasi, Michael Kavanagh, Raymond Kirby, Bruce McAfee and Ben Morgan, Jr. In addition, I am indebted to Drs. Coates, Kirby, Morgan and Kavanagh, who have played an important role in my professional development while in graduate school.

A special note of appreciation is extended to Mr. J. Raymond Comstock, a friend and colleague who has contributed in many important ways, not only to this project, but to my personal and professional development as well.

I am also grateful to Mr. Woody Turpin, who "starred" as director, cameraman, technical advisor, and all-around handyman during the filming of my training programs. In addition, I am indebted to Ms. Tina Broome, Personnel

iii

Training Manager at Old Dominion, for her support and assistance in organizing the workshops and supplying needed information. I would also like to thank Pete Mikulka for betting against me.

Finally, I would like to thank my wife, Carole, and my children, Jesse and Jordan, for their encouragement, understanding and support throughout this project.

TABLE OF CONTENTS

v

vi

vii

LIST OF TABLES

viii

LIST OF FIGURES

x

Introduction

For many years, psychologists have realized the importance of performance measurement in organizational settings. Unfortunately, this knowledge has also been accompanied by a realization that the accurate measurement of job performance is not a simple task. For this reason, some have suggested that measurement should focus on objective indices of job performance, such as production data (i.e., units produced, sales volume) or personnel data (i.e., absenteeism, turnover). While the use of these job performance measures is a logical choice (in that they serve as good indicators of organizational effectiveness), they typically fail to measure individual performance effectiveness. There are several reasons why this is so.

First, there are many situational factors beyond an employee's control (i.e., equipment malfunction, size of a salesperson's territory) that will impact directly on these data. In addition, cost-related measures are often difficult to obtain on employees in many jobs. Consequently, these measures are often useless as performance criteria, and as a result, sole reliance on judgmental indices (such as ratings) has frequently occurred. Use of subjective criteria has not occurred by

default alone, however, but also through the belief that judgmental indices of performance can reflect the complexity of the job, are more likely to minimize situational factors, and can measure more directly what an employee does on the job.

Still, this widespread use of job performance ratings has generated numerous questions concerning the reliability, validity and accuracy of such "subjective" measures of performance. Consequently, an enormous amount of research has been conducted addressing the use of judgmental measures of performance. Researchers have focused on, among other things, the format, the content, or the most appropriate source of appraisal in hopes of answering these questions. Landy and Farr (1980) have provided an extensive review of these efforts. A recent approach to this problem has been to examine the effects of rater training on rating errors and rating accuracy.

Rater training research has typically been concerned with providing information (of one sort or another) to performance appraisers with the hope that they will become "better," "more effective" evaluators of their employees' job performance. "Better" and "more effective" have most frequently been measured by evaluating the frequency of occurrence of a variety of so-called "rating errors." The most often used error measures have been labeled <u>halo error</u> (inappropriate generalization from one aspect of a person's job performance to all aspects of a person's performance),

leniency error (a tendency of the rater to rate all his or her employees too high) and range restriction error (failure of the rater to discriminate among his or her employees in terms of their respective performance levels). Numerous other rating errors have also been defined and measured, including first impression error (a tendency of the rater to evaluate someone on the basis of judgments made primarily after an initial meeting), similarity error (a tendency on the part of the rater to judge more favorably those persons he/she perceives as similar to himself/herself), and contrast error (a tendency by the rater to judge an employee in comparison to the most recently evaluated employee).

In addition to psychometric error measures, the reliability and validity of the ratings have been used as indices of training effectiveness. Reliability information typically is collected in reference to agreement between raters (interrater reliability), while validity information has been gathered by means of a comparison of job performance ratings to known performance scores (or "normative true scores"). Rating validities have also been estimated by using the Kavanagh, MacKinney and Wolins (1971) Analysis of Variance approach, thus providing convergent and discriminate validity indices.

In general, researchers in the area of rater training have relied almost exclusively on rater error measures to assess training effectiveness. A reduction in rater tendency to commit these errors subsequent to training is

viewed as evidence that these raters have become more accurate in rating their employees. This assumed relationship between rater errors and rating accuracy has recently been questioned (Borman, 1975; Bernardin & Pence, 1980). This uncertain relationship between rating errors and accuracy is the focus of the current research effort.

Rater Training to Reduce Psychometric Errors

Stockford and Bissell (1949) and Levine and Butler (1952) provided some of the earliest information on attempts at improving performance ratings by training performance appraisers. Stockford and Bissell (1949) were concerned with making merit ratings more objective in the Lockheed Aircraft Corporation. Toward this end they undertook a series of studies to determine the degree to which certain weaknesses inherent in the current ratings could be reduced or overcome by designing a new scale and training supervisors in the principles and techniques of rating.

Supervisors received either a two-hour general orientation to the new form, or six hours of instruction on the philosophy and principles of rating, participation in the selection of items to be used in the scale, and feedback about how they were rating. The six-hour rater training resulted in significantly more reliable ratings, and significantly fewer halo errors when compared with the general orientation training group. However, the fact that one group received three times as much training time as the

other group reduces the researcher's ability to infer positive changes in rater behavior as a result of training content.

Levine and Butler (1952) dealt with supervisors in a large manufacturing plant who had been overrating employees in the higher job grades, yet underrating employees in the lower job grades. These researchers classified this problem as a type of halo error (in reality, this problem is more accurately described as a context error). To reduce or eliminate this problem, Levine and Butler (1952) randomly assigned supervisors to a control group, a lecture group or a discussion group. While the control group received no training or information, the lecture group was presented detailed information on rating theory and technique, including the causes of the previous problems, and how to correct them. Supervisors in the discussion group met as a group and discussed the nature of the problem and how to resolve it. A discussion leader was present, but acted only as a moderator. Rating data collected subsequent to training showed only supervisors participating in the discussion group changed their ratings of subordinates in an appropriate manner.

Since these initial studies, rater training research typically has been designed to determine whether a particular type of training program will significantly reduce certain rating errors when an experimental group is compared to a no-training control group. In addition, most

of these studies have utilized a posttest-only design. Both of these approaches must be seen as serious deficiencies in rater training research. The only conclusion that can be drawn from the first design is that something is better than nothing, while the second design (posttest-only) ignores pre-training data.

For example, Brown (1968), trained a group of student nurses in an attempt to reduce rating error. The one-hour training session consisted of discussion of: (1) the different kinds of rating scales and rating procedures, (2) the problems in obtaining sound ratings, and (3) the errors often committed by supervisors. They were also given some practice using the rating scale. The data collected (peer ratings) from the trained group were then compared with data gathered from a group of untrained raters. The student nurses who had gone through training subsequently were able to discriminate better between employees, which was interpreted by Brown (1968) as proof that halo had been reduced.

Several other empirical studies have focused on the reduction of halo as a measure of training program success. Taylor and Hastman (1956) compared four groups of subjects that differed on the method of completing ratings (rate all ratees on one dimension before proceeding to the next dimension, or rate a ratee on all dimensions and then move to the next ratee). They found no differences between groups using a measure of halo, but they concluded that, in fact, none of the groups displayed a tendency toward halo.

Borman (1975) also trained supervisors to reduce halo error, but developed only a five-minute training program for this purpose. Using a one-group pretest/posttest design, Borman demonstrated a significant reduction in halo errors after this short period of training. The main focus of the Borman (1975) study was on the side effects of rater training—on how training effects reliability and validity of the ratings. These criteria of rating quality will be discussed more fully in the next section.

In a much more extensive, systematic approach to training performance appraisers, Latham, Wexley and Pursell (1975) trained employees in a large corporation to minimize halo, first impression, similarity, and contrast errors; and then, measured the extent of these errors six months after training. Raters received training via a workshop or group discussion approach. Their workshop treatment provided participants with an opportunity to practice observing and rating actual videotaped ratees. The group discussion format was similar to the Levine and Butler (1952) discussion group approach.

Latham and his colleagues (1975) evaluated these training approaches by comparing both the extent of errors and the subjective reactions to training with a no-training control group. Both the workshop and the group discussion approaches resulted in the reduction or elimination of all four rating errors, while the control group exhibited significant similarity and contrast errors. In addition,

employees reacted more favorably to the workshop approach, and Latham et al. (1975) interpreted these findings as strong support for the workshop training. Once again, a posttest-only design was used.

A recent rater training study by Faye and Latham (1982) also used the Latham et al. (1975) workshop approach. Half of the subjects (business students) received training, while the other half served as a control group. Subsequent to training, all subjects rated videotapes of applicants in job interviews using a trait scale, a Behavioral Expectation Scale (BES) or a Behavioral Observation Scale (BOS). Results showed that rating errors were reduced regardless of the rating scale used, although trainees who used the BES or BOS committed fewer psychometric errors than did trainees using the trait scale.

A different approach to training raters has been used by Bernardin and Walters (1977). They asked college students to record behavioral examples of teacher performance during the semester in a diary, and then use the information as an aid in making detailed performance ratings at the end of the term. Four experimental groups were used in the study. All four groups received some variation of the typical psychometric error-reduction training. The diary-keeping group received a one-hour lecture on rating errors at the beginning of the semester, with practice using the scale. A second group received similar training at the beginning of the semester, but without practice using the

scale. Group Three received the same treatment as Group Two, but immediately prior to formal evaluation. Group Four served as a quasi-control, receiving only minimal instructions prior to rating. The results revealed that the diary-keeping group, which had received psychometric error training and exposure to the scale early in the semester, showed significantly less halo and leniency than all other groups. This study also utilized a posttest-only design.

A later study by Bernardin (1978) compared a short psychometric error training program (similar to Borman's 1975 study) with a more comprehensive approach (consisting of a one-hour training session). Halo, leniency and central tendency were measured across three consecutive rating periods. Immediately after training, the one-hour training was found to be significantly more effective in reducing both halo and leniency than the five-minute training, and both groups were superior to the two control groups. However, after 13 weeks, the training effectiveness had dissipated and no differences were found for the four groups.

Ivancevich (1979) also evaluated the effects of psychometric error training, using a longitudinal design. An intense training group received a lecture on psychometric errors, and how to avoid them, as well as practice evaluating high and low performers. A lecture/discussion format provided a second group with psychometric error training. Each of these groups received approximately 14

hours of training. The tendency for raters to exhibit halo and leniency errors was measured six months before training, and six and twelve months after training. The findings revealed that the intense training group was superior to the discussion group and a control group in reducing halo and leniency error after six months. However, at twelve months after training, much of the training effect had dissipated for the intense training group.

With a somewhat similar focus, Warmke and Billings (1979) evaluated the generalizability of effects from lecture and group discussion formats by comparing experimental ratings collected immediately after training, with administrative ratings collected two months later. Higher levels of halo were found in the administrative ratings compared to those collected experimentally, and no differences were found between trained and untrained groups.

Not all research on psychometric error training has produced such significant training effects as suggested by the majority of the studies cited. For example, Vance, Kuhnert and Farr (1978) trained students using the typical lecture/discussion format, yet found no difference in level of halo or leniency between a trained and no-training control group. In addition, Sauser and Pond (1981) found no halo or leniency differences between subjects who were given psychometric error training (a two-hour lecture/discussion/-practice format) and subjects who received no such training. Still, while several studies have suggested no rating error

differences between trained and untrained raters, rater training has generally been shown to be effective in reducing rating errors soon after training.

Exactly how and why rater training reduces psychometric errors in ratings has become a recent topic of discussion. Bernardin and his colleagues (Bernardin, 1978; Bernardin & Pence, 1980; Bernardin & Buckley, 1981) have contended that typical rater training programs focus on changing rater response distributions by presenting certain rating distributions as representative of rating errors. Thus, training causes the raters to adopt new (but possibly inappropriate) response sets in order to eliminate these errors when they rate their employees. Bernardin has also suggested that rating errors can be reduced or eliminated with training programs of relatively short (one hour or less) duration.

Latham and his colleagues (Latham & Wexley, 1981; Faye & Latham, 1982) have disagreed with Bernardin's approach, and have suggested that rating errors are well-ingrained habits that are quite difficult to extinguish. Consequently, it should take many hours of training to eliminate these rating errors. In addition, these researchers have concluded that training must include practice in observing ratees committing these errors, rather than the presentation of appropriate and inappropriate response sets. Accordingly, the results of such efforts are believed to include not only a reduction in rating errors,

but an improvement in rating accuracy. While this continuing debate may serve to generate additional research questions, when viewed from the perspective of rating accuracy, its worth as a central research issue may be rather limited. The limitations of such a perspective are discussed in the following section.

Accuracy and the Judgment of Performance

While researchers have evaluated training effects on psychometric errors, reliability, and validity, it is apparent from the foregoing literature review that the majority of research has focused on psychometric errors. Consequently, researchers have assumed that the more accurate ratings are those with reduced levels of psychometric errors. To clarify this reasoning, a discussion of the relationship between psychometric errors, reliability, validity, and accuracy is needed.

The most straightforward way to view the relationship between these variables is through a discussion of measurement theory. Central to this discussion is the recognition that performance ratings are not error-free. Rather, performance measurement contains both elements of error and elements of truth. In this context, reliability can be defined as the proportion of true variance in a set of ratings, while validity can be defined as the proportion of true variance that is relevant to the purpose of the measurement procedure (Campbell, 1976). This implies that true variance can be separated into two components,

systematic relevant variance, and systematic error variance. Thus, observed variance in a set of ratings is determined by the proportion of true variance, systematic error variance and random error variance. As random error variance decreases, the ratings become more reliable, and the potential for valid variance increases. However, since systematic variance may be relevant or irrelevant, high reliability does not guarantee high validity.

With this framework in mind, rater training researchers have assumed that rating errors represent error variance, and training that reduces these rating errors should have a positive effect on validity and accuracy. However, the relationship between validity and accuracy is not as well defined. Generally, these two terms are used synonymously in the literature, yet theoretically (as Guion, 1965, noted), these terms are not equivalent. This is so, because systematic errors in ratings can contribute to validity as much as, or more than true variance. Thus, perfect validity would not be evidence of accuracy. Still, when defined operationally, accuracy and construct validity become synonymous, especially when construct validation is approached from a multi-trait-multi-rater perspective (see Kavanagh, et al., 1971).

Unfortunately, a major deficiency in most of these studies on performance rating errors, is that the researcher is required to assume that rating errors and random error variance are equivalent. However, what has been termed

rating errors might be conceptualized more appropriately as rating effects. As Bingham (1939) noted, it is not at all clear if halo is valid or invalid. The same line of reasoning applies to the other rating errors. It may be that a portion of the rating effect (i.e., leniency) represents true score variance, and the remainder is error variance.

Still, as Borman (1979a) noted, past rater training research has failed to investigate directly rating accuracy or validity. In fact, only three published studies dealing with rater training (Borman, 1975; Borman, 1979a; Bernardin & Pence, 1980) have used accuracy as a dependent measure. Before discussing each of these studies in detail, it is useful to focus on the concept of accuracy, and its use as a dependent measure.

Accuracy as a measure of performance. Given the foregoing discussion, the usefulness of information concerning how reliable our measures of performance are, and whether raters exhibit a tendency toward leniency, halo and the like is apparent. Yet, we are ultimately concerned with the accuracy of our performance measures. As Borman (1980) noted, accuracy is critical in personnel research involving employees' performance as a criterion. In regards to administrative ratings, accuracy is necessary in ensuring fair personnel decisions made on the basis of performance appraisals--be it for promotion, merit pay increases or training purposes.

Still, a major problem confronts the researcher attempting to measure accuracy. This problem is related to the classic criterion problem in Industrial/Organizational psychology, namely, how we arrive at "true scores" indicative of the dimensions of job performance. In a recent paper, Borman (1930) discussed three possible approaches to studying performance rating accuracy.

One approach to addressing the criterion problem has been through the use of "paper people." This method involves the development of vignettes or stories about persons performing on a job. Given knowledge concerning the relevant dimensions for a particular job, vignettes depicting an employee performing at various levels on each of these dimensions can be generated. As a result, normative true scores may be developed for each ratee on each dimension. With the development of a number of these "paper people," it is then possible to evaluate the similarity or accuracy of a particular rater's actual ratings compared to these true scores. The flexibility afforded in generating ratees with different performance profiles is a major advantage of this approach. However, the main disadvantage to this method involves the lack of realism associated with "paper people."

A second possible solution to this problem involves the identification and use of some external criterion of performance. Unfortunately, the major pitfall to this approach is the inability to find external criteria that

correspond conceptually to the various performance dimensions of a job. Consequently, accuracy on performance dimensions typically included in performance appraisal instruments cannot easily be studied using this approach. When faced with this dilemma in a field setting, the Kavanagh et al. (1971) MTMR approach may be a viable alternative.

A final approach used by Borman and his colleagues (Borman, Hough & Dunnette, 1976), involves the development of videotaped vignettes of persons performing on the job. This approach combines the flexibility of the "paper people" approach with the notion that watching people performing on the job is more realistic. Still, it can be argued that the short duration of the performance episodes viewed, and the opportunity to record ratings immediately after viewing the ratee perform his or her job is less than what would be expected in a real life performance situation. In any event, normative true scores can then be generated, and a measure of accuracy derived by comparing actual ratings of these tapes (provided by supervisors/raters) to the true scores.

The paper people approach and the videotaped vignette approach have been used in the three studies cited (Borman, 1975; Borman, 1979a; Bernardin & Pence, 1980) in an attempt to assess the accuracy of performance ratings.

Rater training and accuracy. In a 1975 study, Borman asked first-time supervisors to evaluate written vignettes

describing ratees performing on the job, both before and after a five-minute training session on reducing halo errors in their ratings. As noted earlier, results showed that halo was significantly reduced after training. In addition to measuring extent of psychometric errors, Borman (1975) also computed an index of validity and interrater reliability in order to assess accuracy more directly. These results revealed that validity was unaffected by rater training. However, performance ratings completed after training possessed lower reliability. While the absence of a formal control group was a definite weakness in this study, it has been instrumental in focusing attention on the relationship between rater errors and rating accuracy.

Borman (1979a) investigated further this error-accuracy issue using a much more elaborate training program (the workshop approach of Latham et al., 1975). Borman provided training to half of the student subjects, and then asked them to rate two sets of videotaped ratees (a group of eight managers in a problem-solving session with a troublesome subordinate, and a group of eight interviewers). In addition to measuring halo effects, Borman (1979a) generated a measure of accuracy (differential accuracy) proposed by Cronbach (1955).

After receiving psychometric error-reduction training, halo errors were significantly reduced. However, training had no positive effect on the accuracy of the ratings, although the accuracy of the ratings did not drop

significantly after training. Borman (1979a) suggested it may be easier to teach persons to eliminate or reduce rating errors than to teach them to be more accurate, and, in fact, the relationship between these dependent variables may not be as clear as previously assumed. Accuracy may not be increased automatically when rating errors are reduced.

The most recent study that has examined this relationship between rating errors and accuracy was conducted by Bernardin and Pence (1980). These researchers provided student subjects with one of two types of rater training. One group received the typical type of rater error training (see Bernardin, 1978). The second group of subjects was lectured on the multidimensionality of most types of work performance, and the need to distinguish each dimension when evaluating performance. The importance of fair, unbiased and accurate ratings was also stressed. In addition, discussion centered around seeking consensus on stereotypes of effective and ineffective teacher performance.

A posttest-only design allowed a comparison of the two training groups and a no-training control group after evaluation of written vignettes depicting performance of two faculty members. Measures of halo and leniency effects were collected, as well as a measure of accuracy (difference between actual scores and true scores). Ratings from the psychometric error training group had significantly less leniency and halo error than ratings from the other two

groups. However, significantly less accuracy was also found for this group than for the control or "generalized training" group. Bernardin and Pence (1980) speculated that psychometric error training fosters a response set in raters that results in lower levels of leniency and halo, but lower levels of accuracy as well. In addition, they suggested the need for further research to develop rater training programs that increase rating accuracy rather than train rater response sets.

Since the publication of these three studies, several researchers have begun to address the relationship between rating errors and accuracy. Borman (1977) correlated differential accuracy scores for each rater with his or her halo, leniency and range restriction scores (using data from the Borman, 1979a study), and found very little correspondence between accuracy and psychometric errors. In addition, Murphy and Balzar (1981) evaluated the relationship between six rater error measures and four measures of rating accuracy across three laboratory studies. Once again, none of the error measures showed consistent correlations with any of the accuracy measures across all three studies.

More recent studies by these authors (Borman, 1979b; Murphy, Garcia, Kerkar, Martin & Balzar, 1982) attempted to evaluate more closely accuracy in judging performance. Borman (1979b) focused on valid predictors of accuracy, and concluded that certain individual difference variables are

related to accuracy across a variety of situations. Murphy and his colleagues (1982) focused on the relation between observational accuracy and performance rating accuracy, and concluded that the two are correlated to some extent. Raters who overestimated the frequency of favorable teacher behaviors also tended to give higher performance ratings.

Rater Training Programs

Three separate training programs were developed or adapted for use in this research. Each differed in their content and focus. The first training program was chosen as an imitation of the typical psychometric error training found in the literature. In addition, several researchers (Borman, 1979a; Landy & Farr, 1980) have proposed training to improve observational skills. A second rater training program reflects these suggestions. Finally, it is the belief of the present author that this new approach must be altered to include behaviors that occur subsequent to observation, namely, decision-making processes. Consequently, the third training program was developed to fill this void. In addition, a no-training control group was included in the experiment.

Training to reduce psychometric errors. This training approach reflects the rater error training typically found in the literature. A lecture/discussion format was used to introduce subjects to the meaning and prevention of four comon rater errors (halo, leniency, range restriction and

similarity). A videotaped lecture, approximately 20 minutes in length, introduced each of the four rating errors (For all three training groups, a videotaped lecture format was chosen to reduce experimenter bias, a problem typically ignored in the rater training literature). Definitions, graphic illustrations, examples and suggestions for preventing these errors were presented and then discussed at the end of the lecture. In addition, a discussion section followed (moderated by the present author). Discussion was initiated by the reading of two case studies, designed to demonstrate supervisors committing these errors in a work setting (see Appendix A for a transcript of the lecture, as well as the case studies).

Training to improve observational skills. This training approach is similar to the Thorton and Zorich (1980) observer training program. Once again, a videotaped lecture, approximately 20 minutes long, introduced subjects to the importance of being a good observer of behavior. Training included instructions to observe carefully, watch for specific behaviors, and take notes, as well as, a discussion of several systematic errors of observation (contamination from prior information, and over-reliance on a single source of information). The discussion session focused on two exercises. First, subjects were given performance dimensions relevant to the job of recruit interviewer (see Borman et al., 1976). These were discussed, and it was suggested that these dimensions be

used to help focus their observations of two videotaped recruit interviewers that were to follow. Taking notes was also suggested as an aid in the subsequent rating of these ratees. Subjects then viewed each ratee and rated them on the behaviorally anchored rating scales provided. A second exercise involved the reading of a case study illustrating errors of observation. Discussion followed each of these exercises, moderated by the present author. In general, then, this approach attempted to improve observation processes, such as detection, perception, and recall or recognition of specific behavioral events (see Appendix B for a transcript of the lecture, as well as the stimulus materials used in the discussion session).

Training to improve decision-making skills. Given the lack of success that has surrounded efforts to improve the accuracy of ratings, it was believed that the best hope for success in the future lies in the area of decision-making training. It seems apparent that attempts to effect accuracy by reducing or eliminating psychometric errors is questionable. However, a shift in focus toward the cognitive processes that occur prior to actual rating of employees may be fruitful in influencing the accuracy of those ratings. Data-driven and theory-driven inferences from diverse areas of behavioral research point to such a conclusion. For example, Cooper (1981), in a recent review of the "halo effect" suggested a study of the clinical training literature on diagnostic accuracy (i.e., Goldberg,

1968). In addition, initial research efforts by Thorton and Zorich (1980), have attempted, with some success, to improve observation processes of raters. While such a focus is an important step in the right direction, it neglects the processes that occur subsequent to observation of ratee behavior—namely, the interpretation and weighting of those behaviors. Research efforts from the selection interview domain have addressed this weighting phenomenon and its effect on the decision process. Springbett (1958), Bolster and Springbett (1961) and Holland (1972) have all concluded that negative and positive information is processed differently, with negative information being weighted much more heavily.

Finally research from the social-psychological, behavioral decision-making and cognitive domains have generated information supporting a shift in focus to rater decision-making training. The work of Kahneman and Tversky (i.e., Kahneman & Tversky, 1972, 1973, 1981; Tversky & Kahneman, 1971, 1974, 1978) on the use of simple heuristics, or of cognitive strategies have demonstrated both the important role these heuristics play in accurate judgments, and the judgmental inaccuracies that result when such strategies are misapplied. Hogarth (1981), in a recent review of judgmental heuristics has elaborated the conditions under which heuristics can be a valuable aid in the decision process. In addition, Nisbett and Ross (1980) have dealt extensively with the notion of strategies and

shortcomings of human inference, outlining possible approaches to improving human inference. Through a consideration of these related concepts and inferences from numerous, diverse areas of research, it seems necessary that training directed at improving accuracy confront the strategies (both formal and informal) used by raters to arrive at final rating decisions. This training approach is an attempt to deal with these issues and concerns.

A lecture/discussion format was used to introduce subjects to the idea of intuitive and formal decision-making strategies. A videotaped lecture, approximately 20 minutes in length, included a discussion of judgmental heuristics, as well as the costs and benefits of formal versus intuitive strategies. In addition, various inferential errors were illustrated, such as insensitivity to the perils of biased data, inappropriate causal inference, over-reliance on previously formed theories, and inappropriate weighting of observed behaviors. A discussion session (moderated by the author) followed the lecture, and consisted of two exercises. Exercise One involved the reading of two scenarios prior to viewing, and subsequently rating two videotaped recruit interviewers (see Borman et al., 1976). Each scenario presented information irrelevant to job performance (i.e., personality information, recent personal life events or crises). After rating the performance of each ratee, a discussion of scenario effect on ratings was initiated.

The second exercise involved viewing still-life scenes depicting people in a work setting. Subjects were asked to first generate a list of behaviors observed in the picture, and then list inferences drawn from these observations. A discussion followed, focusing on the differences between behaviors and inferences, and how inferences can be made inappropriately in a given situation (see Appendix C for a transcript of the videotaped lecture, as well as the stimulus materials used in the discussion sessions).

Purposes of the Study

Overall, then, while it has been established that psychometric error training reduces rating errors, questions remain about what effect this training has on accuracy. It appears that not enough is known about the potential usefulness of training to enhance rating accuracy. Results of studies reviewed above suggest, however, that improvements in accuracy using established rater training programs, may be more difficult to bring about than simply changes in rater behavior. This question is the focus of the current research. Thus, the questions of interest here are the following:

(1) How effective are different types of rater training at reducing psychometric errors, and improving the accuracy of ratings?

(2) Do effects of different types of rater training in the laboratory transfer to performance ratings on the job?

(3) To what extent do employees react differently to different types of rater training?

Method

This section describes the research design and procedures used in conducting the experiment. In addition, a discussion of the dependent variables chosen for use in this study is included.

Research Design and Procedure

Supervisory personnel working at Old Dominion University (i.e., Food Services, Financial Aid, Buildings and Grounds, Library Services, Personnel) were randomly assigned to one of three training groups, or the no-training control group. In all, 52 supervisors participated in the two-part workshop on performance appraisal. In addition, each group was subdivided into two subgroups, such that half of the members of each group met on one day, and the other half of the group met another day of the week. In all, then, eight training sessions were held during the first week of training, with subgroups randomly assigned to either a morning or afternoon session. Sessions for the three training groups lasted approximately three-and-one-half hours, while the no-training control group session lasted approximately two-and-one-half hours.

During Week Two of the workshop a similar procedure was followed, resulting in eight training sessions. Sessions for the three training groups lasted approximately two-and-one-half hours, while the control group sessions lasted approximately three-and-one-half hours (after data collection was completed, the control group received training as well). In all, then, the performance appraisal workshop for each of the four groups lasted approximately six hours over a two-week period (see Appendix D for the schedule of training sessions).

Procedure. During the initial training session, supervisors were given a general introduction to the purposes of the workshop, and then viewed five videotaped managerial vignettes one by one, making their ratings after each performance. Each of the training groups then received their specific rater training program. In addition, the three training groups were asked to fill out a short questionnaire concerning their reactions to the training program. Supervisors then returned the following week, and once again viewed and rated the five managerial tapes. Also, they were asked to evaluate several of their current employees, using the Commonwealth of Virginia's Performance Evaluation form. Subsequent to all data collection the Control Group was given rater training. This training included viewing the Psychometric Error Training videotaped lecture, plus two case studies. A discussion session followed (moderated by the present author), incorporating

comments and suggestions on how to be better decision-makers, and observers of behavior. Thus, the research employed a pretest/posttest design, with a control group. The data collection design and procedure is shown in Figure 1.

Use of Borman videotapes. As noted previously, few rater training studies have been designed to evaluate rating accuracy, due to the absence of performance dimension true scores. This problem was dealt with in the current study by using the Borman videotapes previously developed (Borman et al., 1976). Subsequent to development, Borman (1979a) revalidated the dimensional true scores by having a group of expert raters evaluate all taped ratees. Interrater agreement between experts ranged from .84 to .97, while correlations between these new expert ratings and intended true scores were also high (median r = .81). Consequently, the means of the new experts' ratings were adopted as the normative true scores. For purposes of this study, five of the managerial videotapes were used to collect performance ratings from supervisors attending the workshop. The seven dimensions and the normative true scores are shown in Table 1.

Use of actual job performance ratings. In addition to ratings gathered from use of the Borman videotapes, actual job performance ratings of university employees were collected from each supervisor subsequent to rater training. The Commonwealth of Virginia's Employment Performance Evaluation form is included in Appendix E.
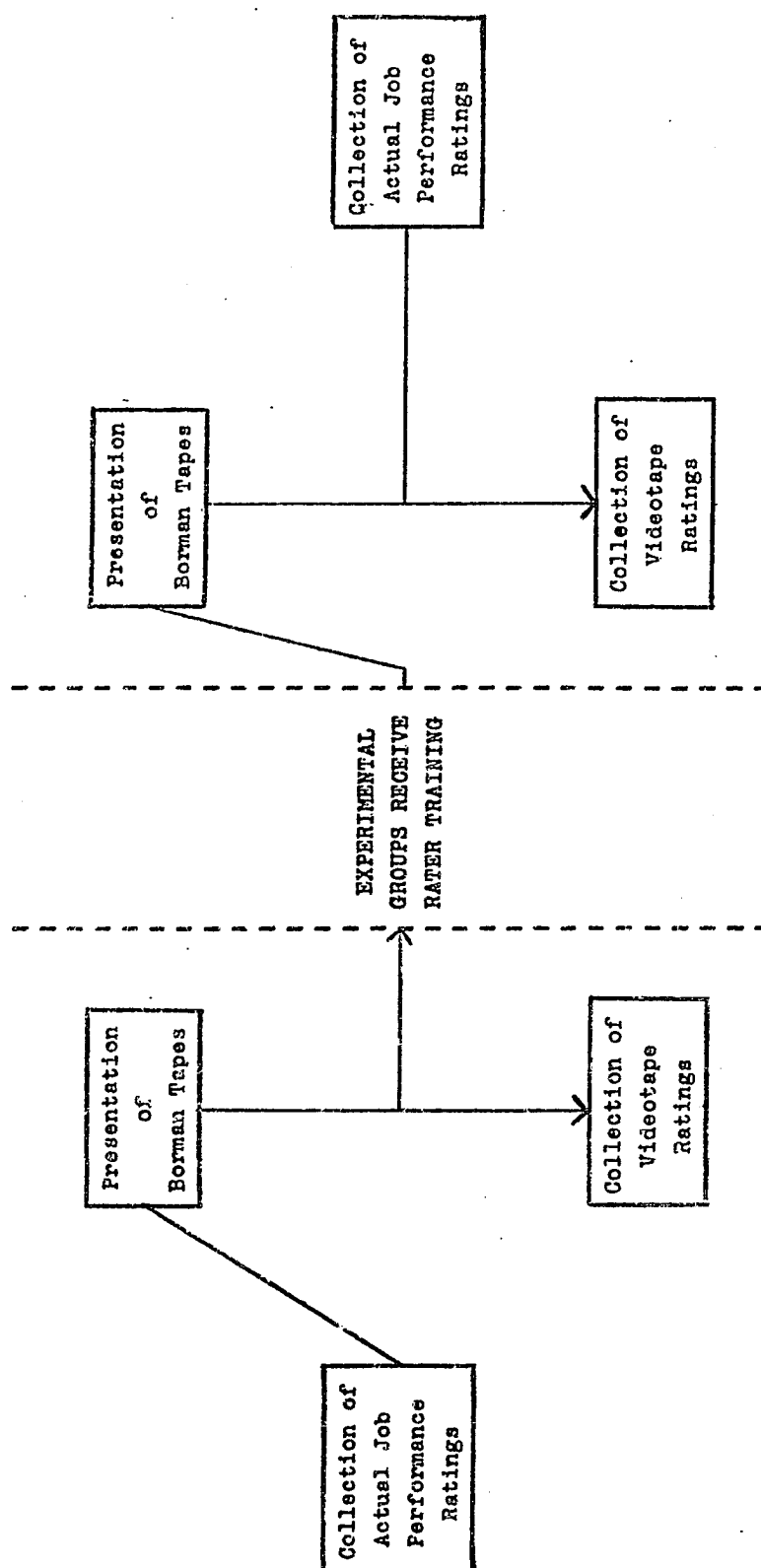
Figure 1: Research Design

Table 1

EXPERTS' RATINGS OF MANAGER PERFORMANCE

			RATEES			
PERFORMANCE FACTORS	1	2	3	4	5
A. STRUCTURING/CONTROLLING INTERVIEW	5.0	2.5	6.0	2.0	3.0
B. ESTABLISHING/MAINTAINING RAPPORT	2.5	5.5	4.5	6.0	4.0
C. REACTING TO STRESS	1.5	4.5	5.0	3.5	2.5
D. OBTAINING INFORMATION	3.5	3.5	6.0	2.5	2.0
E. RESOLVING CONFLICTS	1.5	2.0	6.0	5.0	5.0
F. DEVELOPING EMPLOYEES	2.5	3.5	3.5	3.5	5.5
G. MOTIVATING EMPLOYEES	2.0	5.0	5.0	2.5	6.0

Dependent Measures

Dependent measures assessed in the present study are grouped according to the focus of measurement.

Psychometric considerations. Measures of halo, leniency and range restriction were gathered for both pre-training and post-training videotape performance ratings. In addition, these same psychometric variables were measured on the actual job performance ratings. Both pre-training (the most recently completed Employment Performance Evaluations were supplied by the Personnel Department) and post-training evaluations were collected. For purposes of this study, halo was operationally defined as the variance across dimensions of the rater's ratings of a particular ratee. Leniency was operationally defined as a shift in the mean ratings from the midpoint of the scale in the favorable or higher rating direction. Restriction of range was designated as the standard deviation of the rating distribution, over ratees and dimensions.

Accuracy. Utilizing data collected from ratings of the videotaped ratees both before and after training, and compared to the normative true scores, accuracy data were derived. The measure of accuracy chosen for use in this study (per Cronbach, 1955) was differential accuracy (DA). The DA measure provided accuracy scores for each rater on each job dimension. The DA for a rater on a dimension was computed by correlating the rater's rating of the five videotaped managers on that dimension with the mean true

scores provided by the expert judges. The Fisher r to z transformation was then applied to each DA correlation. Each rater, therefore, had a total of seven pre-training accuracy scores, and seven post-training accuracy scores, corresponding to the seven managerial dimensions.

Trainee reaction measures. Three items on the Trainee Reaction Questionnaire administered to all employees after training (see Appendix F) were used to measure participants' reactions to (1) the overall training program, (2) the videotaped lecture portion of training, and (3) the discussion portion of the training program.

Results

Laboratory Data

Leniency, halo, range restriction and differential accuracy measures were calculated from performance ratings of the Borman managerial videotapes. The pretest/posttest design allowed a comparison of means across groups and time. Consequently, psychometric errors were analyzed using a Group (4) x Time (2) x Rater (12) Analysis of Variance (ANOVA), with raters nested within groups. Differential accuracy was measured by computing a Group (4) x Time (2) x Rater (12) Multivariate Analysis of Variance (MANOVA) (with raters nested within groups), with the seven dimensions designated as dependent measures. In addition, because the central focus of the research concerned identifying the changes in error rate and accuracy within groups across time, orthogonal comparisons were applied to all Group x Time interactions, regardless of statistical significance. In fact, wherever orthogonal comparisons are analyzed, the ANOVA interaction effect is of secondary concern. In addition, because of unequal levels of the Rater factor, four subjects were randomly excluded from the analyses (two subjects from the Psychometric Error Training Group, and one subject each from the Decision-Making and Control groups).

Leniency.   Leniency was operationally defined as a shift in the mean ratings from the midpoint of the scale in a higher rating direction.   Thus, mean ratings for all raters in the four groups, both before and after training were compared.   Results of the 4 x 2 x 12 ANOVA indicated significant Time and Time x Group leniency effects (see Table 2).   Orthogonal comparisons between each group's pre-training and post-training ratings were subsequently performed, and indicate a significant change in level of leniency for the Psychometric Error Training group ($p <$ .05).   In addition, this change was in the expected direction, with error-training causing a drop in rater leniency (Table 3).   However, no other groups showed a significant increase or decrease in level of leniency as a result of training.

Halo.   Halo was operationally defined as the variance across dimensions of the rater's ratings of a particular ratee.   Thus, the variance of the ratings across ratees and dimensions was also analyzed using the Group x Time x Rater ANOVA.   Significant Group and Group x Time effects resulted, as shown in Table 4.   Once again, orthogonal comparisons were used to test for significant time differences within each group.   Just as with leniency, significant halo differences were found in the group that received training to reduce halo.   This significant difference was also in the expected direction (see Table 3), with a drop in halo occurring after training.   A statistically significant

Table 2

Summary Table of Analysis of Variance on Leniency

Scores from Laboratory Ratings

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
GROUP	0.5744	3	0.1915	0.6777
RATER(G)	12.4307	44	0.2825	------
TIME	0.9429	1	0.9429	13.1182*
TR(G)	3.1626	44	0.0719	------
TG	1.1339	3	0.3780	5.2586*
Total	18.2445	95	------	------

$p < .05$

Table 3

Orthogonal Comparisons Between Pre-Training and

Post-Training Mean Scores for Each Group

Type of Training	Leniency	Halo	Range Restriction
Control (1)	0.0170	3.1256	4.7840*
Psychometric Error (2)	26.1269*	17.5971*	3.1066*
Observation (3)	0.6828	4.8276*	6.9015*
Decision-Making (4)	2.0600	0.0219	0.7649
MS error	0.0719	0.1447	0.7457
Direction of Significant F Values**	21 > 22	21 > 22 31 < 32	11 < 12 31 < 32

**These values reflect significant changes from Pre-Training to Post-Training for each group. For example, the leniency score for Group 2 (psychometric error training) prior to training was significantly larger than leniency after training.

* $p < .05$

Table 4

Summary Table of Analysis of Variance on Halo

Scores from Laboratory Ratings

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
GROUP	10.9113	3	3.6371	10.2466*
RATER(G)	15.6181	44	0.3550	------
TIME	0.0052	1	0.0052	0.0357
TR(G)	6.3883	44	0.1447	------
TG	3.6966	3	1.2322	8.5135*
Total	36.6195	95	------	------

* $p < .05$

difference (p < .Ø5) between pre- and post-training means was also discovered for the Observation Training group. This change, however, was in the opposite direction, with degree of halo in the performance ratings increasing after

Observation Training. The Control group and the Decision-Making Training group once again showed no significant change in level of halo.

Range Restriction. Range restriction was operationally defined as the standard deviation of the rating distribution over ratees and dimensions. Thus, a comparison of range restriction levels was once again accomplished by using a 4 x 2 x 12 ANOVA. A summary of the Analysis of Variance of these data are presented in Table 5. These results show a statistically significant Time effect (p < .Ø5). In addition, orthogonal comparisons within groups across time found significant differences between levels of range restriction for the Control group and the Observation Training group. These results indicate increased range restriction for these two groups subsequent to training. Levels of range restriction did not change significantly for either the Psychometric Error Training group or the Decision-Making group (see Table 3).

Leniency, halo and range restriction means collected before and after training from each of the four groups are presented in Table 6. A summary of the psychometric findings suggests, then, that training to reduce psychometric errors did, in fact, cause a significant

Table 5

Summary Table of Analysis of Variance on Range Restriction

Scores from Laboratory Ratings

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
GROUP	0.5699	3	0.1900	0.1738
RATER(G)	48.0853	44	1.0929	------
TIME	10.3549	1	10.3549	13.8867*
TR(G)	32.8093	44	0.7457	------
TG	1.2504	3	0.4168	0.5590
Total	93.0698	95	------	------

* p < .05

Table 6

Means of Leniency, Halo and Range Restriction

Scores for Field Ratings

		Leniency	Halo	Range Restriction
Control Group	Time 1	4.0405	1.2262	3.0964
	Time 2	4.0548	0.9516	2.3250*
Psychometric Error Group	Time 1	4.4142	1.4857	2.9143
	Time 2	3.8548*	2.1373	2.2929
Observation Training Group	Time 1	4.3071	1.0936	2.9560
	Time 2	4.2167	0.7525	2.0298*
Decision -Making Group	Time 1	4.2571	1.1580	2.7607
	Time 2	4.1000	1.1809	2.4524

* denotes significant mean changes from Time 1 to Time 2

reduction in halo and leniency. However, Observation Training caused an increase in halo errors and range restriction errors, while Decision-Making Training appeared to have no significant effect on psychometric error rate.

Accuracy. Differential accuracy scores for each rater on each dimension were compared using a Group x Time x Rater MANOVA, with the seven dimensions as multiple dependent measures. Results of the 4 x 2 x 12 MANOVA showed a statistically significant Group x Time effect (p < .05). Table 7 presents the Summary MANOVA Table. Subsequently, seven separate univariate analyses were computed for each of the dependent measures on this factor. As shown in Table 8, three of the seven ANOVAs were also statistically significant at the .05 level (Dimensions 4, 5, & 7). In addition, given the rationale presented earlier, orthogonal comparisons were performed on all seven dimensions for each group, comparing pre-training and post-training DA scores. Table 9 presents the results of these planned comparisons. Orthogonal comparisons on Dimensions One, Three and Six resulted in no significant differences in accuracy for the four groups across time. Dimension Two, however, shows a significant change in accuracy for the Decision-Making Training group. In addition, this change in accuracy was in a positive direction, increasing significantly after training. No other group showed a significant training effect on this dimension. Thus, the results indicate that training aimed at improving a rater's decision-making skills also improves the accuracy of the ratings on Dimension Two.

Table 7

Summary Table of Multivariate Analysis of Variance

on Accuracy Scores from Laboratory Ratings

Source of Variation	Degrees of Freedom	Approximate F-Value
Group	21, 110	1.0524
Time	7, 38	1.3059
G x T	21, 110	2.0846*

* p < .05

Table 8

Summary Table of Univariate F-Tests on Accuracy

Scores from Laboratory Ratings for Group x Time Effect

Variable	Sum of Squares	Mean Squares	$\underline{F}$ Ratio**
Dimension 1	0.2800	0.0933	0.3868
Dimension 2	1.0147	0.3382	2.3325
Dimension 3	0.6865	0.2288	1.2075
Dimension 4	5.1840	1.7280	5.1896*
Dimension 5	1.9080	0.6360	3.7583*
Dimension 6	3.3692	1.1231	2.5903
Dimension 7	2.3285	0.7762	2.9951*

** df = 3, 44 in each case   * $\underline{p}$ < .05

Table 9

Orthogonal Comparisons Between Pre-Training and Post-Training

Mean Scores for Each Dimension and Group

Type of Training	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6	Dimension 7
Control (1)	0.0186	1.4884	0.0168	0.4164	4.7365*	2.2489	1.8214
Psychometric (2)	0.0768	0.0204	1.2747	37.4126*	2.0773	3.4041	10.0055*
Observation (3)	1.4257	0.8195	0.0221	8.4442*	5.0689*	0.2866	0.6877
Decision-Making (4)	0.0972	5.9481*	2.3447	0.0011	0.1075	2.2469	1.0884
MS error	0.2413	0.1450	0.1895	0.3330	0.1692	0.4336	0.2592
Direction of Significant $\underline{F}$-Values**		41 < 42		21 > 22 31 < 32	11 > 12 31 < 32		21 > 22

* $\underline{p}$ < .05

** denotes significant change from pre-training to post-training

As shown in Table 9, significant differences were also uncovered for the remaining three dimensions (Dimensions 4, 5, & 7). For both Dimension Four and Dimension Seven, significant decline in accuracy was found in subjects exposed to the psychometric error training. In addition, the Observation Training group changed significantly in Dimensions Four and Five. The direction of change for the Observation Training group on these two dimensions was opposite that of the scores for the Psychometric Error Training group; namely, an improvement in the accuracy of the ratings occurred after training. Finally, the Control group was found to have decreased in accuracy at Time Two (after training) on Dimension Five.

Overall, significant changes in accuracy were found on four of the seven performance dimensions. The Psychometric Error Training group became significantly less accurate on two of the seven dimensions, the Observation Training group became more accurate on two of the dimensions, the Decision-Making group was significantly more accurate on one dimension, and the Control group was found to be less accurate on one dimension. The mean DA scores for all groups both before and after training are listed in Appendix G.

Additional insight into the effects of the different training approaches on psychometric errors and accuracy can be gained by looking at a summary table of <u>trends</u> depicting mean changes across training. While these changes do not

represent statistically significant differences, they do suggest direction of change for each group. Table 10 illustrates these changes for all laboratory dependent measures. A "+" signifies an improvement in scores on that dependent variable subsequent to training. A "-" denotes a decrement in performance, and an "=" signifies no noticeable change in pre-to-post training performance ratings (For the accuracy, halo, and range restriction measures a .03 fluctuation was considered a change, while for the leniency measure, a .30 fluctuation was considered a change).

As is apparent, while psychometric error training reduced halo and leniency, it also tended to reduce accuracy on five of the seven performance dimensions. However, for the other two training groups, psychometric errors either persisted or increased, but more importantly, accuracy improved. For the Observation Training group, an improvement in accuracy is noted on four of the seven dimensions, while the Decision-Making Training group improved on five of the seven dimensions. Thus it appears from this analysis of trends, that while psychometric error training does indeed reduce rating errors, it has a negative impact on accuracy. Observation and decision-making training, on the other hand, appear to have a positive effect on accuracy.

Field Data (Actual Performance Evaluations)

Leniency, halo and range restriction measures were calculated from actual performance ratings. After the

Table 10

Trends of Means From Time One to Time Two

PSYCHOMETRIC ERRORS / ACCURACY

	CONTROL	PSYCHOMETRIC	OBSERVATION	DECISION-MAKING
LENIENCY	=	+	=	=
HALO	−	+	−	=
RANGE RESTRICTION	−	−	−	−
DIMENSION 1	=	−	+	+
DIMENSION 2	−	=	+	+
DIMENSION 3	=	−	=	+
DIMENSION 4	+	−	+	=
DIMENSION 5	−	−	+	−
DIMENSION 6	−	+	−	+
DIMENSION 7	−	−	=	+

"+" improvement   "−" decrement   "=" stays same

Note: These signs indicate direction of change, not statistical significance.

completion of training, supervisors attending the workshop were asked to fill out evaluations for the employees they evaluated on the job. Employees rated were then identified, and the most recent pre-training evaluation was supplied by the Personnel Department. However, because some employees had never been rated previously, some supervisors in the workshop were new and had never rated before, and a supervisor needed to evaluate more than one employee for psychometric measures to be computed, a significant number of raters had to be dropped from the pre-training/post-training analyses (five subjects remained in the Control and Decision-Making Groups, six subjects in the Observation Training Group, and eight subjects in the Psychometric Error Training Group). Once again, unequal levels of the rater factor required additional elimination of subjects. As a result, only five raters per group were used in analyses of the field data.

Leniency. A Group x Time x Rater ANOVA (with raters nested in groups) was used to evaluate mean ratings. As illustrated in Table 11, no significant main or interaction effects were found. Applying the same rationale used in analyzing the laboratory data, orthogonal comparisons were computed to test for pre-training-to-post-training differences within each group. Table 12 presents these comparisons. Both the Psychometric Error Training group and the Observation Training group showed significant changes in leniency after training. However, the change for the two

Table 11

Summary Table of Analysis of Variance on Leniency

Scores from Field Ratings

Source of Variation	Sum of Squares	df	Mean Squares	$\underline{F}$ Ratio
GROUP	0.5419	3	0.1806	1.4992
RATER(G)	1.9277	16	0.1205	------
TIME	0.0043	1	0.0043	0.3135
TR(G)	0.2212	16	0.0138	------
TG	0.1151	3	0.0384	2.7755
Total	2.8102	39	------	------

* $\underline{p}$ < .05

Table 12

Orthogonal Comparisons Between Pre-Training and

Post-Training Mean Scores for Each Group

Type of Training	Leniency	Halo	Range Restriction
Control (1)	0.1287	0.5486	0.5388
Psychometric (2)	6.7599*	0.0805	4.9105*
Observation (3)	5.2242*	0.0851	0.5529
Decision-Making (4)	0.0723	0.4920	44.3992*
MS error	0.0138	0.0045	0.0031
Direction of Significant F-Values**	21 > 22 31 < 32		21 > 22 41 < 42

* $p < .05$

** denotes significant changes from Pre-Training to Post-Training for each group

groups differed in direction. While the Psychometric Error Training group reduced significantly their leniency errors, the group of supervisors trained to improve observational skills became more lenient after training. There were no significant changes in level of leniency for either the Control group or the Decision-Making group.

Halo. A 4 x 2 x 5 ANOVA was computed using the variances of ratings across ratees and dimensions. A significant Group effect was found in these halo data (see Table 13), but using orthogonal comparisons, no statistically significant difference were obtained within groups across training (see Table 12). Thus, training had no impact on level of halo for the four groups.

Range restriction. Finally, the same ANOVA procedure was used to test for differences in restriction of range. As noted in Table 14, no statistically significant effects resulted from the 4 x 2 x 5 ANOVA. Orthogonal comparisons, however, uncovered two significant changes in level of range restriction. First, the Psychometric Error Training group demonstrated a significant decrease in amount of range restriction (evidenced by an increase in the standard deviation of the performance ratings). In addition, the group of supervisors who underwent decision-making training showed an increase in restriction of range (see Table 12). Thus, it appears that psychometric error training was beneficial in reducing leniency and range restriction, while observation training caused an increase in leniency, and

Table 13

Summary Table of Analysis of Variance on

Halo Scores from Field Ratings

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
GROUP	0.0440	3	0.0147	3.2827
RATER(G)	0.0715	16	0.0045	------
TIME	0.0004	1	0.0004	0.1218
TR(G)	0.0513	16	0.0032	------
TG	0.0052	3	0.0017	0.5353
Total	0.1724	39	------	------

* $p < .05$

Table 14

Summary Table of Analysis of Variance on Range

Restriction Scores from Field Ratings

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
GROUP	0.0133	3	0.0044	0.9327
RATER(G)	0.0762	16	0.0048	------
TIME	0.0002	1	0.0002	0.0686
TR(G)	0.0492	16	0.0031	------
TG	0.0269	3	0.0090	2.9167
Total	0.1658	39	------	------

* $\underline{p} < .05$

decision-making training caused an increase in range restriction. A table of means for these three psychometric errors appears in Table 15.

Trainee Reaction Measures

Supervisors' reactions to the four training programs (the Control group completed a questionnaire following their post-data collection training) were evaluated on the basis of three trainee reaction measures (items 1, 2, & 3 of the questionnaire that appears in Appendix F).

Trainee reactions to whether they benefitted from the training program were evaluated by means of a One-Way Analysis of Variance. As shown in Table 16, there were no significant differences in the perception of overall training worth across the four groups. The second ratee reaction measure concerned the worth of the videotaped lecture. Once again, a One-Way ANOVA resulted in no significant group differences (see Table 17). Finally, reactions to the practice/discussion section of training were evaluated using a One-Way ANOVA, but no significant Group differences were found (see Table 18). In summary, then, no single training program was perceived by the trainees as more beneficial. Means and standard deviations on these three trainee reaction measures are shown in Table 19.

Table 15

Means of Leniency, Halo and Range Restriction

Scores for Field Ratings

		Leniency	Halo	Range Restriction
Control Group	Time 1	3.4333	0.1313	0.0833
	Time 2	3.4600	0.1627	0.1089
Psychometric Error Group	Time 1	3.3867	0.2053	0.0856
	Time 2	3.1933*	0.2173	0.1633*
Observation Training Group	Time 1	3.17333	0.1753	0.1467
	Time 2	3.3433*	0.1877	0.0216
Decision-Making Group	Time 1	3.5500	0.2500	0.1750
	Time 2	3.5700	0.2203	0.1161*

* denotes significant change in mean from Pre-Training to Post-Training

Table 16

Summary Table of Analysis of Variance for Each

Group's Trainee Reaction (Item One)

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
Between Groups	5.2715	3	1.7572	0.8795
Within Groups	97.8984	49	1.9979	------
Total	103.1699	52	------	------

$\underline{p} < .05$

Table 17

Summary Table of Analysis of Variance for Each

Group's Trainee Reaction (Item 2)

Source of Variation	Sum of Squares	df	Mean Squares	F Ratio
Between Groups	8.5244	3	2.8415	1.5022
Within Groups	92.6832	49	1.8915	------
Total	101.2076	52	------	------

$p < .05$

Table 18

Summary Table of Analysis of Variance for Each

Group's Trainee Reaction (Item Three)

Source of Variation	Sum of Squares	df	Mean Squares	$\underline{F}$ Ratio
Between Groups	7.7999	3	2.6000	1.5050
Within Groups	84.6529	49	1.7276	-------
Total	92.4528	52	------	-------

$\underline{p} < .05$

Table 19

Table of Means and Standard Deviations for

Trainee Reaction Measures

		Group 1	Group 2	Group 3	Group 4
Item 1	Mean	2.7143	3.1428	2.2500	2.6154
	S.D.	1.3260	1.4601	1.6026	1.2609
Item 2	Mean	2.7143	2.9286	1.8333	2.5385
	S.D.	1.3828	1.4917	1.4035	1.9829
Item 3	Mean	2.8571	2.8571	1.9167	2.7692
	S.D.	1.4064	1.2315	1.3114	1.3009

Discussion and Conclusions

The present study was designed to investigate differential effects of three training programs on psychometric errors and accuracy in performance ratings. As noted in the previous section, psychometric error training significantly reduced levels of leniency and halo, thus supporting earlier research in this area (i.e., Levine & Butler, 1952; Borman, 1975; Bernardin, 1978). From such results, researchers in the past have inferred that the accuracy of the ratings would be increased. Instead, it appears that differential accuracy scores decrease after rater error training. These findings lend support to the Bernardin and Pence (1980) study, and suggest that such training has an adverse impact on rating accuracy. Perhaps training subjects to be aware of certain rating errors oversensitizes them to where on the scale they are rating, rather than how accurately they are rating.

When viewed from a measurement theory approach, these findings suggest an alternative interpretation. As noted previously, rater training researchers classify rating errors as error variance, and consequently, assume training to reduce these errors will increase accuracy. However, if these rating errors contain both error variance and true variance, training that reduces these errors would not only

reduce error variance, but would affect true variance as well. When viewed from this perspective, rater error training could, in fact, reduce the accuracy of the ratings, rather than improving accuracy as previously assumed.

Results from this study also support the recent recommendations by Borman (1979a) and Bernardin and Buckley (1981) concerning observation training, and the findings of Thorton and Zorich (1980). Supervisors receiving training to improve observation skills showed overall improvement in the accuracy of their ratings, but this was not reflected in the psychometric error measures. A closer look at the content and focus of such training is an important area for future research. For example, Borman (1979a) has recommended that more emphasis be placed on training individuals to observe performance-related behaviors, and to agree on and to learn correct performance standards.

Consonant with this line of thinking, Bernardin and Buckley (1981) have suggested a formal, standardized diary-keeping system as an aid in increasing observational skills. In addition, they have recommended the use of frame-of-reference training whereby raters with idiosyncratic work standards could be identified, and attempts made to bring their perceptions into closer congruence with the rest of the organization.

In relation to the Decision-Making Training Program developed for this study, results suggest such an approach may be a fruitful avenue for further investigation. While

supervisors trained using the decision-making approach did not show reduced levels of psychometric errors, an improvement in rating accuracy was evident. Much additional work is needed to clarify and improve decision-making training in the area of performance appraisal. Much can be learned from existing cognitive and decision-making literature. It is the belief of the present author that training to improve observational skills deals only with the early stages of the Decision-Making Model (namely, the reception and storage stages), and that increased emphasis on the recall and/or response selection stages of decision-making are required. In a recent review article, Feldman (1981) focused on cognitive processes in performance appraisal and suggested further research efforts in this direction.

In general, both the Observation and Decision-Making Approaches to training offer some hope for improving the accuracy of performance ratings. Both of these training approaches view performance appraisal as a dynamic process that occurs throughout the year. The typical psychometric error training takes a much more static approach to appraisal. Consequently, the focus remains on an awareness of what errors "look like" when the evaluation is filled out. When these three training approaches are viewed in this manner, it is not illogical to propose that a longitudinal design may even uncover more dramatic changes. Thus, while psychometric errors may return to former levels

as time since training increases (Ivancevich, 1979), the accumulation of additional data points may increase the power and effect of training to improve observation and decision-making skills, by increasing the accuracy of decisions about performance.

Results of field data analyses suggest a closer look at transfer of training. While significant reduction in leniency and range restriction levels were found, sample size (n = 5) limits the significance of the results. In addition, raters were aware that post-training evaluations would impact in no way on merit increases, thus possibly confounding the results. Still, the direction of psychometric error changes for the field data were consistent with those in the laboratory data (namely, the Psychometric Error Training group showed decreases in errors, while the other groups either increased or stayed the same).

Finally, the results of the current study suggest several other avenues for future research. Because of an increased concern for assessing the accuracy of performance ratings, and the relation between psychometric errors and accuracy, additional research needs to focus on identifying predictors of rating accuracy. Borman (1977; 1979b; 1980) and Murphy and his colleagues (Murphy & Balzar, 1981; Murphy, et al., 1982), as noted earlier, have begun to address some of these issues.

Also, additional research needs to evaluate the relation between training, accuracy and dimensions of performance. Borman (1979a), found a significant dimension effect when analyzing differential accuracy scores using a Format x Dimension x Training ANOVA. He suggested that certain kinds of dimensions may be inherently more difficult than others for evaluating others accurately. Given the results of the present study, a Dimension x Type of Training Interaction is suggested. As shown in Table 10, the accuracy of performance ratings varied widely across the four groups. Consequently, different types of rater training may improve or limit one's ability to make accurate judgments about different aspects of job performance.

In general, the current research makes a fairly strong indictment against the traditional psychometric error approach to rater training, at least when viewed from the perspective of performance rating accuracy. In addition, further research is necessary to determine whether observation (and decision-making training are viable new approaches. Research may, in fact, determine that none of these approaches are worthwhile from a management perspective. The possibility also exists that different approaches to rater training may be more effective at different supervisory levels, or in some combination. Finally, while psychometric error training may not prove useful as a means for improving rating accuracy, its worth may lie in how such training affects rater attitudes.

Programs of rater training must meet an acceptability criterion if they are to be deemed useful. While no differences were found in trainee reactions to the three training approaches used in this study, this variable must be considered before abandoning a particular approach to rater training.

References

Bernardin, H.J. Effects of rater training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 1978, 63, 301-308.

Bernardin, H.J. & Buckley, M.R. Strategies in rater training. Academy of Management Review, 1981, 6, 205-212.

Bernardin, H.J. & Pence, E.C. The effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 1980, 65, 60-66.

Bernardin, H.J. & Walter, C.S. Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 1977, 62, 64-69.

Bingham, W.V. Halo, invalid and valid. Journal of Applied Psychology, 1939, 23, 221-228.

Bolster, B.F. & Springbett, B.M. The reaction of interviewers to favorable and unfavorable information. Journal of Applied Psychology, 1961, 45, 97-103.

Borman, W.C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 1975, 60, 556-560.

Borman, W.C. Consistency of rating accuracy and rating errors in the judgment of human performance. Organizational Behavior and Human Performance, 1977, 20, 258-272.

Borman, W.C. Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 1979(a), 64, 410-421.

Borman, W.C. Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 1979(b), 3, 103-115.

Borman, W.C. Performance judgments: The quest for accuracy in ratings of performance effectiveness. Paper presented at First Annual Scientist-Practitioner Conference. Old Dominion University, Norfolk, Virginia, 1980.

Borman, W.C., Hough, L.M. & Dunnette, M.D. Performance ratings: An investigation of reliability, accuracy, and relationships between individual differences and rater error. Minneapolis: Personnel Decisions, Inc., 1976.

Brown, E.M. Influence of training, method and relationship on halo effect. Journal of Applied Psychology, 1968, 52, 195-199.

Campbell, J.P. Psychometric theory. In M.D. Dunnette (Ed.), Handbook of Industrial and Organizational Psychology, Chicago: Rand-McNally, 1976.

Cooper, W.H. Ubiquitous halo. Psychological Bulletin, 1981, 90, 218-244.

Cronbach, L.J. Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, 1955, 52, 177-193.

Faye, C.H. & Latham, G.P. Effects of training and rating scales on rating errors. Personnel Psychology, 1982, 35, 105-116.

Feldman, J.M. Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 1981, 66, 127-148.

Goldberg, L.R. Man versus model of man: A rationale, plus some evidence, for a method of improving clinical inferences. Psychological Bulletin, 1970, 73, 422-432.

Guion, R.M. Personnel testing. New York: McGraw-Hill, 1965.

Hogarth, R.M. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, 1981, 90, 197-217.

Ivancevich, J.M. A longitudinal study of the effects of rater training on psychometric errors in ratings. Journal of Applied Psychology, 1979, 64, 502-508.

Kahneman, D. & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.

Kahneman, D. & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 251-273.

Kahneman, D. & Tversky, A. Intuitive prediction: Biases and corrective procedures. In S. Makridakis & S.C. Wheelright (Ed.), Studies in the management sciences: Forecasting. New York: John Wiley, 1981.

Kavanagh, M.J., MacKinney, A. & Wolins, L. Issues in managerial performance: Multitrait-multimethod analysis of variance. Psychological Bulletin, 1971, 75, 34-49.

Landy, F.J. & Farr, J.L. Performance rating. Psychological Bulletin, 1980, 87, 72-107.

Latham, G.P. & Wexley, K.N. Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley, 1981.

Latham, G.P., Wexley, K.N. & Pursell, E.D. Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 60, 550-555.

Levine, J. & Butler, J. Lecture vs. group discussion in changing behavior. Journal of Applied Psychology, 1952, 36, 29-33.

Murphy, K. & Balzar, W. Rater errors and rating accuracy. Paper presented at the Annual Meeting of the American Psychological Association, Los Angeles, August, 1981.

Murphy, K., Garcia, M., Kerkar, S., Martin, C., & Balzar, W. Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 1982, 67, 320-325.

Nisbett, R. & Ross, L. Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall, 1980.

Sauser, W.I. & Pond, S.B. Effects of rater training and participation on cognitive complexity: An exploration of Schneier's cognitive reinterpretation. Personnel Psychology, 1981, 34, 563-577.

Springbett, B.M. Factors affecting the final decision in the employment interview. Canadian Journal of Psychology, 1958, 12, 13-22.

Stockford, L. & Bissell, H.W. Factors involved in establishing a management-rating scale. Personnel, 1949, 26, 94-116.

Taylor, E.K. & Hastman, R. Relation of format and administration to the characteristics of graphing rating scales. Personnel Psychology, 1956, 9, 181-206.

Thorton, G.C. & Zorich, S. Training to improve observer accuracy. Journal of Applied Psychology, 1980, 65, 351-354.

Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Tversky, A. & Kahneman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.

Tversky, A. & Kahneman, D. Causal schemata in judgment under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillsdale, NJ: Lawrence Erlbaum, 1978.

Vance, R.J., Kuhnert, K.W. & Farr, J.L.   Interview
     judgments: Using external criteria to compare
     behavioral and graphic scale ratings. Organizational
     Behavior and Human Performance, 1978, 22, 279-294.

Warmke, D. & Billings, R.S.   Comparison of training methods
     for improving the psychometric quality of experimental
     and administrative performance ratings.   Journal of
     Applied Psychology, 1979, 64, 124-131.

APPENDIX A

Psychometric Error Training Lecture and Discussion Materials

PSYCHOMETRIC ERROR TRAINING

Training Appraisers

Evaluating employee performance is an important and necessary part of any supervisor's job. Regardless of whether your organization employs a formal system of employee evaluation, judgments about how individual employees are performing are made almost daily. People are constantly making judgments about others. Unfortunately, many of these informal judgments may be erroneous.

Consequently, a formal system of performance evaluation is usually adopted to help reduce the possibilities of bias and uninformed judgments; to standardize the types of information that will be forthcoming; and to ensure that the resulting appraisal information that will be forthcoming; and to ensure that the resulting appraisal information is gathered in a form that permits its use across the entire organization.

While a formal system of appraisal helps to standardize this process in the organization, it in no way guarantees consistent, accurate evaluation. Therefore, the purpose of today's talk will focus on helping you be more accurate in your judgment of employee performance. We will deal with this problem by focusing on some of the most common types of errors supervisors make in evaluating their employees. By becoming familiar with these errors, and discussing ways to avoid them, we hope to improve the accuracy of performance evaluations in general.

Insert overhead about here

Psychometric Errors

Halo Effect

Probably the most common rater error encountered is known as the halo effect. The halo effect refers to inappropriate generalization from one aspect of a person's performance to all aspects of a person's job performance. By attending to a global impression of each ratee (employee) rather than carefully distinguishing between performance factors, the supervisor commits a halo error.

For example, a person who is quite outstanding in one area of the job (e.g., job knowledge/skills) may be rated inaccurately as outstanding in all areas of the job (e.g., quality of work, productivity, initiative, dependability, etc.). Conversely, if a person is rated as deficient in one area of the job, that person may be rated incorrectly as doing poorly on all aspects of the job.

As you can see, ratings plagued by this error often do not provide an accurate portrayal of an individual's performance on different factors.

Insert overhead here

This does not necessarily mean that certain individuals cannot be superior on all performance factors, only that certain strengths or weaknesses can sometimes influence your ratings. Remember, people have both strengths and weaknesses, and each needs to be evaluated separately. Don't let one strength or weakness influence your ratings of all performance factors.

Leniency/Severity

A second type of error often committed by those persons who must evaluate their employees is known as a leniency error, and its converse--severity error. These types of error reflect a tendency by a supervisor to be either too easy or too hard in rating all their employees.

Insert overhead about here

For example, a supervisor may rate all his/her people at one end of the scale, or the other. The problem with doing this is that in the performance evaluation process, leniency may raise unwarranted expectations of the employee for raises, promotions or challenging job assignments.

On the other hand, with severity, the employee may get tired of banging his/her head against the wall, because no matter how hard the individual tries, the supervisor cannot be satisfied. Thus, it is the rater who is either too harsh, or too lenient on subordinates. The harsh rater tends to give evaluations that are lower than what they

should be, while the easy supervisor tends to give ratings that are higher or better than they should be--better than their performance warrants.

Restriction of Range

A third rating error which in some ways is similar to the leniency/severity error is known as restriction of range. When a supervisor rates all his/her employees harshly, or all of them leniently, or all of them about average, it's difficult to distinguish between employees. Thus, restriction of range refers to rating all your employees at about the same level. Range restriction errors are committed by supervisors who want to play it safe. Consequently, the obtained ratings will not allow the supervisor or the organization to differentiate between employees according to levels of performance.

Insert overhead about here

As you can see, just as with leniency/severity, restriction of range occurs when the rater does not use the whole scale when rating. By not singling out certain individuals as exemplary or overly deficient in certain areas, the supervisor avoids (at least temporarily) having to deal with what he feels will be unhappy or jealous employees.

Similarity

A fourth and final rating error that you need to be aware of is what's known as the "similar-to-me" effect, or similarity error. This type of error involves a tendency on the part of raters to judge more favorably those people whom they perceive as similar to themselves. The more closely an employee resembles the rater in attitudes or background, the stronger the tendency of the rater to judge that individual favorably.

Why might this error occur? We all tend to like, and think more highly of others whom we perceive as like us, rather than unlike us because it is flattering and reinforcing. While it's true that in social situations we tend to associate with, and like those who are "similar-to-me," when we let these sorts of impressions influence our evaluations, we hurt the accuracy of our evaluations.

Summary

In summary, error can be involved anytime we attempt to evaluate other people. For this reason, it is important that we be as objective as possible when we rate our subordinates. An awareness of the most common rating errors is an important step toward increasing the accuracy of our

evaluations. We discussed four of the most common rating errors:

Insert overhead about here

---

1) <u>Halo error</u> -- which comes into play when a rater feels that a particular performance factor is extremely important. Ratings are then assigned on the other factors that are consistent with the rating on the most important factor.

2) <u>Leniency/Severity error</u> -- results when the rater gives ratings that are unusually harsh or unusually easy to all his/her subordinates.

3) <u>Range Restriction error</u> -- comes into play when supervisors play it safe, and rate all their employees at a fairly even level, without even using the full range of scale values available.

4) <u>Similarity errors</u> -- result when we let similar attitudes and background of our employees influence our ratings of those employees. Thus, the important thing to keep in mind is that we need to evaluate each performance factor separately and make sure to concentrate on actual job-related behaviors.

COMMON RATING ERRORS

1) <u>Halo error</u> -- which comes into play when a rater feels that a particular performance factor is extremely important. Ratings are then assigned on the other factors that are consistent with the rating on the most important factor.

2) <u>Leniency/Severity error</u> -- results when the rater gives ratings that are unusually harsh or unusually easy to all his/her subordinates.

3) <u>Range Restriction error</u> -- comes into play when supervisors play it safe, and rate all their employees at a fairly even level, without even using the full range of scale values available.

4) <u>Similarity errors</u> -- result when we let similar attitudes and background of our employees influence our ratings of those employees. Thus, the important thing to keep in mind is that we need to evaluate each performance factor separately and make sure to concentrate on actual job-related behaviors.

CASE ONE

The Case of Ambition Exceeding Ability

In 1961, John Senn was hired by the University as a bookkeeper trainee. Prior to this job, he had worked for a short time on a weekly newspaper, but he had been replaced by a man who could sell as well as write. He also had a brief job as an apprentice sign painter, but had quit due to lack of interest. From 1961 to 1967, he had shown little promise of success in his job, advancing only one step from bookkeeper to clerk. In addition, during this time John and his supervisor had several bitter arguments, the result of personality clashes.

In 1967, a Public Relations Department was organized at the University. A woman was brought in from the outside to initiate and develop this activity. She had a considerable amount of experience in the public relations field, having worked for a large newspaper in that capacity. In addition, she had been a business writer for the Wall Street Journal, and more recently headed her own advertising agency.

John Senn had asked for a transfer to this new department because of his earlier experience in writing for the weekly newspaper. Based on information gathered from John's supervisor, the new head of Public Relations was reluctant to approve the transfer, but was persuaded to do so by her boss. John was assigned to writing publicity for the department. During the period from 1967 to 1972, he

handled his tasks reasonably well. He was also sent to special workshop school during two summers to study public relations. After his completion of this workshop program he was given his second pay-grade promotion in the organization or group. He did not wish to appear before any group to give speeches or to attend any other functions that were required in the public relations program.

In the past few years, the Public Relations Department has grown and expanded. Several young people have been added to the department. A couple of these have already passed the level of authority that is still held by John Senn. John performs the duties assigned to him quite adequately, and has developed into a capable writer. In general, Management is convinced that he performs his assigned tasks well enough, but that he does not have an outgoing personality. He is considered to be too introverted to attain a position of higher authority. He also has a tendency to receive rather than to initiate.

One month ago, you were made Department Head when the past department head took a job with another university. At the beginning of this week, John approached you and said that he wanted to be given more important tasks to handle. He believes that as a long-term employee of the University, he should be given more responsibility. In short, he feels that it is quite unfair and ungrateful of the University to promote younger and less-experienced people to positions higher than his.

The past department head did not feel that John was qualified to handle problems any more complex than those he deals with at present. However, she did not want to hurt his feelings or discourage him in his present job. Consequently, her annual ratings of him reflected an above-average employee. In addition, both informal and formal meetings between the two were congenial, and she rarely mentioned any need for improvement in performance. Still, she was convinced that John could not now, nor in the future, be promoted to a position involving more responsible work. Yet, John is a steady and dependable worker. If John should quit, his departure would constitute a loss to the department. You have asked John for some time to think over the request. You are concerned that he will not receive the department head's appraisal of him positively. Still, he should be told that he has, in upper management's opinion, realized all his potential and will not go any higher in the organization.

CASE TWO

The Case of the Ambitious Unhappy Instructor

In 1970, the Carnation Simulation Training Center was started with a complement of three employees and very little equipment. At present, the company has over 400 employees and assets in excess of $10 million. The rapid growth has been due in large to the higher cost for maintaining in-house training programs and the recognition that simulation is a bona fide training technique.

At the Center, there is a head of each branch who reports directly to the vice president in charge of the Simulation Division. At present, the Personnel Department occupies a relatively low position in the structure of the Carnation Center. It is small in size and is largely manned by former technical personnel. Its activities are confined primarily to screening applicants, administering employee benefits, directing company security, and maintaining personnel records. All personnel who hold supervisory or executive positions at Carnation have strong technical backgrounds.

Bob Rose, who is 29 years of age, and has been with the company for six years is an Assistant Instructor (a non-supervisory position). Assistant Instructors are responsble for helping Executive Instructors run training sessions. Part of Bob's job involves writing training programs. Bob feels that he is entitled to a position of Executive

Instructor (a supervisory position that involves supervising Assistant Instructors, making assignments to new simulation tasks, as well as supervising some clerical personnel).

In the opinion of Bob's supervisor, he is highly competent technically. He always completes his work on time, but prefers to keep to himself. While his fellow employees recognize Bob's technical expertise as an Assistant Instructor, and in fact, come to him for advice on matters pertaining to writing training programs and helping out at training sessions, he is not overly popular (but not disliked either).

According to his supervisor, Bob's biggest weakness is his ability to supervise others. To some he gives too much and too detailed instruction, and they soon feel that their intelligence is being insulted. Others feel that they are not getting enough information, and that they are lost and do not know what to do.

Bob is unhappy in his position as Assistant Instructor and has appealed to his supervisor for support in being promoted to Executive Instructor. While Bob's boss feels unsure about the promotion, Bob's supervisor's boss believes Bob's technical expertise would make him a fine Executive Instructor. Consequently, based on these recommendations, Bob is promoted to Executive Instructor.

Once promoted, however, Bob soon begins to have problems with his subordinates. He has a hard time making assignments, and often, when technical problems occur, he

prefers to do the work himself. Six months after being promoted Bob leaves the Carnation Simulation Center, unhappy with his job, and upset about complaints from his subordinates and criticism from his bosses.

Common Rating Errors

- HALO EFFECT

- LENIENCY/SEVERITY ERRORS

- RESTRICTION OF RANGE

- SIMILARITY ERROR

Illustration of Halo Effect

PERFORMANCE FACTORS	RATER ONE	RATER TWO
Job Knowledge	4	4
Quality of Work	4	3
Productivity	4	3
Record Keeping	3	4
Dependability	4	2
Adaptability	4	3
Initiative	3	2
Attendance	4	4
Relations with Others	4	3
Safety	4	4

Illustration of Leniency/Severity

Overall Evaluation

EMPLOYEES	RATER ONE	RATER TWO
1	3.80	3.80
2	3.75	2.75
3	3.90	3.50
4	3.80	3.75
5	3.75	1.75
6	3.70	3.25

Illustration of Range Restriction

Overall Evaluation

EMPLOYEES	RATER ONE	RATER TWO
1	2.75	3.50
2	2.60	3.70
3	2.70	2.20
4	3.00	3.20
5	2.85	3.80
6	2.90	1.90

APPENDIX B

Observation Training Lecture and Discussion Materials

OBSERVATION TRAINING

Training Appraisers

Evaluating employee performance is an important and necessary part of any supervisor's job. Regardless of whether your organization employs a formal system of employee evaluation, judgments about how individual employees are performing are made almost daily. People are constantly making judgments about others. Unfortunately, many of these informal judgments may be erroneous.

Consequently, a formal system of performance evaluation is usually adopted to help reduce the possibilities of bias and uninformed judgments; to standardize the types of information that will be forthcoming; and to insure that the resulting appraisal information is gathered in a form that permits its use across the entire organization.

While a formal system of appraisal helps to standardize this process in the organization, it in no way guarantees consistent, accurate evaluation. Therefore, the purpose of today's talk will be to begin to help you be more accurate in your judgment of employees. We will deal with this problem by focusing on the importance of developing good observation skills.

General Instructions/Hints/Precautions

First, I'd like to focus on three things to keep in mind when observing behavior, namely, careful observation, the observation of specific behaviors, and the need to take notes if possible.

Insert overhead about here

An extremely important part of evaluating an employee's performance is being a careful observer.

Careful Observation of Behavior

Prior to filling out the formal performance evaluation, you periodically come in contact with (often daily/sometimes less frequently) that employee in the course of performing his or her job duties. It is important that at these times you observe carefully their job-related behavior.

In addition, it may be helpful to think of the behaviors that you have observed during an evaluation period, as a sample of all the job behaviors exhibited by your subordinate during the rating period. Consequently, the behaviors you're actually seeing are only a small portion of the total number of behaviors, and therefore it's important to observe these behaviors carefully.

In addition, keep in mind that when observing an employee's job behavior, you do not necessarily have to be

physically present. There are many ways you can obtain sound information on performance. For instance, you might rely on a subordinate's oral or written reports that might reflect employee performance. Also, another supervisor may have had occasion to observe directly one of your employees, and thus can provide you with feedback concerning the subordinate's behavior. In general, then, the key is to collect as many relevant observations as possible, both through direct, careful observation and from other relevant observers.

Watch for Specific Behaviors

Insert overhead about here

It would be nice to believe that the task of making specific, accurate observations could be done objectively with only minimal interference from subjective factors. Obviously, however, the subjectivity involved in evaluating people is always going to be a factor, simply because we choose to pay attention to certain things or activities while we ignore others. It is impossible to observe everything in a given situation at the same time; while we are focusing on some attributes of a situation, we are naturally missing others. One way to use this selective

attention to our advantage in terms of evaluating employees, is to keep in mind those performance factors on the evaluation form on which we rate employees.

Insert overhead about here

For example, your evaluation form consists of factors like job knowledge, dependability, initiative, relation with others; or work habits, managerial skills, communication skills, planning and development skills. By keeping these performance categories in mind, they can help us to focus on those specific job behaviors that relevant when it comes time to evaluate our employees.

Take Notes

While it is not feasible to write down continually all observed behaviors, it's often beneficial to jot down (and file) behaviors you observe from time to time.

Insert overhead about here

Keep in mind that 12 months is a long time between evaluations, and many important behaviors occur, most of which will be forgotten unless recorded in some fashion. In addition, if nothing is written down, what will tend to be

COMMONWEALTH OF VIRGINIA

CONFIDENTIAL
EMPLOYEE PERFORMANCE EVALUATION

Name_____ Soc. Sec. No._____ Position No._____

Agency Name_____ Sub. Division_____ Agency Code_____

Class Title_____ Class Code_____ Date Entered Present Position_____

Date of Evaluation_____

Describe Briefly the Principal Duties in Present Job_____

_____

_____

PART I - PERFORMANCE FACTORS - CIRCLE THE APPROPRIATE PERFORMANCE LEVEL

1- JOB KNOWLEDGE/SKILLS - To what extent does the employee maintain a satisfactory level of job   4   3   2   1
knowledge and/or job skills?

Remarks_____

_____

_____

2- QUALITY OF WORK - To what extent does the employee's work meet the required quality standards: 4   3   2   1
i.e., accuracy, neatness and thoroughness?

Remarks_____

_____

_____

3- PRODUCTIVITY - To what extent does the employee accomplish the quantity of work expected of the   4   3   2   1
job assignment?

Remarks_____

_____

_____

4- RECORD KEEPING/DOCUMENTATION - To what extent does the employee adequately prepare and   4   3   2   1
maintain records, written reports, correspondence, and files?

Remarks_____

_____

_____

| | 4 | 3 | 2 | 1 |

5- <u>DEPENDABILITY</u> - To what extent does the employee perform work without close supervision or assistance?

Remarks_____

_____

_____

| | 4 | 3 | 2 | 1 |

6- <u>ADAPTABILITY</u> - To what extent does the employee readily adapt to new situations and changes in routines, work load, and/or work assignments?

Remarks_____

_____

_____

| | 4 | 3 | 2 | 1 |

7- <u>INITIATIVE</u> - To what extent does the employee present new ideas, improve procedures or otherwise demonstrate an awareness of clerical or technical changes related to the job?

Remarks_____

_____

_____

| | 4 | 3 | 2 | 1 |

8- <u>ATTENDANCE</u> - To what extent does the employee maintain satisfactory attendance performance in regard to tardiness, early departures, and/or absences?

Remarks_____

_____

_____

| | 4 | 3 | 2 | 1 |

9- <u>RELATIONS WITH OTHERS</u> - To what extent does the employee establish effective working relationships when dealing with supervision, co-workers, and/or the public?

Remarks_____

_____

_____

| | 4 | 3 | 2 | 1 |

10- <u>SAFETY</u> - To what extent does the employee work in a safe manner and observe safety practices?

Remarks_____

_____

_____

PART II - PERFORMANCE FACTORS - The following performance factors tend to reinforce the performance levels identified in Part I. The supervisor in completing Part II should indicate the employee's performance level by circling the appropriate level of performance. Use the remarks section to record your comments.

1- WORK HABITS - To what extent does the employee demonstrate adaptability, and a sense of priorities?   4   3   2   1

Remarks_____

_____

_____

_____

_____

2- PLANNING AND ANALYTICAL ABILITY - To what extent does the employee demonstrate the skills to analyze and solve problems?   4   3   2   1

Remarks_____

_____

_____

_____

_____

3- MANAGERIAL SKILLS - To what extent does the employee effectively work well with and through others to complete assignments in a timely and productive manner?   4   3   2   1

Remarks_____

_____

_____

_____

_____

4- COMMUNICATIONS SKILLS - To what extent can the employee effectively express himself/herself orally and in writing including correspondence and reports and presentations at conferences, seminars, workshops, etc., as required by the job?   4   3   2   1

Remarks_____

_____

_____

_____

5- DEVELOPMENT OF OTHERS - To what extent does the employee develop others to become more effective in work assignments and better prepared for future job opportunities?   4   3   2   1

Remarks_____

_____

_____

_____

_____

remembered will be those especially negative events, and the most recently observed behaviors—neither of which may be very representative of a particular employee's job performance during the entire year.

Insert overhead about here

Remember, observe performance carefully, watch for specific behaviors, and take notes when possible. If, when you're in contact with a particular employee, you are careful in what you observe or think you observe; if you focus on the behaviors relevant to the performance factors you'll be rating on, and if you jot down a few notes when possible, it ought to help you be more accurate when you sit down to formally evaluate them.

Systematic Errors of Observation

In addition to talking about what sorts of things we should do to be more accurate observers, we must also discuss some of the errors observers of behavior often make. I'd like to talk about two general areas where errors in observation may occur, and consequently, adversely affect your ratings.

Insert overhead about here

Contamination from Prior Information

Several sorts of common observational errors result from contamination from prior information about the employee being evaluated.

First, it is often an unintentional tendency of people to distort information observed, in a way that makes it similar to previously received information. Thus, for example, a supervisor might have noticed that a particular employee has left work 10-15 minutes before quitting time several times in the last several weeks. Now, whenever the end of the workday approaches and the supervisor notices the employee away from his/her assigned station, it is assumed immediately that the employee has left work early again, and subsequently, is marked down on attendance on the next performance evaluation. Thus, prior information, regardless of how accurate it is can influence your expectations, which may influence your observations.

In addition, one aspect of observed behavior may tend to influence unduly your overall observation and evaluation of an employee. Consequently, while the evaluation form has ten separate areas to evaluate each employee, a poor score on one observed factors (such as attendance) may influence you to give low ratings on many of the other performance factors, regardless of whether poor performance was observed on the other factors. So, be aware that prior observation can and may affect future observations.

Insert overhead about here

Overdependence on a Single Source of Information

Another prevalent observational source of error is that generated by an overreliance on a single source of information. While this source is often the most reliable source of information, it may also be a major source of error. This is so because in many instances what causes one observational source to take precedence, and to be relied on almost exclusively, is ease of acquisition. In other words, whatever way some information about an employee can be gathered most quickly and easily (regardless of whether it's accurate) that way is often relied upon. Unfortunately, as I'm sure you're aware, this can lead to misleading and inaccurate evaluations. In addition, if you collect only a limited amount of observational information, your judgments have to be based on what's available (and not necessarily what's a more complete, accurate picture).

Summary

Insert overhead about here

Thus, due to the ease and frequency with which these observational errors are committed, it is important to remember the things we have talked about today:

Things to do -- observe carefully.

watch for relevant behaviors.

take notes whenever possible.

Things to avoid -- contamination from prior information

overreliance on a single source of information

Therefore, if you keep in mind some of these things to do to be better observers, and be aware of some errors that can occur, they should help you be more accurate when you evaluate your employees.

DO'S AND DON'TS

Things to Do --   OBSERVE CAREFULLY

WATCH FOR RELEVANT BEHAVIORS

TAKE NOTES WHENEVER POSSIBLE

Things to Avoid --   CONTAMINATION FROM PRIOR INFORMATION

OVER-RELIANCE ON A SINGLE DATA SOURCE

EXERCISE ONE

RECRUITER PERFORMANCE FACTORS

1. Creating a Favorable Image of the Company
        presenting a positive, but realistic image of GCI;
        spelling out clearly the advantages of working for
        GCI.

2. Organizing the Interview
        structuring the interview to allow for an
        appropriately balanced information exchange
        between recruiter and interviewee; giving the
        interviewee a chance to ask questions; defining
        the purpose of the interview.

3. Providing Relevant Information About the Company
        giving the interviewee specific information about
        the characteristics of various jobs so that he/she
        can make informed decisions; displaying
        familiarity with programs at GCI and their
        requirements; demonstrating knowledge about
        benefits, promotions, pay, etc.

4. Asking Relevant Questions
        asking questions which maximize the amount of
        meaningful information available to the
        interviewer; asking the interviewee questions
        he/she can understand and respond to readily;
        making clear the information desired.

5. Answering Recruitees' Questions
        providing complete, clear, concise and accurate
        answers to interviewees' questions; answering
        interviewees' questions so that they have the
        information desired; ensuring that the interviewee
        understands the recruiter's answer.

6. Establishing Rapport with Interviewees
        developing a nonthreatening relationship with the
        interviewee; creating a relaxed atmosphere;
        gaining the friendship and trust of the
        interviewee.

A. CREATING A FAVORABLE IMAGE FOR THE COMPANY

Presenting a positive, but realistic image of GCI; spelling out clearly the advantages of working for GCI versus presenting a negative or misleading positive image of GCI; failing to outline positive aspects such as available programs and opportunities at GCI.

High Level Performance

- Gives the Interviewee a very broad and accurate picture of the history and current features of the company.

- Tells the Interviewee about special company features that make it well fitted to the capabilities and interests of the Interviewee.

What a high level performer might do:

7. This interviewer can be expected to discuss the history of GCI and how it is currently organized, to describe several ways in which GCI is better than other companies, and to point out specific features of GCI which make it fit the ability and experience of the Interviewee better than most other companies.

6. Would expect this Interviewer to comment on the many training and development programs that GCI has and to point out that few major companies offer so many exceptional opportunities.

(vertical: EXPECTED PERFORMANCE)

Average Performance

- Gives the Interviewee a reasonably good general picture of the company.

- Tells the interviewee about some of the things that are especially good about the company.

What an average performer might do:

5. Can be expected to tell Interviewees about GCI's excellent record in the area of environmental pollution to show how progressive and socially aware GCI's top management is.

4. This Interviewer can be expected to emphasize one or two central reasons why the Interviewee should join GCI and to state that other aspects aren't important.

3. Would expect this Interviewer to steer conversation away from areas where GCI does not excel and to spend time describing all the favorable things about the company.

(vertical: EXPECTED PERFORMANCE)

Low Level Performance

- Refrains from talking much about the company and provides some facts that are not entirely accurate.

- Provides no solid reasons for joining GCI and may inappropriately mention one or more negative aspects of working at GCI.

What a low level performer might do:

2. This interviewer can be expected to mention many bad points about GCI before getting around to mentioning anything good about it.

1. Can be expected to tell the interviewee almost nothing about reputation, benefits, or opportunities offered by GCI and to say nothing about reasons for joining the company.

B. ORGANIZING THE INTERVIEW

Structuring the interview to allow for an appropriately balanced information exchange between recruiter and interviewee; giving the interviewee a chance to ask questions; defining the purpose of the interview versus displaying inadequate organization or planning for the interview; providing inadequate time to ask questions; failing to provide a definition of the interview's purpose.

High Level Performance

- Starts the interview by outlining with the interviewee exactly the kinds of things they will be talking about during the interview and then follows the plan closely.

- Structures the interview so that both the recruiter and the interviewee will have enough time to ask questions and to provide information.

What a high level performer might do:

7. Can be expected to begin by telling the interviewee that he/she will ask some questions to obtain an idea of the interviewee's qualifications and interests, then to discuss why GCI is a good place to work, and finally, to allow the interviewee to ask whatever questions he/she wants to.

6. This interviewer can be expected to state at the beginning of the interview that he/she wants to spend an equal amount of time discussing the opportunities at GCI, answering the interviewee's questions, ans asking some of his/her own.

PERFORMANCE EXAMPLES

Average Performance

- Starts the interview by suggesting a general plan and then follows this plan through most of the session.

- Starts the interview without spelling out a firm structure but manages to provide a reasonably good balance of information exchange anyway.

What an average performer might do:

5. Would expect this interviewer to state after a few pleasantries, "Let's talk about you and GCI," and then to ask the interviewee about interests. Can also be expected to ask the interviewee what he/she wants to get out of GCI, what his/her qualifications are, and then explain how the interviewee can fit into one of GCI's training programs.

4. Can be expected to state carefully the purpose of the interview and then to follow a check list of "things to cover" during the session.

3. Can expect this interviewer to appear somewhat rushed and harried during the interview, would also expect this interviewer to provide enough time to describe GCI but to cause the interviewee to remind him/her that he/she has some questions.

PERFORMANCE EXAMPLES

Low Level Performance

- Starts the interview in a conversational manner without suggesting any plan of things to be covered and maintains this loose organization throughout the interview.

- Conducts the interview in a rambling and disorganized way so that the exchange of information between recruiter and interviewee becomes unbalanced.

What a low level performer might do:

2. This interviewer can be expected to tell the interviewee to talk about anything he/she wants to and then to sit back and wait. Can also be expected to provide direct answers to questions but to rely on the interviewee to direct and lead the interview.

1. Can expect this interviewer to start talking and asking questions about one thing before finishing up preceding comments such that some questions, answers, and explanations are run together resulting in the interviewee becoming extremely confused.

6. PROVIDING RELEVANT INFORMATION ABOUT THE COMPANY

Giving the interviewee specific information about the characteristics of various jobs so that he/she can make informed decisions; displaying familiarity with programs at GCI and their requirements; demonstrating knowledge about benefits, promotions, pay, etc. versus presenting inadequate information about programs relevant to the interviewee's background and interests; displaying a lack of knowledge about benefits, promotions, pay, etc.

High Level Performance

• Provides complete information about all facets of the company including various jobs that might be appropriate for the interviewee.

• Gives comprehensive details of jobs and programs available in the company.

What a high level performer might do:

PERFORMANCE EXAMPLES

7. Can be expected to give specific details about the requirements of the management trainee program such as possible continued training, salary, fringe benefits, promotion possibilities, job duties, etc.

6. This interviewer will be expected to display considerable familiarity with GCI's training and benefit programs and to provide basic information about a wide variety of jobs.

Average Performance

• Provides a broad overview of the company and gives details about some of the programs and jobs that might interest the interviewee.

• Has sufficient knowledge about company to answer most of the interviewee's questions but usually doesn't provide specifics.

What an average performer might do:

PERFORMANCE EXAMPLES

5. Would expect this interviewer to display considerable information about most jobs the interviewee is interested in, except for some of the job content changes in engineering divisions.

4. When asked specific questions about a certain mechanical engineering position, this interviewer would be expected to give the interviewee a general idea and to offer to find out more particulars for him/her after the interview.

3. This interviewer, although conversant with the general content of most jobs at GCI, would be expected to refer the interviewee to the recruiting brochure when questions arise about pay, training, or promotion opportunities.

Low Level Performance

• Seems to have knowledge about some but not all facets of the company and provides only a limited amount of information to the interviewee.

• Seems to lack knowledge about most jobs and programs relevant to the interviewee and does not give much useful knowledge to the interviewee.

What a low level performer might do:

PERFORMANCE EXAMPLES

2. When asked about positions outside the technical area, can expect this interviewer to state that he/she has come from a technical division at GCI and knows only about jobs in that division.

1. This interviewer may be expected to display little or no knowledge about training opportunities interviewees are interested in and to express ignorance about the existence of GCI's management training program.

D. ASKING RELEVANT QUESTIONS

Asking questions which maximize the amount of meaningful information available to the interviewer; asking the interviewee questions he can understand and respond to readily; making clear the information desired; versus asking questions irrelevant to the job or difficult to answer; unnecessarily confusing the interviewee concerning the information desired.

High Level Performance

Asks easily understood questions that are relevant to the interviewee and to the job for which he/she is being considered.

Asks clear questions in a logical way so that the maximum amount of useful information is obtained.

What a high level performer might do:

Would expect this interviewer to ask simple, open-ended questions, enabling the interviewee to give rich, yet pertinent information about himself/herself.

Can be expected to ask relevant, straightforward questions which leave the interviewee certain of what is being asked and which yield answers the interviewer can use to make a judgment about the interviewee's suitability for GCI.

PERFORMANCE EXAMPLES

Average Performance

Asks clear questions and obtains good information, but some seem somewhat irrelevant to the job or to the interviewee.

Asks questions that are clear and easily understood but sometimes gets somewhat "off track" in getting the most meaningful information.

What an average performer might do:

5. This interviewer would be expected to ask short, to-the-point questions such as "Why did you like that particular course best?"

4. For the most part, this interviewer would ask questions relevant to determining the interviewee's potential for an opening in sales at GCI, only occasionally confusing the interviewee about what information was being asked for.

3. Can be expected to ask somewhat vague and general questions, such as "Tell me about yourself" without expanding the questions further.

PERFORMANCE EXAMPLES

Low Level Performance

Asks questions that are rather confusing and often difficult to answer.

Asks vague questions that often seem irrelevant so that only a limited amount of meaningful information is obtained.

What a low level performer might do:

2. This interviewer may be expected to ask long, involved questions which often confuse the interviewee.

1. Would expect this interviewer to ask several questions, one after the other, without giving the interviewee a chance to respond fully to any of them.

107

E. ANSWERING RECRUITEES' QUESTIONS

Providing complete, clear, concise, and accurate answers to interviewees' questions; answering interviewees' questions so that they have the information desired; ensuring that the interviewee understands the recruiter's answer _versus_ providing incomplete, confusing or inaccurate answers; attempting to avoid questions

High Level Performance

- Carefully answers all questions, making certain the answers are complete and accurate and that the interviewee understands the answers.

- When appropriate provides important extra information related to a question.

What a high level performer might do:

7. Would expect this interviewer to be meticulous about answers, to go into detail, and to circle important information in the brochures. Would also expect this interviewer to ask if he/she had answered the questions completely.

6. Can be expected to expand the scope of the interviewee's questions, resulting in lengthier, but also more complete answers.

Average Performance

- Usually answers interviewees' questions competently but may get sidetracked during an explanation or, may sometimes answer questions only indirectly.

- Occasionally provides incomplete or unclear answers but generally ensures that interviewees' receive the information desired.

What an average performer might do:

5. Would be expected to answer most questions completely, to note carefully questions he/she cannot answer, and to tell the interviewee that the answer will be obtained from someone who knows.

4. When asked about GCI's pension plan, can be expected to provide a complete explanation but to ramble on somewhat about the virtues of the plan.

3. This interviewer would be expected to answer most questions competently but to try to bypass a few of the interviewee's questions.

Low Level Performance

- Frequently provides incomplete and superficial answers to legitimate questions about GCI.

- Avoids interviewees' questions or provides confusing responses.

What a low level performer might do:

2. Can be expected to downplay the interviewee's questions and give only prepared statements in place of answers.

1. Would expect this interviewer, when asked a tough question, to tell the interviewee to read the recruiting brochure.

PERFORMANCE SCALE

PERFORMANCE SCALE

F. ESTABLISHING RAPPORT WITH INTERVIEWEES

Developing a nonthreatening relationship with the interviewee; creating a relaxed atmosphere; gaining the friendship and trust of the interviewee versus failing to establish rapport with the interviewee; creating a cold or threatening atmosphere; failing to put the interviewee at ease.

High Level Performance

- Develops a relaxed atmosphere by talking about a common interest or by asking questions which set the interviewee at ease.

- Greets the interviewee with courtesy and gains the interviewee's trust by being sincere, warm, and personable.

What a high level performer might do:

7. Would expect this interviewer to begin by talking about an interest in common with the interviewee and to ask questions only after the interviewee is talking freely.

6. This interviewer would greet the interviewee warmly, offer him/her a chair, and spend a short time conversing about his/her alma mater. Then he/she would get down to business and start asking questions.

P
E
R
F
O
R
M
A
N
C
E

E
X
A
M
P
L
E
S

Average Performance

- Is relaxed and friendly during portions of the interview but also comes on in a very business-like, task oriented way at other times in the session.

- Sets the interviewee somewhat at ease by engaging in small talk at the beginning of the interview or by joking with him/her at appropriate times.

What an average performer might do:

5. Can be expected to laugh freely when the interviewee makes a joke about his/her past experiences.

4. This interviewer would be expected to begin the interview by making small talk about sports after noticing that the interviewee was a college football player.

3. Can expect this interviewer to be somewhat skeptical and reserved when the interview begins but to become more relaxed and talkative once into the interview.

P
E
R
F
O
R
M
A
N
C
E

E
X
A
M
P
L
E
S

Low Level Performance

- Interacts in a cold and detached manner during the interview, and is generally unresponsive to the interviewee.

- Creates a threatening atmosphere by immediately asking personal questions or by appearing suspicious of interviewees and their credentials.

What a low level performer might do:

2. As soon as the interviewee sits down, would expect this interviewer to begin asking him/her questions about his/her background.

1. This interviewer can be expected to appear detached throughout the interview and not to smile, speak, or nod to the interviewee other than to ask or answer questions.

P
E
R
F
O
R
M
A
N
C
E

E
X
A
M
P
L
E
S

EXERCISE TWO

The Case of Ambition Exceeding Ability

In 1961, John Senn was hired by the University as a bookkeeper trainee. Prior to this job, he had worked for a short time on a weekly newspaper, but he had been replaced by a man who could sell, as well as write. He also had a brief job as an apprentice sign painter, but had quit due to lack of interest. From 1961 to 1967, he had shown little promise of success in his job, advancing only one step from bookkeeper to clerk. In addition, during this time John and his supervisor had several bitter arguments, the result of personality clashes.

In 1967, a Public Relations Department was organized in the University. A woman was brought in from the outside to initiate and develop this activity. She had a considerable amount of experience in the public relations field, having worked for a large newspaper in that capacity. In addition, she had been a business writer for the Wall Street Journal, and more recently headed her own advertising agency.

John Senn asked for a transfer to this new department because of his earlier experience in writing for the weekly newspaper. Based on information gathered from John,s supervisor, the new head of Public Relations was reluctant to approve the transfer, but was persuaded to do so by her boss. John was assigned to writing publicity for the department.

During the period from 1967 to 1972, he handled his tasks reasonably well. He was also sent to special workshop school during two summers to study public relations. After his completion of this workshop program, he was given his second pay-grade promotion in the organization.

In the past few years, the Public Relations Department has grown and expanded. The department has been divided into two divisions, a Writing Division (of which John is a member), and a Public Lecture Division (of which you are a member). Even though you are in separate divisions you see John daily, and rate samples of his work above-average. You feel John has developed into a capable writer. In general, the Writing Division supervisor is convinced that he performs his assigned tasks well enough, but that he does not have an outgoing personality. Also, his file indicates he has a tendency to receive rather than initiate.

One month ago, you were named the new supervisor of the Writing Division, when the past supervisor took a job in another city. At the beginning of this week, John approached you and said that he wanted to be given more important tasks to handle. He believes that as a long-term employee of the University, he should be given more responsibility. In short, he feels that it is quite unfair and ungrateful of the University to promote younger and less-experienced people to higher positions than his.

_____

Remember, you are the supervisor ... what do you do? Do you tell him you agree with him, you disagree with him, tell him you will think about it, or tell him something else?

APPENDIX C

Decision-Making Training Lecture and Discussion Materials

DECISION-MAKING TRAINING

Training Appraisers

Evaluating employee performance is an important and necessary part of any supervisor's job. Regardless of whether your organization employs a formal system of employee evaluation, judgments about how individual employees are performing are made almost daily. People are constantly making judgments about others. Unfortunately, many of these informal judgments may be erroneous.

Consequently, a formal system of performance evaluation is usually adopted to help reduce the possibilities of bias and uninformed judgments; to standardize the types of information that will be forthcoming; and to ensure that the resulting appraisal information is gathered in a form that permits its use across the entire organization.

While a formal system of appraisal helps to standardize this process in the organization, it in no way guarantees consistent, accurate evaluation. Therefore, the purpose of today's talk will focus on helping you be more accurate in your judgment of employees. Specifically, we will talk about some common strategies that are used by supervisors to arrive a final evaluation decisions. In addition, we will talk about how some of these strategies are effective, while others can cause errors in judgments.

Intuitive-vs-Formal Decision-Making Strategies

In the work setting, suprevisors are continually confronted with information they must identify and use in making decisions. However, each decision-maker utilizes only a limited amount of information in arriving at any particular decision. Applied to the area of performance evaluation, this tendency for people to make decisions without full use of the available information is greatly increased by the fact that each supervisor can devote only a limited amount of time to evaluating employees.

As a result of this time limitation, decision-makers often rely on simple cognitive or judgmental strategies to make quick, hopefully accurate decisions. For example, when a supervisor is told that a particular person has behaved in a particular way, he may quickly review the behaviors that stand out in memory, and make a decision about whether to believe the information. These simple intuitive strategies (or rules of thumb) are quite often used unconsciously, and yet they are used appropriately and effectively in the great majority of cases. Such strategies allow us to deal with problems and make decisions without processing a great deal of information. However, our concern in the remainder of this lecture will focus on the times when these intuitive strategies are used inappropriately, and consequently, effect out ability to make accurate judgments when evaluating our employees.

Judgmental Errors

_____

Insert overhead about here

_____

Insensitivity to Biased Data

An overreliance on intuitive strategies can result in a judgmental error, for example, when a supervisor relies on too few samples of behavior. If you were to check samples of an employee's work only a few times during the rating period, your conclusions (and therefore, your evaluation) might be based on biased information. For example, if you were to observe her four times during the year (and two of these times you found work quality below par), it's possible that work quality might have slipped just two times out of 50, but two of your (say) four pieces of information were collected on those days.

_____

Insert overhead about here

_____

In addition to this bias in how and what we observe, judgmental error can also result from what we remember. It often occurs that we do not remember all that we observe. However, what we do remember is typically the information that confirms our beliefs about the evidence we currently possess. In other words, we possess a biased system of

recollection as well. This biased system can also be affected by the vividness of these events, in that vivid information (an employee caught dozing) will stand out in memory (even if it was only one time).

One way to deal with the fact that biased information can and may affect the accuracy of our evaluations, is to collect as much information as possible about an individual's behavior--by increasing the frequency of observing work samples and collecting information from other sources (e.g., co-workers). In addition, it is important that we observe behavior carefully so that we are collecting accurate information on which to base our evaluation.

Insert overhead about here

Inappropriate Causal Inference

Another cognitive strategy that is often beneficial, yet when used inappropriately can cause us to err in our judgments about the behavior of others, is often used when we observe behaviors, objects or situations together. In order to make sense of the scene we observe, we often link things together in a cause-and-effect relationship. This view that the events are related is then strengthened when the two are observed together more than one time.

For example, if you were to observe a particular employee on several occasions, and each time a piece of

equipment he/she was using broke down, you might interpret the information to mean that the employee was the cause of equipment breakdown.

Once again, let me emphasize I'm not trying to imply that making these cause and effect judgments is always inappropriate--but rather, it's an error that can occur, and therefore, we should keep it in mind when we are evaluating the behaviors we've observed.

Insert overhead about here

Overreliance on Previously Formed Theories

Another way that we make inappropriate decisions is through an overreliance on previously formed beliefs. The behaviors we have seen in the past, and the information read or heard concerning a particular employee, as well as any stereotypes or prejudices we possess will influence the way we look at current behaviors, or if we notice them at all. Research has shown that once people have applied a particular label to a given object, or formed a particular opinion about a set of behaviors, they place too much emphasis on that opinion when making future evaluations.

Insert overhead about here

For example, if you are supervising a group of Clerk Typists, and you believe females are superior to males at this position, this stereotypic view may effect how you evaluate a male clerk typist--regardless of the actual behaviors you observe.

Consequently, it is important to be aware that previous views can effect what is observed and how those behaviors are evaluated.

Insert overhead about here

Inappropriate Weighting

A final intuitive strategy that is frequently used involves the weighting of information that is available to us. This strategy is often valuable in helping us make accurate evaluations. However, how a supervisor weights the importance of a particular behavior or event is often due to the vividness of the information. And it is this vividness emphasis that can tend to distort our decisions, or the information we use to make our decisions. This distortion occurs because the vividness of the information is often not related to its true value as evidence to be used in an evaluation.

Insert overhead about here

For example, if during the past year, a particular employee of yours was late for work two times, that information probably would not figure significantly in our evaluation of that employee's behavior. However, if it so happened that the two mornings the employee was late, coincided with emergencies that arose requiring timely completion of a letter or work assignment, those two late arrivals may be vividly remembered and weighted quite heavily when evaluating that employee's dependability. In addition, even though dependability is only one of the performance factors you evaluate an employee on, an especially well-remembered behavior may affect how you rate that person on the other performance factors as well. In fact, research has shown us that negative information is typically weighted more heavily than is positive information.

Factors such as you emotional interest or involvement with the event or employee, the concreteness of the event, and how close the person or information is to us all tend to affect vividness, and consequently your weighting. Therefore, the key here seems to be to observe behaviors carefully, and if you know you cannot avoid giving a particular incident far more weight then is justified, then avoid using that piece of information, and rely on evidence that will be more truthful, and allow you to be more accurate in your evaluation.

<u>Summary</u>

There are several important things to remember when evaluating your employees.

1) We frequently "fall back" on judgmental strategies that help us to make decisions quickly and accurately with as little information as possible.

2) However, this can lead to problems--and therefore, when we are evaluating others we should be aware of these potential problems resulting from:

a) an insensitivity to biased data, including a reliance on too few examples of behavior, and the frequent recollection of biased information.

b) allowing prior beliefs/theories about an employee effect the information we use to make decisions.

c) inappropriately establishing cause-effect relationships between employees and incidents that are observed together.

d) inappropriate weighting of behavioral incidents.

Insert overhead about here

EXERCISE ONE

A. **INFORMATION ABOUT RECRUIT INTERVIEWER ONE**

Recruit Interviewer Bill Smith has been with GCI for five years, during which time he has received three promotions and four pay increases. At the present time, Bill is in charge of P e r s o n n e l Recruitment at GCI's branch office located near the University. Recently, events in Bill's life have left him quite confused and troubled. Six months ago, Bill's wife was diagnosed as having terminal cancer, and given less than one year to live. In addition, Bill's mother was killed in a tragic train—automobile collision less than two weeks ago. Needless to say, Bill is still in the process of trying to get his life in order, and prepare himself and his four children for the possible death of his wife.

B. **INFORMATION ABOUT RECRUIT INTERVIEWER TWO**

Recruit Interviewer Daniel Reeves has been with GCI for six weeks in his present position. Prior to coming to GCI, Daniel had spent several years in a similar capacity with one of GCI's competitors. He was terminated from that job, however, because of his inability to establish rapport with the prospective employees. In addition, his file indicated that he had an inability to "sell" these interviewees on the benefits of working with his company.

RECRUITER PERFORMANCE FACTORS

1. <u>Creating a Favorable Image of the Company</u>
         presenting a positive, but realistic image of GCI;
         spelling out clearly the advantages of working for
         GCI.

2. <u>Organizing the Interview</u>
         structuring the interview to allow for an
         appropriately balanced information exchange
         between recruiter and interviewee; giving the
         interviewee a chance to ask questions; defining
         the purpose of the interview.

3. <u>Providing Relevant Information About the Company</u>
         giving the interviewee specific information about
         the characteristics of various jobs so that he/she
         can make informed decisions; displaying
         familiarity with programs at GCI and their
         requirements; demonstrating knowledge about
         benefits, promotions, pay, etc.

4. <u>Asking Relevant Questions</u>
         asking questions which maximize the amount of
         meaningful information available to the
         interviewer; asking the interviewee questions
         he/she can understand and respond to readily;
         making clear the information desired.

5. <u>Answering Recruitees' Questions</u>
         providing complete, clear, concise and accurate
         answers to interviewees' questions; answering
         interviewees' questions so that they have the
         information desired; ensuring that the interviewee
         understands the recruiter's answer.

6. <u>Establishing Rapport with Interviewees</u>
         developing a nonthreatening relationship with the
         interviewee; creating a relaxed atmosphere;
         gaining the friendship and trust of the
         interviewee.

A. CREATING A FAVORABLE IMAGE FOR THE COMPANY

Presenting a positive, but realistic image of GCI; spelling out clearly the advantages of working for GCI versus presenting a negative or misleading positive image of GCI; failing to outline positive aspects such as available programs and opportunities at GCI.

High Level Performance

· Gives the interviewee a very broad and accurate picture of the history and current features of the company.

· Tells the interviewee about special company features that make it well fitted to the capabilities and interests of the interviewee.

What a high level performer might do:

7. This interviewer can be expected to discuss the history of GCI and how it is currently organized, to describe several ways in which GCI is better than other companies, and to point out specific features of GCI which make it fit the ability and experience of the interviewee better than most other companies.

6. Would expect this interviewer to comment on the many training and development programs that GCI has and to point out that few major companies offer so many exceptional opportunities.

Average Performance

· Gives the interviewee a reasonably good general picture of the company

· Tells the interviewee about some of the things that are especially good about the company.

What an average performer might do:

5. Can be expected to tell interviewees about GCI's excellent record in the area of environmental pollution to show how progressive and socially aware GCI's top management is.

4. This interviewer can be expected to emphasize one or two central reasons why the interviewee should join GCI and to state that other aspects aren't important.

3. Would expect this interviewer to steer conversation away from areas where GCI does not excel and to spend time describing all the favorable things about the company.

Low Level Performance

· Refrains from talking much about the company and provides some facts that are not entirely accurate.

· Provides no solid reasons for joining GCI and may inappropriately mention one or more negative aspects of working at GCI.

What a low level performer might do:

2. This interviewer can be expected to mention many bad points about GCI before getting around to mentioning anything good about it.

1. Can be expected to tell the interviewee almost nothing about reputation, benefits, or opportunities offered by GCI and to say nothing about reasons for joining the company.

C. ORGANIZING THE INTERVIEW

Structuring the interview to allow for an appropriately balanced information exchange between recruiter and interviewee; giving the interviewee a chance to ask questions; defining the purpose of the interview versus displaying inadequate organization or planning for the interview; providing inadequate time to ask questions; failing to provide a definition of the interview's purpose.

High Level Performance

- Starts the interview by outlining with the interviewee exactly the kinds of things they will be talking about during the interview and then follows the plan closely.

- Structures the interview so that both the recruiter and the interviewee will have enough time to ask questions and to provide information.

What a high level performer might do:

7. Can be expected to begin by telling the interviewee that he/she will ask some questions to obtain an idea of the interviewee's qualifications and interests, then to discuss why GCI is a good place to work, and finally, to allow the interviewee to ask whatever questions he/she wants to.

6. This interviewer can be expected to state at the beginning of the interview that he/she wants to spend an equal amount of time discussing the opportunities at GCI, answering the interviewee's questions, and asking some of his/her own.

Average Performance

- Starts the interview by suggesting a general plan and then follows this plan through most of the session.

- Starts the interview without spelling out a firm structure but manages to provide a reasonably good balance of information exchange anyway.

What an average performer might do:

5. Would expect this interviewer to state after a few pleasantries, "Let's talk about you and GCI," and then to ask the interviewee about interests. Can also be expected to ask the interviewee what he/she wants to get out of GCI, what his/her qualifications are, and then explain how the interviewee can fit into one of GCI's training programs.

4. Can be expected to state carefully the purpose of the interview and then to follow a check list of things to cover during the session.

3. Can expect this interviewer to appear somewhat rushed and harried during the interview, would also expect this interviewer to provide enough time to describe GCI but to cause the interviewee to remind him/her that he/she has some questions.

Low Level Performance

- Starts the interview in a conversational manner without suggesting any plan of things to be covered and maintains this loose organization throughout the interview.

- Conducts the interview in a rambling and disorganized way so that the exchange of information between recruiter and interviewee becomes unbalanced.

What a low level performer might do:

2. This interviewer can be expected to tell the interviewee to talk about anything he/she wants to and then to sit back and wait. Can also be expected to provide direct answers to questions but to rely on the interviewee to direct and lead the interview.

1. Can expect this interviewer to start talking and asking questions about one thing before finishing up preceding comments such that some questions, answers, and explanations are run together resulting in the interviewee becoming extremely confused.

C. PROVIDING RELEVANT INFORMATION ABOUT THE COMPANY

Giving the interviewee specific information about the characteristics of various jobs so that he/she can make informed decisions; displaying familiarity with programs at GCI and their requirements; demonstrating knowledge about benefits, promotions, pay, etc. versus presenting inadequate information about programs relevant to the interviewee's background and interests; displaying a lack of knowledge about benefits, promotions, pay, etc.

High Level Performance

- Provides complete information about all facets of the company including various jobs that might be appropriate for the interviewee.

- Gives comprehensive details of jobs and programs available in the company.

What a high level performer might do:

7. Can be expected to give specific details about the requirements of the management trainee program such as possible continued training, salary, fringe benefits, promotion possibilities, job duties, etc.

6. This interviewer will be expected to display considerable familiarity with GCI's training and benefit programs and to provide basic information about a wide variety of jobs.

E X A M P L E S

P E R F O R M A N C E S

Average Performance

- Provides a broad overview of the company and gives details about some of the programs and jobs that might interest the interviewee.

- Has sufficient knowledge about company to answer most of the interviewee's questions but usually doesn't provide specifics.

What an average performer might do:

5. Would expect this interviewer to display considerable information about most jobs the interviewee is interested in, except for some of the job content changes in engineering divisions.

4. When asked specific questions about a certain mechanical engineering position, this interviewer would be expected to give the interviewee a general idea and to offer to find out more particulars for him/her after the interview.

3. This interviewer, although conversant with the general content of most jobs at GCI, would be expected to refer the interviewee to the recruiting brochure when questions arise about pay, training, or promotion opportunities.

E X A M P L E S

P E R F O R M A N C E S

Low Level Performance

- Seems to have knowledge about some but not all facets of the company and provides only a limited amount of information to the interviewee.

- Seems to lack knowledge about most jobs and programs relevant to the interviewee and does not give much useful knowledge to the interviewee.

What a low level performer might do:

2. When asked about positions outside the technical area, can expect this interviewer to state that he/she has come from a technical division at GCI and knows only about jobs in that division.

1. This interviewer may be expected to display little or no knowledge about training opportunities interviewees are interested in and to express ignorance about the existence of GCI's management training program.

E X A M P L E S

P E R F O R M A N C E S

D. ASKING RELEVANT QUESTIONS

Asking questions which maximize the amount of meaningful information available to the interviewer; asking the interviewee questions he can understand and respond to readily; making clear the information desired; versus asking questions irrelevant to the job or difficult to answer; unnecessarily confusing the interviewee concerning the information desired.

High Level Performance

Asks easily understood questions that are relevant to the interviewee and to the job for which he/she is being considered.

Asks clear questions in a logical way so that the maximum amount of useful information is obtained.

What a high level performer might do:

- Would expect this interviewer to ask simple, open-ended questions, enabling the interviewee to give rich, yet pertinent information about himself/herself.

- Can be expected to ask relevant, straightforward questions which leave the interviewee certain of what is being asked and which yield answers the interviewer can use to make a judgment about the interviewee's suitability for GCI.

Average Performance

Asks clear questions and obtains good information, but some seem somewhat irrelevant to the job or to the interviewee.

Asks questions that are clear and easily understood but sometimes gets somewhat "off track" in getting the most meaningful information.

What an average performer might do:

5. This interviewer would be expected to ask short, to-the-point questions such as "Why did you like that particular course best?"

4. For the most part, this interviewer would ask questions relevant to determining the interviewee's potential for an opening in sales at GCI, only occasionally confusing the interviewee about what information was being asked for.

3. Can be expected to ask somewhat vague and general questions, such as "Tell me about yourself," without expanding the questions further.

Low Level Performance

Asks questions that are rather confusing and often difficult to answer.

Asks vague questions that often seem irrelevant so that only a limited amount of meaningful information is obtained.

What a low level performer might do:

2. This interviewer may be expected to ask long, involved questions which often confuse the interviewee.

1. Would expect this interviewer to ask several questions, one after the other, without giving the interviewee a chance to respond fully to any of them.

PERFORMANCE EXAMPLES

PERFORMANCE EXAMPLES

E. ANSWERING RECRUITEES' QUESTIONS

Providing complete, clear, concise, and accurate answers to interviewees' questions; answering interviewees' questions so that they have the information desired; ensuring that the interviewee understands the recruiter's answer versus providing incomplete, confusing or inaccurate answers; attempting to avoid questions

High Level Performance

- Carefully answers all questions, making certain the answers are complete and accurate and that the interviewee understands the answers.

- When appropriate provides important information related to a question.

What a high level performer might do:

7. Would expect this interviewer to be meticulous about answers, to go into detail, and to circle important information in the brochures. Would also expect this interviewer to ask if he/she had answered the questions completely.

6. Can be expected to expand the scope of the interviewee's questions, resulting in lengthier, but also more complete answers.

PERFORMANCE LEVELS

Average Performance

- Usually answers interviewees' questions competently but may get sidetracked during an explanation or, may sometimes answer questions only indirectly.

- Occasionally provides incomplete or unclear answers but generally ensures that interviewees' receive the information desired.

What an average performer might do:

5. Would be expected to answer most questions completely, to note carefully questions he/she cannot answer, and to tell the interviewee that the answer will be obtained from someone who knows.

4. When asked about GCI's pension plan, can be expected to provide a complete explanation but to ramble on somewhat about the virtues of the plan.

3. This interviewer would be expected to answer most questions competently but to try to bypass a few of the interviewee's questions.

PERFORMANCE LEVELS

Low Level Performance

- Frequently provides incomplete and superficial answers to legitimate questions about GCI.

- Avoids interviewees' questions or provides confusing responses.

What a low level performer might do:

2. Can be expected to downplay the interviewee's questions and give only prepared statements in place of answers.

1. Would expect this interviewer, when asked a tough question, to tell the interviewee to read the recruiting brochure.

F. ESTABLISHING RAPPORT WITH INTERVIEWEES

Developing a nonthreatening relationship with the interviewee; creating a relaxed atmosphere; gaining the friendship and trust of the interviewee versus failing to establish rapport with the interviewee; creating a cold or threatening atmosphere; failing to put the interviewee at ease.

High Level Performance

- Develops a relaxed atmosphere by talking about a common interest or by asking questions which set the interviewee at ease.

- Greets the interviewee with courtesy and gains the interviewee's trust by being sincere, warm, and personable.

What a high level performer might do:

7. Would expect this interviewer to begin by talking about an interest in common with the interviewee and to ask questions only after the interviewee is talking freely.

6. This interviewer would greet the interviewee warmly, offer him/her a chair, and spend a short time conversing about his/her dingo major. Then he/she would get down to business and start asking questions.

PERFORMANCE EXAMPLES

Average Performance

- Is relaxed and friendly during portions of the interview but also comes on in a very business-like, task oriented way at other times in the session.

- Sets the interviewee somewhat at ease by engaging in small talk at the beginning of the interview or by joking with him/her at appropriate times.

What an average performer might do:

5. Can be expected to laugh freely when the interviewee makes a joke about his/her past experiences.

4. This interviewer would be expected to begin the interview by making small talk about sports after noticing that the interviewee was a college football player.

3. Can expect this interviewer to be somewhat cynical and reserved when the interview begins but to become more relaxed and talkative once into the interview.

PERFORMANCE EXAMPLES

Low Level Performance

- Interacts in a cold and detached manner during the interview, and is generally unresponsive to the interviewee.

- Creates a threatening atmosphere by immediately asking personal questions or by appearing suspicious of interviewees and their credentials.

What a low level performer might do:

2. As soon as the interviewee sits down, would expect this interviewer to begin asking him/her questions about his/her background.

1. This interviewer can be expected to appear detached throughout the interview and not to smile, speak, or nod to the interviewee other than to ask or answer questions.

PERFORMANCE EXAMPLES

EXERCISE TWO

INSTRUCTIONS: This exercise consists of two parts. First, you are asked to look closely at the picture presented to you, and write down as many observations as you feel relevant. Secondly, please write down relevant inferences drawn from the people, setting and objects in the picture.

A. Observations Made

   1.

   2.

   3.

   4.

   5.

   6.

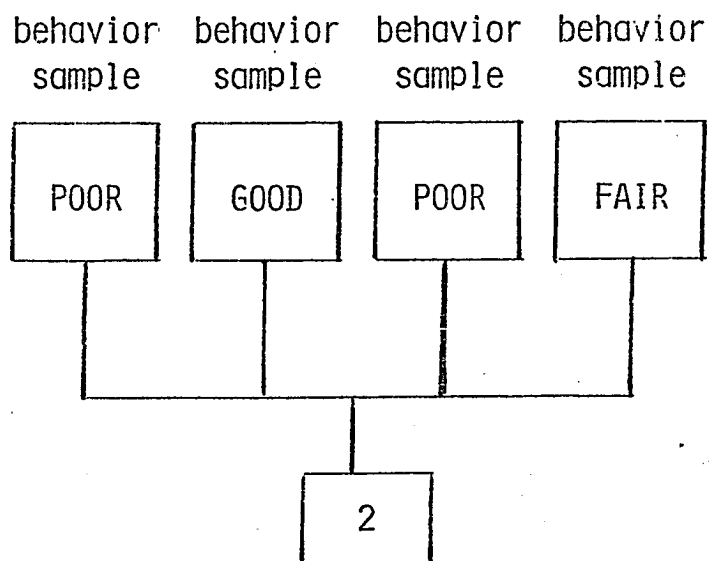B. Inferences Drawn

   1.

   2.

   3.

   4.

   5.

   6.

JUDGMENTAL ERRORS

* Insensitivity to Biased Data

* Inappropriate Causal Inference

* Over-reliance on Previously Formed Theories
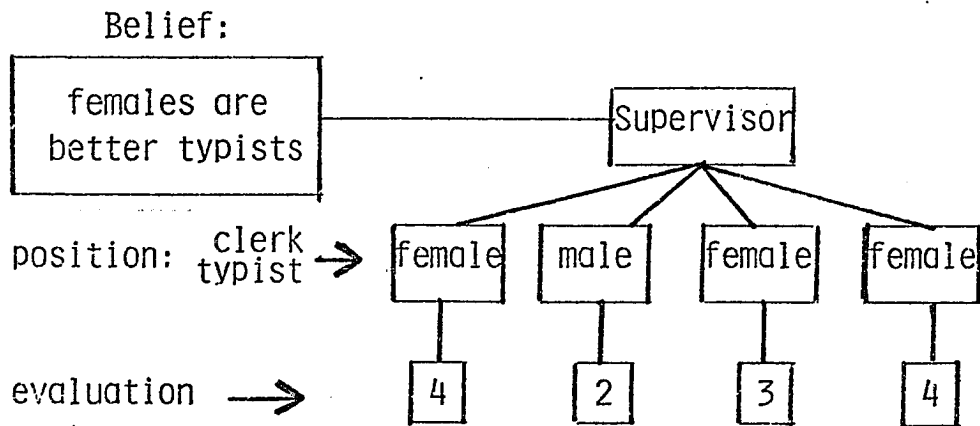
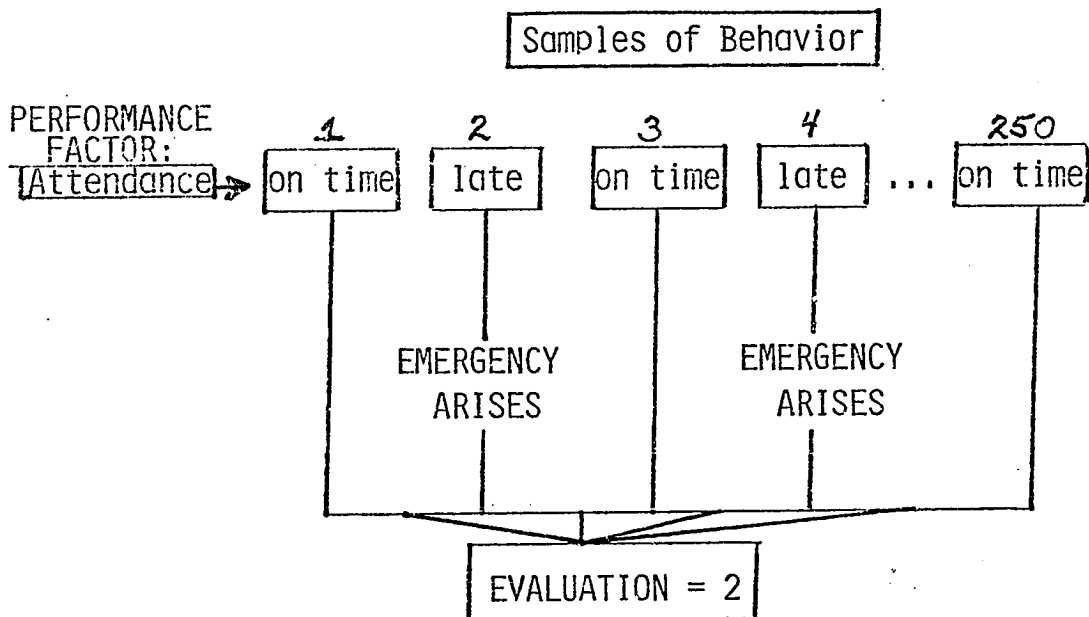* Inappropriate Weighting of Information

Illustration of Biased Data Use

Performance Factor:
Quality of Work

behavior sample	behavior sample	behavior sample	behavior sample
POOR	GOOD	POOR	FAIR

2

OVER-RELIANCE ON PREVIOUSLY FORMED THEORIES

Belief:

| females are better typists |———————— Supervisor

position: clerk typist → | female | | male | | female | | female |

evaluation → | 4 | | 2 | | 3 | | 4 |

ILLUSTRATION OF INAPPROPRIATE WEIGHTING

| Samples of Behavior |

PERFORMANCE
   FACTOR:

Attendance →

1	2	3	4	250
on time	late	on time	late	... on time

EMERGENCY
ARISES

EMERGENCY
ARISES

| EVALUATION = 2 |

APPENDIX D

Schedule of Performance Evaluation Training

PERFORMANCE EVALUATION TRAINING

Training Sessions

Session A

   April 12...8:30 - noon
   April 22...8:30 - noon

Session B

   April 14...8:30 - noon
   April 22...1:00 - 4:30 p.m.

Session C

   April 13...1:00 - 4:30 p.m.
   April 19...8:30 - noon

Session D

   April 15...1:00 - 4:30 p.m.
   April 20...1:00 - 4:30 p.m.

Session E

   April 14...1:00 - 4:30 p.m.
   April 20...8:30 - noon

Session F

   April 15...8:30 - noon
   April 21...1:00 - 4:30 p.m.

Session G

   April 12...1:00 - 4:30 p.m.
   April 19...1:00 - 4:30 p.m.

Session H

   April 13...8:30 - noon
   April 21...8:30 - noon

PERFORMANCE EVALUATION TRAINING

Registration Form

Name: _____   Session: _____

Dept.: _____

Name: _____   Session: _____

Dept.: _____

Name: _____   Session: _____

Dept.: _____

Name: _____   Session: _____

Dept.: _____

If session times are a problem, please call extension 3063
to arrange a different combination of times.

RETURN TO TRAINING & DEVELOPMENT, PERSONNEL OFFICE

APPENDIX E

Old Dominion University Performance Evaluation Form

COMMONWEALTH OF VIRGINIA

CONFIDENTIAL
EMPLOYEE PERFORMANCE EVALUATION

Name_____ Soc. Sec. No._____ Position No._____

Agency Name_____ Sub. Division_____ Agency Code_____

Class Title_____ Class Code_____ Date Entered Present Position____

Date of Evaluation_____

Describe Briefly the Principal Duties in Present Job_____

PART I – PERFORMANCE FACTORS – CIRCLE THE APPROPRIATE PERFORMANCE LEVEL

1- JOB KNOWLEDGE/SKILLS - To what extent does the employee maintain a satisfactory level of job knowledge and/or job skills?   4   3   2   1

Remarks_____

2- QUALITY OF WORK - To what extent does the employee's work meet the required quality standards; i.e., accuracy, neatness and thoroughness?   4   3   2   1

Remarks_____

3- PRODUCTIVITY - To what extent does the employee accomplish the quantity of work expected of the job assignment?   4   3   2   1

Remarks_____

4- RECORD KEEPING/DOCUMENTATION - To what extent does the employee adequately prepare and maintain records, written reports, correspondence, and files?   4   3   2   1

Remarks_____

| | | 4 | 3 | 2 | 1 |

5- __DEPENDABILITY__ - To what extent does the employee perform work without close supervision or assistance?

Remarks_____

_____

_____

| | | 4 | 3 | 2 | 1 |

6- __ADAPTABILITY__ - To what extent does the employee readily adapt to new situations and changes in routines, work load, and/or work assignments?

Remarks_____

_____

_____

| | | 4 | 3 | 2 | 1 |

7- __INITIATIVE__ - To what extent does the employee present new ideas, improve procedures or otherwise demonstrate an awareness of clerical or technical changes related to the job?

Remarks_____

_____

_____

| | | 4 | 3 | 2 | 1 |

8- __ATTENDANCE__ - To what extent does the employee maintain satisfactory attendance performance in regard to tardiness, early departures, and/or absences?

Remarks_____

_____

_____

| | | 4 | 3 | 2 | 1 |

9- __RELATIONS WITH OTHERS__ - To what extent does the employee establish effective working relationships when dealing with supervision, co-workers, and/or the public?

Remarks_____

_____

_____

| | | 4 | 3 | 2 | 1 |

10- __SAFETY__ - To what extent does the employee work in a safe manner and observe safety practices?

Remarks_____

_____

_____

DETERMINING THE OVERALL EVALUATION: ADD the number circled for each performance factor, DIVIDE the total by ten (10) to determine the overall evaluation. Indicate the overall evaluation score by circling, or inserting and circling, the overall evaluation on the scale provided.

Performance Levels		Scale
Employee's performance regularly exceeds the job requirements.	(3.50 & above)	4.00
		3.75
		3.50
Employee's performance meets normal job requirements on a sustained basis.	(2.75 to 3.49)	3.25
		3.00
		2.75
Employee's performance reflects that there is a need for improvement on a sustained basis.	(2.00 to 2.74)	2.50
		2.25
		2.00
Employee's performance fails to meet the job requirements.	(1.99 & below)	1.75
		1.50
		1.25

SUPERVISOR'S COMMENTS CONCERNING THE OVERALL EVALUATION:

PART II - DEVELOPMENTAL TRENDS

1- SIGNIFICANT CHANGES - Indicate any significant changes in performance since the employee's last evaluation.

2- DEVELOPMENT AND TRAINING: (a) Indicate recommendations for further development and training for purposes of preparing the employee for additional responsibilities or for the improvement of current job performance.

(b) Identify any training or developmental activities the employee has completed since his/her last performance evaluation. Such training was (check one) taken as a result of the supervisor's recommendation ___, or the employee's initiative ___.

EVALUATED BY_____ TITLE_____

REVIEWED BY_____ TITLE_____

TO THE EMPLOYEE:
You are requested to sign on the line provided below to indicate only that you have had an opportunity to review and discuss your performance evaluation with your supervisor. YOUR SIGNATURE DOES NOT INDICATE THAT YOU AGREE WITH THE EVALUATION.

EMPLOYEE'S COMMENTS:

EMPLOYEE'S SIGNATURE_____ DATE_____

Form No. 129-00-000

APPENDIX F

Trainee Reaction Questionnaire

REACTIONS TO PERFORMANCE EVALUATION TRAINING

INSTRUCTIONS: Rate each of the following questions, using
the scale provided below. Place the number
which corresponds to your answer in the blank
beside the question.

1   Not at all
2   To a small extent
3   To a moderate extent
4   To a great extent
5   To a very great extent

_____1. To what extent was the performance evaluation
training beneficial to you?

_____2. To what extent was the videotaped <u>lecture</u> portion
of the training beneficial to you?

_____3. To what extent was the <u>practice/discussion</u> portion
of training beneficial to you?

_____4. To what extent did you feel like you were in need
of some formal performance evaluation training?

_____5. To what extent do you believe all new supervisory
personnel should receive formal training in
performance evaluation?

_____6. To what extent do you believe all supervisory
personnel (both new and old) should receive
performance evaluation training?

_____7. To what extent do you believe all supervisory
personnel should receive performance evaluation
"refresher" training on a regular basis?

_____8. How frequently do you believe performance
evaluation "refresher" training should be
conducted?

1   Every six months or less
2   Once a year
3   Once every two years
4   Less than once every two years
5   No performance evaluation "refresher"
training is needed

_____9. Have you ever received formal performance
evaluation training?

1   No, not at all
2   Yes, within the last year
3   Yes, but more than a year ago
4   Yes, but more than two years ago

APPENDIX G

Table of Mean Accuracy Scores for

Laboratory Data

Table G-1

Mean Differential Accuracy Scores Before-and-After Training

		Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5	Dimension 6	Dimension 7
Control	Time 1	0.8740	0.4800	0.6354	0.5016	1.0469	1.1100	1.0082
	Time 2	0.9014	0.2903	0.6584	0.6536	0.6813	0.7069	0.7277
Psycho-metric	Time 1	0.7483	0.2009	0.5930	0.8208	0.5429	0.5331	0.9461
	Time 2	0.6929	0.2223	0.3923	0.2006	0.3006	1.0291	0.2886
Obser-vation	Time 1	0.6195	0.4373	0.7442	0.4192	0.5599	0.9865	1.1847
	Time 2	0.8599	0.5780	0.7178	1.1038	0.9380	0.8426	1.0123
Decision-Making	Time 1	0.8624	0.3965	0.5644	0.7853	0.8004	0.8902	1.3972
	Time 2	0.9249	0.7757	0.8366	0.7931	0.7453	1.2945	1.6141