

20189049



This is to certify that the
dissertation entitled
Performance Appraisal in Context: Motivational
Influences on Performance Ratings
presented by

Margaret Youtz Padgett

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Management


Major professor

Date May 13, 1988



RETURNING MATERIALS:
Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

AUG 14 1997
198

MAY 23 2000

PERFORMANCE APPRAISAL IN CONTEXT: MOTIVATIONAL
INFLUENCES ON PERFORMANCE RATINGS

By

Margaret Youtz Padgett

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Management

1988

ABSTRACT

PERFORMANCE APPRAISAL IN CONTEXT: MOTIVATIONAL INFLUENCES ON PERFORMANCE RATINGS

By

Margaret Youtz Padgett

The purpose of this study was to gain an understanding of some of the determinants of accuracy in performance ratings. Traditional explanations for inaccuracy have focused on rater ability, arguing that better rating formats or more effective rater training should make raters more able to evaluate performance accurately. The central thesis of this study was that an equally important, but generally ignored, determinant of accuracy is the motivation of raters to provide accurate ratings. A causal model detailing the relationship between some conditions likely to reduce the motivation to rate accurately was developed and submitted to empirical test using latent variable structural equation analysis.

One hundred and twenty four managers completed a questionnaire assessing the hypothesized motivational influences and participated in a short interview during which they were asked to provide an honest appraisal of the performance of one employee. This private evaluation was then compared to the most recent public evaluation (obtained from organizational records) to obtain a behavioral index of the extent to which intentional distortion of performance appraisals occurs.

The hypothesized structural model was found to fit the data well (Goodness of Fit Index = .883). Results indicated that the perceived freedom of the rater to be honest and the expected reaction of the

ratee to the appraisal were direct influences on the amount of difference between public and private ratings and thus, the accuracy of ratings. Other important motivational influences included the credibility of the rater to the ratee, the ability of the rater to document the evaluation, and the expected consequences of the appraisal for the ratee. Implications of these results and suggestions for future research are discussed.

Copyright by

MARGARET YOUTZ PADGETT

1988

In loving memory of Sunny,
our family cat and faithful friend for eighteen years

ACKNOWLEDGMENTS

It is impossible to look back over the years I have spent working on my dissertation without recognizing the invaluable assistance of those people who helped make the completion of this project possible. First, I must express my thanks to Dan Ilgen, my chairperson, who provided an immeasurable amount of help and guidance throughout the entire dissertation process. I especially appreciated his speed in providing feedback to me after reading each draft of the proposal and final write-up, particularly toward the end when time was of the essence. Let me also express my gratitude for his constant patience and for those gentle "prods" he provided to keep me working and on target.

I also want to recognize Ken Wexley and John Hollenbeck, my other two committee members. Both Ken and John were always very encouraging and supportive and their substantial contribution helped improve the final product a great deal. Finally, to all three committee members I want to express my appreciation for their guidance and friendship throughout my (many!) years in graduate school. I have learned a great deal from all of them. And even though all memories of graduate school are not positive, mine of these three will always be among my most pleasant.

In addition to my committee, there are several other people who cannot go without recognition. Foremost among these is my husband,

Bob, whose computer and data-analytic skills were absolutely invaluable to me. His frequent help while analyzing my data made this portion of the project go much more smoothly than it would otherwise. It would not be an exaggeration to say that I would probably still be trying to figure out LISREL VI and MTS, Wayne State's computer system, if I hadn't been so fortunate as to have his knowledge and experience at my disposal whenever I needed it. Beyond this, I want to thank him for his constant support, tolerance and love while I was working on my dissertation.

I would also like to recognize my parents, who instilled in me the desire for and value of a good education. They were always in the background cheering me on and encouraging me to keep working even when I was discouraged or it seemed as though little progress was being made. Their frequent encouragement made the long process more bearable.

Last, but by no means least, I want to recognize the 124 managers who contributed their time by participating in this study, for without their help I truly would have had no dissertation.

TABLE OF CONTENTS

List of Tables.....	xi
List of Figures.....	xii
CHAPTER 1: INTRODUCTION.....	1
Statement of the Problem.....	1
Factors Influencing Rater Ability.....	9
The Rating Instrument.....	9
The Roles of Rater and Ratee.....	13
The Rating Process.....	16
The Rating Context.....	18
Conclusion.....	20
Factors Influencing Rater Motivation.....	21
Purpose of the Appraisal and Appraisal Consequences.....	21
Trust in the Appraisal Process.....	28
Conclusion.....	31
CHAPTER 2: MODEL AND HYPOTHESES.....	32
Overview of Model.....	33
Expected Consequences of the Appraisal for the Ratee.....	36
Purpose of the Appraisal.....	36
Reaction of the Ratee to the Appraisal.....	38
Expected Consequences of the Appraisal for the Ratee.....	38
Credibility of the Rater to the Ratee.....	40
Appraisal Visibility.....	41

Appraisal Visibility.....	43
Task Interdependence Among Employees.....	43
Perceived Freedom to be Honest.....	43
Reaction of the Ratee to the Appraisal.....	44
Rater's Desire to be Liked by the Ratee.....	44
Ability to Document the Appraisal.....	45
Appraisal Visibility.....	46
Occurrence of Rendering Errors.....	47
Perceived Freedom to be Honest.....	47
Summary.....	47
CHAPTER 3: METHOD.....	49
Overview of Methodology.....	49
Participants.....	49
Procedure.....	51
The Pilot Study.....	51
The Primary Study.....	53
Variables.....	58
Rendering Errors.....	58
Motivational Influences.....	61
Data Analysis.....	68
Overview of Linear Structural Equation Analysis.....	68
Assessment of Fit.....	69
Assumptions Underlying the Use of Structural Equation Analysis.....	71
Description of Diagram Depicting the Measurement and Structural Models.....	73
CHAPTER 4: RESULTS.....	77
Assessment of the Measurement Model.....	77

Assessment of the Structural Model.....	78
Exploratory Analysis.....	87
CHAPTER 5: DISCUSSION.....	96
Summary and Implications of Findings.....	96
Informal Observations.....	96
Formal Analyses: Supported Hypotheses.....	98
Formal Analyses: Unsupported Hypotheses.....	106
Limitations in the Study.....	111
Suggestions for Future Research.....	114
Conclusion.....	117
APPENDIX A: Evaluation Forms Used by the Organization.....	118
APPENDIX B: Questionnaire Completed by Study Participants.....	121
APPENDIX C: Procedures for Measuring Expected Consequences of the Performance Appraisal for the Ratee.....	134
APPENDIX D: Questionnaire Items Measuring Each Motivational Influence.....	136
Footnotes.....	143
List of References.....	145

LIST OF TABLES

Table 1:	Means, Standard Deviations, Reliabilities and Correlations between Scales Measuring the Motivational Influences.....	79
Table 2:	Factor Loadings for Confirmatory Factor Analysis - The Lambda Matrix.....	80
Table 3:	Intercorrelations between Latent Variables - The Phi Matrix.....	81
Table 4:	Structural Coefficients and T-Values for the Originally Hypothesized Model.....	84
Table 5:	Structural Coefficients and T-Values for the Modified Model.....	91
Table 6:	Goodness of Fit Indices for the Original Model and Sequential Modifications.....	92
Table 7:	Structural Coefficients and T-Values for the Final Model.....	95

LIST OF FIGURES

Figure 1: Factors Influencing the Accuracy of Performance Appraisals.....	4
Figure 2: Landy and Farr's (1980) Component Model of Performance Rating.....	10
Figure 3: Performance Appraisal Behaviors and Possible Outcomes of those Behaviors for Raters and Ratees (from Mohrman and Lawler, 1983).....	24
Figure 4: Model of the Factors Influencing Rater Motivation to Provide Accurate Performance Evaluations.....	34
Figure 5: Hypothesized Measurement and Structural Model of Rater Motivation.....	74
Figure 6: Structural Parameters for Hypothesized Model of Rater Motivation.....	82
Figure 7: Structural Parameters for Modified Model of Rater Motivation.....	90
Figure 8: Structural Parameters for Final Model of Rater Motivation.....	94

CHAPTER 1: INTRODUCTION

Statement of the Problem

The appraisal of human performance has been a concern of researchers for many years as demonstrated by the large number of empirical studies on performance appraisals (cf. Landy & Farr, 1980, for a review of much of this literature). Yet, in spite of the vast amount of research conducted on the appraisal process, it is not clear that much progress has been made toward improving the quality of ratings that result from a typical appraisal system. Understanding and improving the performance evaluation process is particularly important, given the extent to which appraisals are used in organizations. The results of a 1977 survey, for example, revealed that over 90% of those organizations sampled had an appraisal system (Locher & Teel, 1977). Furthermore, in most organizations that have appraisal systems, they are used for purposes that have important implications for employees (Ilgen & Feldman, 1983; Kane & Lawler, 1979). For example, performance appraisals may be used as a basis for promotion and placement decisions, as well as reward allocation and termination decisions. Evaluations may also serve as the criteria against which training and selection programs are validated and be used to provide developmental feedback to employees. Given the number, diversity and importance of the situations utilizing performance appraisal information, it is necessary that this information be as accurate as possible.

In order to understand the evaluation process and some of the factors that can affect appraisal accuracy, it is helpful to examine performance appraisal from a job behavior, or task, perspective.

Researchers examining human performance have argued that effective performance on some task is a function of two factors: a person's ability to perform the task and his/her motivation to do so. The basis of this assumption is Lewin's (1935) interactive model of performance which states that both ability and motivation must be present in order for a person to perform well on some task. In a performance appraisal context, the central task is evaluating the performance of employees. The goal is to obtain ratings that reflect, to the extent possible, the actual behavior of the ratee (Borman, 1978; Bernardin & Pence, 1980). Adapting Lewin's general model of performance to the performance appraisal task suggests that performance rating accuracy is affected by two conditions, a rater's ability to provide accurate ratings of performance and his/her motivation to do so.

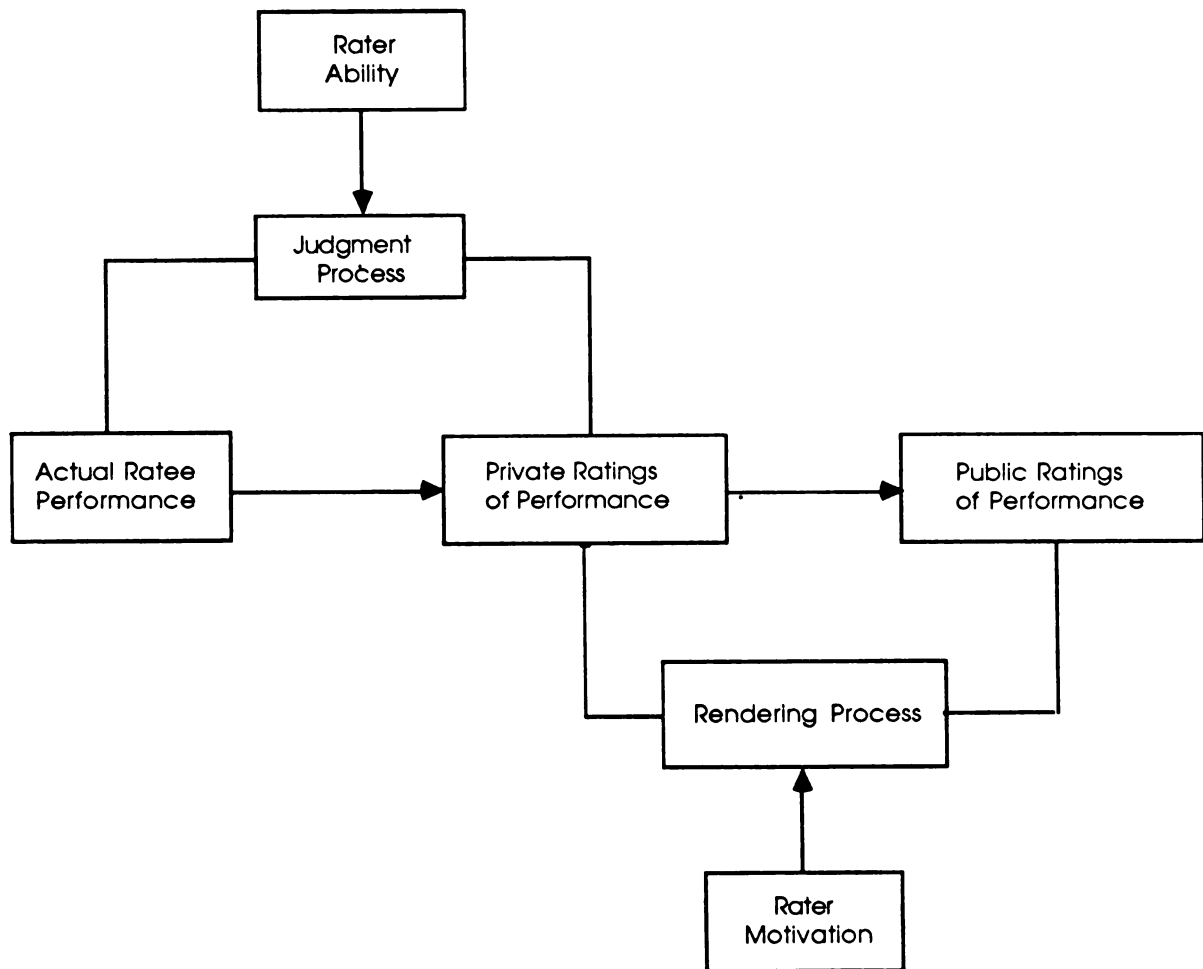
In order to better understand how rater ability and rater motivation influence the accuracy of performance ratings it is necessary to recognize the existence of three potentially distinct views of ratee performance. These are: (1) the actual performance of the ratee, (2) the rater's private evaluation of ratee performance, and (3) the rater's public evaluation of ratee performance. Although the first of these requires little clarification, the distinction between private and public evaluations of ratee performance needs some explanation. According to Mohrman and Lawler (1983), private performance appraisal behaviors include any internal acts of cognition, judgment, perception, evaluation or attribution on the part of raters about some ratee. Private behaviors might also include the making and retention of private notes or other documents about the

ratee. These private rating behaviors, therefore, reflect what raters actually think about the ratee's performance. On the other hand, public performance appraisal behaviors involve verbally communicating appraisals to other people, such as ratees, or recording the appraisal on a form that is seen and used by other people in the organization (Mohrman & Lawler, 1983). Public ratings of performance indicate what the rater wants other people to know about the ratee's performance. The public evaluation is what is typically referred to when the term "performance appraisal" is used.

Based on this distinction, it can be seen that the relationship between actual ratee performance and written ratings of performance (i.e., the extent of appraisal accuracy) really contains two linkages: (1) a linkage between actual ratee performance and rater private judgments about performance and (2) a linkage between rater private judgments about performance and his/her public ratings of performance (see Figure 1). The first linkage has been termed the judgment, or evaluation, process and the second linkage the rating, or rendering, process (Banks & Murphy, 1985). Clearly, both linkages must be strong if performance appraisals (i.e., public ratings) are to be accurate.

The traditional explanation for inaccurate ratings has focused on rater ability. Inaccuracy in performance ratings due to low rater ability is likely to be reflected in a lack of correspondance between actual ratee performance and rater private judgments of performance (linkage 1). Underlying this explanation has been the implicit, but rarely stated, assumption that most inaccuracies in performance ratings occur unintentionally (i.e., without the awareness of the

Figure 1: Factors that Influence the Accuracy of Performance Ratings



rater who, in fact, is trying to rate performance as accurately as possible). In other words, raters are believed to accidentally form inaccurate private judgments about ratee performance through various rating errors and biases (e.g. selecting inappropriate performance information, interpreting this information incorrectly, forgetting relevant aspects of ratee performance etc.). As a result, inaccuracies of this sort are likely to be unsystematic or random (i.e., sometimes resulting in evaluations that are higher than the ratee's actual performance and sometimes leading to lower ratings). Attempts to improve the accuracy of performance appraisals by increasing the ability of raters to evaluate performance (e.g. developing new appraisal instruments, rater training and research on rater cognitive processes) are clearly important since evaluations cannot be accurate if raters lack the requisite ability to appraise performance.

However, the necessity of distinguishing between private and public performance ratings indicates that rater motivation is also an important determinant of appraisal accuracy. Motivation, in general, reflects the level, direction and persistence of behavior (Campbell & Pritchard, 1976). While in a performance appraisal context the level of effort exerted by raters toward actually doing performance evaluations is a relevant concern, even more important is the direction of that motivational force. Specifically, it is important that the motivation of raters be directed toward rating performance accurately rather than toward producing a rating at a particular level. If the rater's objective when doing the performance evaluation is to get the employee a large raise or to avoid an unpleasant

confrontation, then the rater might be motivated to intentionally provide a public rating that he/she believes is inaccurate (i.e., that differs from his/her private evaluation). Thus, low motivation to rate accurately is reflected in intentional discrepancies between public and private ratings, and therefore, is likely to produce systematic biases in performance appraisals (i.e., appraisals that either consistently overstate or understate ratee performance).

Most previous performance appraisal researchers have failed to recognize the potential impact of rater motivation on performance appraisals (Banks & Murphy, 1985). Perhaps one reason for this is the dominant paradigm for research examining performance appraisal accuracy. Specifically, the majority of this research has been conducted in laboratory settings (where standards for determining the accuracy of performance ratings can be developed), particularly since the shift in the last few years toward studying the cognitive processes of raters. While laboratory studies are likely to be helpful in illuminating processes affecting rater ability, they may be less useful in understanding factors influencing rater motivation. This is because laboratory settings may reduce or eliminate the effects of motivational influences, such as personal or political agendas, (Banks & Murphy, 1985), since raters have little to lose by rating accurately or to gain by rating inaccurately. Thus, the motivation to record public ratings that differ from private judgments is likely to be lower.

When performance appraisals are conducted in an organizational setting, however, there is more likely to be a discrepancy between

public and private ratings because of organizational pressures placed on raters to intentionally distort public evaluations of ratee performance. Consider the following situation. Suppose that organizational policies and procedures require that employees be given developmental feedback based on performance appraisal data. Assume also that a particular manager has to provide negative feedback to a poor performing employee whom he or she knows has a tendency to get very defensive and hostile, no matter how constructively the criticism is given. Finally, assume that the manager does not feel that she/he has adequate documentation (i.e., specific examples of ineffective job behavior) to support his/her evaluation.

Several motivational influences are operating in this example. The ratee's anticipated defensiveness, the lack of performance documentation and the way in which appraisal information is used are all likely to affect the extent to which the rater is motivated to provide an accurate rendering of performance. Since these kinds of pressures only exist in real organizations, identification of conditions that affect the motivation to rate accurately requires research conducted in field settings.

The purpose of the present study was to gain an understanding of the rendering process and of some of the reasons for intentional discrepancies between private and public evaluations of performance, termed rendering errors. Although, in theory, low rater motivation could result in public evaluations that either consistently overstate or understate performance, in practice, the former are likely to be more common (see Dayal, 1969; Rowe, 1964; Thayer, 1981). This is because raters' personal goals (e.g. getting an employee a promotion

or making themselves appear favorably to superiors) are more likely to be achieved by inflating, rather than deflating, ratings. In addition, managers have been found to express a great deal of reluctance to intentionally deflate ratings because of the high probability that such an action will lead to subsequent problems (Longenecker, Gioia & Sims, 1987). Therefore, the emphasis in this study was on identifying conditions likely to result in intentionally over-rating performance. Following the suggestion of Bartlett (1983), several motivational influences were identified and a model detailing their interrelationships was tested in a field setting.

Before discussing the motivational issue in greater detail, however, literature examining factors that influence the ability of raters to provide accurate ratings is briefly reviewed. The purpose of this discussion is not to provide an in-depth and critical review of the large volume of empirical research examining ability effects on performance appraisal. Such a review is beyond the scope of this paper and has been conducted by others (e.g. Landy & Farr, 1980; Wexley & Klimoski, 1984). Rather, this overview is intended to demonstrate the pervasiveness of the belief that a lack of rater ability accounts for most of the inaccuracy in performance ratings. In addition, the review introduces the major issues relating to rater ability in order to provide a point of contrast for the major focus of this study, which is the examination of motivational influences on performance ratings.

Factors Influencing Rater Ability

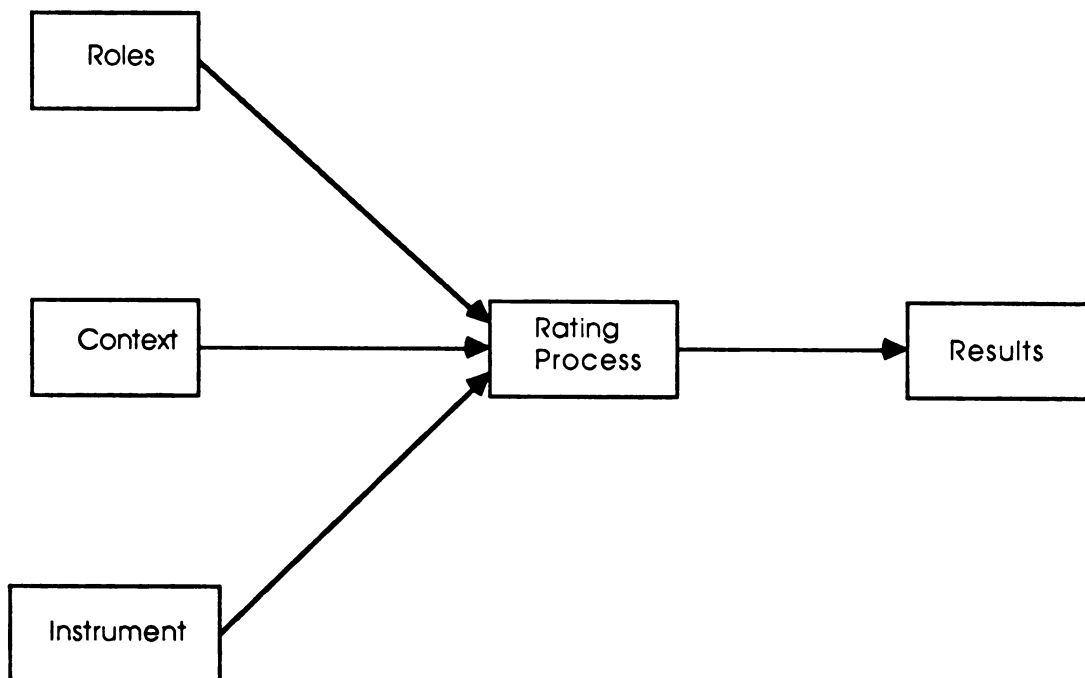
In describing the large body of research examining performance appraisal from the standpoint of rater ability, Landy and Farr (1980) suggest a model that includes several determinants of performance rating results (see Figure 2). These include the vehicle (the rating instrument), the roles (rater and ratee), the rating process and the rating context (e.g. the type of job or organization, the purpose for the appraisal). These components provide a useful structure for briefly reviewing the research dealing with rater ability.

The Rating Instrument

Much of the early performance appraisal research focused on the rating instrument used to record judgments about performance. The assumption behind this research was that how information about a person's performance was elicited (i.e., the design of the form) would influence the ability of raters to make accurate judgments of performance. The different rating formats that have been developed can be distinguished in terms of whether they measure people (i.e., traits), processes (i.e., activities or behaviors) or products (i.e., results) (Wexley & Klimoski, 1984).

The measurement of people typically involves assessing the personal characteristics or traits which they possess. The most pervasive format for measuring traits is the graphic rating scale, introduced by Paterson (1922). This format consists of several rating scales, each associated with a different trait label, a brief definition of the trait and an unbroken line with varying types and numbers of anchors on which the rating is marked. Research on graphic rating scales has involved varying the presence or absence of trait

Figure 2: Landy and Farr's (1980) Component Model of Performance Rating



definitions, the number of divisions in the scale, and the number and type of anchors to see if this affected the quality of performance ratings (e.g. Barrett, Taylor, Parker & Martens, 1958; Madden & Bourdon, 1964).

A second group of rating formats are those which focus on measuring the observable behaviors or activities of employees. The first format of this type was the Behaviorally Anchored Rating Scale (Smith & Kendall, 1963). Behaviorally anchored rating scales (BARS) differ from graphic rating scales in that they utilize behaviorally-oriented anchors for each job dimension, rather than adjectives or numbers. For each dimension, raters indicate which of the behavioral anchors (scaled in terms of effectiveness) is most similar to how they would expect the ratee to behave. A variant of the BARS format is Behavioral Observation Scales (BOS), developed by Latham & Wexley (1977; 1981). Behavioral observation scales require raters to indicate the frequency with which they have observed each of several specific job behaviors relevant to a given performance dimension. Thus, multiple measures are taken of each dimension, rather than just one, as with BARS and graphic rating scales. Other examples of behavioral rating formats are Behavioral Discrimination Scales (Kane & Lawler, 1979), Behavior Summary Scales (Borman, Hough & Dunnette, 1976) and Behavioral Assessment Approaches (Komacki, 1981).

Behaviorally-oriented rating formats offer a number of potential advantages over the traditional graphic rating scale (see Latham & Wexley, 1981 for a more complete discussion of these advantages). For example, behavioral measures are less ambiguous and subjective than are trait measures since they involve actual observations of behavior

rather than abstractions from behavior. In addition, activity measures are more directly related to what the employee actually does and they facilitate providing explicit performance feedback to ratees.

Product, or results, measures are the final type of rating format. The most common results-oriented rating system is Management by Objectives (Drucker, 1954). Management by Objectives involves joint participation by managers and subordinates in the setting of results-oriented goals. Performance evaluation then consists of measuring the extent to which these goals are achieved. The presumed advantage of results-oriented rating systems is that they do not require as much judgment on the part of raters (and thus, bypass their cognitive processes), which should increase the ability of raters to make accurate judgments (Wexley & Klimoski, 1984).

Studies comparing graphic rating scales and BARS have measured rating quality in several ways, including the absence of rating errors, such as halo and leniency, reliability (interrater agreement), discriminability and rater satisfaction with the format. The results of this research are mixed with some studies suggesting that the BARS format may be superior to the graphic rating scale (e.g. Borman & Dunnette, 1975; Burnaska & Hollmann, 1974), and other studies yielding the opposite conclusion (e.g. Bernardin, Alvares & Cranny, 1976). Little research has been conducted comparing results-oriented systems to the other rating formats. Although the practical utility of identifying rating formats that result in more accurate ratings would be substantial, it is not the case that developing better rating formats necessarily eliminates all bias and error in performance

ratings, as early researchers had hoped (Landy & Farr, 1980). Rather, even when carefully developed rating systems are used (whether they are trait, behavior or results systems), some rating bias still seems to occur. Furthermore, Wexley and Klimoski (1984) suggest that the traits vs. behaviors vs. results controversy is not the real issue since each format may be effective in certain situations.

The Roles of Rater and Ratee

The Rater. Research on the rater has been of two types, both oriented toward improving rater ability. Some research has focused on rater personal characteristics, with the primary aim of identifying raters who are more able to provide accurate ratings. A variety of rater characteristics have been examined, including demographic, psychological and job-related attributes (e.g. Borman, 1979b; Taft, 1955; Wexley & Youtz, 1985).

The most frequently examined rater characteristics have been the sex and race of the rater. While the results have been somewhat mixed, there is no consistent evidence that there are sex differences (e.g. Hamner, Kim, Baird & Bigoness, 1974; Jacobson & Effertz, 1974; Rosen & Jerdee, 1973) or race differences (Schmidt & Johnson, 1973) in the quality of evaluations. The primary race-related bias observed consistently is a tendency for raters to give higher performance ratings to ratees of the same race and to be more confident of ratings given to ratees of the same race (e.g. Cox & Krumboltz, 1958; Hamner et al., 1974; Schmidt & Lippin, 1980).

Although rater psychological characteristics would seem to be a fruitful avenue for identifying individuals who are more able to provide accurate performance ratings, psychological characteristics

have been examined too infrequently to allow definite conclusions (see Taft, 1955 and Landy & Farr, 1980 for reviews). Nevertheless, tentative conclusions suggest that more accurate ratings may occur when raters are intelligent, have artistic interests, possess self insight and social skills, and are emotionally adjusted (Borman, 1979b; Taft, 1955). Furthermore, there is some evidence that raters who believe in the variability of people (i.e., who recognize the extent of individual differences) may rate more accurately (Wexley & Youtz, 1985).

The second major type of research on raters has been to examine the quality of ratings from various rater groups. Although the most common source for ratings is the immediate supervisor, other possibilities include peer, self or subordinate ratings. The results of research comparing the quality of ratings from different sources are mixed. While it is evident the ratings obtained from different sources are usually not the same (e.g. Borman, 1974; Kirchner, 1966; Klimoski & London, 1974; Lawler, 1967; Zedeck, Imparato, Krausz & Oleno, 1974), it is not clear that one source is more valid than another. Rather, each rater group appears to have a unique perspective that contributes valid information about performance (Landy & Farr, 1980). This view is consistent with research indicating that different dimensions of job performance are identified by peers and supervisors in the development of BARS for the same job (e.g. Borman, 1974; Landy, Farr, Saal & Freytag, 1976).

The Ratee. Research on the impact of ratee characteristics on performance ratings has been limited almost exclusively to the

examination of ratee demographic characteristics, such as sex and race, on performance ratings (for reviews, see Ford, Kraiger & Schectman, 1986; Kraiger & Ford, 1985, Nieva & Gutek, 1980, and White, Crino & DeSanctis, 1981). Sex of the ratee has been found in some studies to interact with the sex stereotype of the job, such that females in typically male jobs receive lower performance ratings (e.g. Schmitt & Hill, 1977), or lower salaries and less challenging job assignments (e.g. Terborg & Ilgen, 1975). A meta-analysis of ratee race effects showed that black ratees typically receive lower performance ratings than whites but only when evaluated by white raters (Kraiger & Ford, 1985). Several other studies have shown that ratee performance characteristics, such as performance level and performance consistency, may also affect the quality of performance ratings (e.g. DeNisi & Stevens, 1981; Padgett & Ilgen, 1988; Scott & Hamner, 1975).

Overall, research suggests that rater and ratee characteristics may influence the ability of raters to accurately evaluate performance. More research is needed, however, to clarify the mechanisms by which these effects occur. While, from a practical point of view, it is probably not possible to make major changes in the characteristics of raters and ratees which will improve the quality of evaluations, this perspective on rater ability does suggest which people might benefit more from rater training programs designed to eliminate ratings errors (e.g. Bernardin, 1978; Bernardin & Walter, 1977; Latham, Wexley & Pursell, 1975) or improve accuracy (Pulakos, 1984).

The Rating Process

The newest emphasis for research on performance appraisal has been to examine the cognitive processes of raters when making performance evaluations (DeNisi, Cafferty & Meglino, 1984; Feldman, 1981; Ilgen & Feldman, 1983). This approach views the rater as an active information processor involved in selecting information about ratee performance, organizing and storing this information in memory and then, at some later time, recalling the information in order to complete the evaluation form. Although this approach focuses primarily on understanding the rating process, one outcome of this research, from the perspective of rater accuracy training, may be the identification of more effective strategies for gathering, organizing and retrieving information about ratee performance. It may then be possible to teach raters these strategies so that they are more able to rate performance accurately.

Thus far, more theorizing on rater cognitive processes has occurred than actual research and much of the theorizing has tended to emphasize the general relevance of findings in the area of social cognition for performance appraisal than specific applications of this literature to the performance appraisal process (DeNisi et al., 1984). An exception to this tendency is the large body of research examining attribution processes (e.g. Kelley, 1967; Weiner, Frieze, Kukla, Reed, Rest, Rosenbaum, 1971), the effect of attributions on performance evaluations (e.g. Knowlton & Mitchel, 1980; Mitchell & Wood, 1980; Nieva & Gutek, 1980) and the effect of attributions on the distribution of organizational rewards (e.g. Heilman & Guzzo, 1978).

While attributional processes are important cognitive

determinants of the ability of raters to provide accurate performance ratings, it has been argued that research on cognitive processes needs to go beyond attribution theory to examine how the selection, organization, storage and retrieval of performance information affects the accuracy of appraisals (Feldman, 1981; Ilgen & Feldman, 1983). DeNisi, Cafferty & Meglino (1984) provided a model and a number of specific propositions to guide research in this area. Among the more interesting examples of research from this perspective are studies examining (1) factors that influence the selection, organization and recall of performance information, such as appraisal purpose (e.g. Williams et al., 1985), affect (e.g. Bower, 1981; Cardy & Dobbins, 1986; Park, Sims & Motowidlo, 1986), and categorization (e.g. Favero & Ilgen, 1983; Lord, Foti & Phillips, 1982; Murphy & Balzer, 1986; Padgett & Ilgen, 1988), (2) how performance information is processed and its effect on recall (e.g. DeNisi, Williams, Cafferty & Meglino, 1985; Lance & Woehr, 1986; Murphy, Martin & Garcia, 1982; Nathan & Lord, 1983) and (3) the effect of rater cognitive processes on traditional measures of rating quality, such as rating accuracy and the occurrence of rating errors (e.g. Cafferty, DeNisi & Williams, 1984; Favero & Ilgen, 1983; Mount & Thompson, 1987).

Overall, a cognitive processing perspective on performance appraisal seems to offer a number of potential practical applications for training raters how to rate evaluate performance more accurately. However, as noted by DeNisi and his colleagues (DeNisi, Williams, Cafferty & Meglino, 1985), a great deal more research is before it can be concluded that this perspective is more useful than other

approaches to studying performance appraisal and before specific applications can be developed.

The Rating Context

According to Landy & Farr (1980), the rating context consists of those factors not specifically related to the instrument, rater, ratee, or rating process that are still part of the situation surrounding the appraisal and thus, could affect its accuracy. The most frequently cited contextual factor affecting performance ratings is the purpose of the appraisal. It appears, however, that appraisal purpose can influence performance ratings both through its effect on rater ability (e.g. Crockett, Mahood & Press, 1975; Jeffrey & Mischel, 1979; Williams, DeNisi, Blencoe & Cafferty, 1985; Wyer, Srull, Gordon & Hartwick, 1982) and rater motivation (e.g. Bernardin, Orban & Carlyle, 1981; McIntyre, Smith & Hassett, 1984; Meyer, Kay & French, 1965; Sharon & Bartlett, 1969; Zedeck & Cascio, 1982). Only research on how appraisal purpose influences rater ability (i.e., results in unintentional inaccuracies) is discussed here; that dealing with rater motivation and intentional rating distortions is described later.

Performance appraisals can be used for a variety of purposes in organizations, but the two most commonly mentioned purposes are control and coaching. Performance appraisals are used for control when they help to determine rewards and punishments for employees (e.g. salary, promotion, demotion, transfer and termination decisions). As noted by Ilgen & Feldman (1983), the control purpose of appraisals can either be explicit, as when appraisals are directly tied to rewards via a merit pay system, or implicit, such as when a superior determines job assignments for employees based on his or her

impression of their performance. The coaching function involves providing employees with feedback on their performance in order to facilitate performance improvement and development.

DeNisi et al. (1984) suggested that appraisal purpose is most likely to influence rater ability to provide accurate ratings of performance through its effect on the amount and type of information sought by raters and the way that information is stored in memory. For example, some research suggests that raters seek out more information when appraisals are done for administrative decision-making than when they are done for employee development (Matte, 1982). In addition, raters have been found to select distinctiveness information when appraisals are used for salary decisions but sought out consensus information when the appraisal would influence promotion or remedial training decisions (Williams, DeNisi, Blencoe & Cafferty, 1985) (see DeNisi, Cafferty & Meglino, 1984 for a more complete discussion of this issue).

A second contextual factor that could influence rater ability to provide accurate performance evaluations is characteristics of the employee's task and workgroup. For example, several studies have suggested that performance appraisals tend to be done on a relative, rather than absolute, basis (e.g. Grey & Kipnis, 1976; Knowlton & Mitchell, 1980; Mitchell & Liden, 1982). Furthermore, the amount of task interdependence between members of a workgroup could affect performance ratings. As noted by Kane & Lawler (1979), when the tasks performed by members of a workgroup are interdependent, it is more difficult to evaluate performance because of the difficulty in

determining the contribution of any given individual. The result is likely to be less variance across the performance ratings for the members of the workgroup (as found by Liden & Mitchell, 1983) and possibly lower rating accuracy.

A final contextual factor that could affect the ability of raters to accurately evaluate performance is the opportunity of the rater to observe the performance of the employee (Kane & Lawler, 1979). The less opportunity the rater has to observe relevant job behaviors the more difficult it is to develop an accurate picture of an employee's performance (e.g. Henemen & Wexley, 1983). To some extent, the opportunity to observe is determined by the nature of the employee's job, since some jobs (e.g. sales representatives) require employees to spend a significant amount of time in locations where their behavior cannot be observed by the rater.

Overall, contextual factors represent an important but relatively unexplored influence on the ability of raters to provide accurate evaluations of performance. More research is needed to further elaborate the effect of these and other contextual factors on performance ratings.

Conclusion

In the section above, research dealing with the effect of the rating instrument, the rater and ratee, the rating process and the rating context on the ability of raters to provide accurate performance evaluations was briefly reviewed. This review highlights the pervasiveness of the assumption that inaccuracies typically result from a lack of rater ability. It also demonstrates the enormous complexity of the rating process and the extreme difficulty of

obtaining accurate ratings even when only factors influencing rater ability are considered. Unfortunately, even if the ideal rating instrument could be developed, the best raters selected and trained in the most effective strategies for selecting, organizing and retrieving performance information, and the most effective rating context achieved, it is doubtful that accurate performance ratings would result. Only when the issue of rater motivation is also considered is the goal of accurate performance ratings likely to be realized.

Factors Influencing Rater Motivation

The influence of rater motivation on performance ratings has received little attention, compared to the volumes of research examining issues related to rater ability. However, as disillusionment with typical methods of improving rating accuracy (e.g. developing new instruments or training raters to eliminate rating errors and bias) has increased, there has been a greater realization of the importance of rater motivation (e.g. Banks & Murphy, 1985). The most frequently mentioned motivational influences discussed in the literature are described below.

Purpose of the Appraisal and Appraisal Consequences

A number of researchers have recognized the potential influence of appraisal purpose on the motivation of raters to evaluate performance accurately (e.g. Decotiis & Petit, 1978; Kane & Lawler, 1979; Sharon & Bartlett, 1969; Zedeck & Cascio, 1982), as distinct from its effect on rater ability (discussed above). The research on appraisal purpose comprises the majority of the empirical research done to date which is relevant to understanding rater motivation.

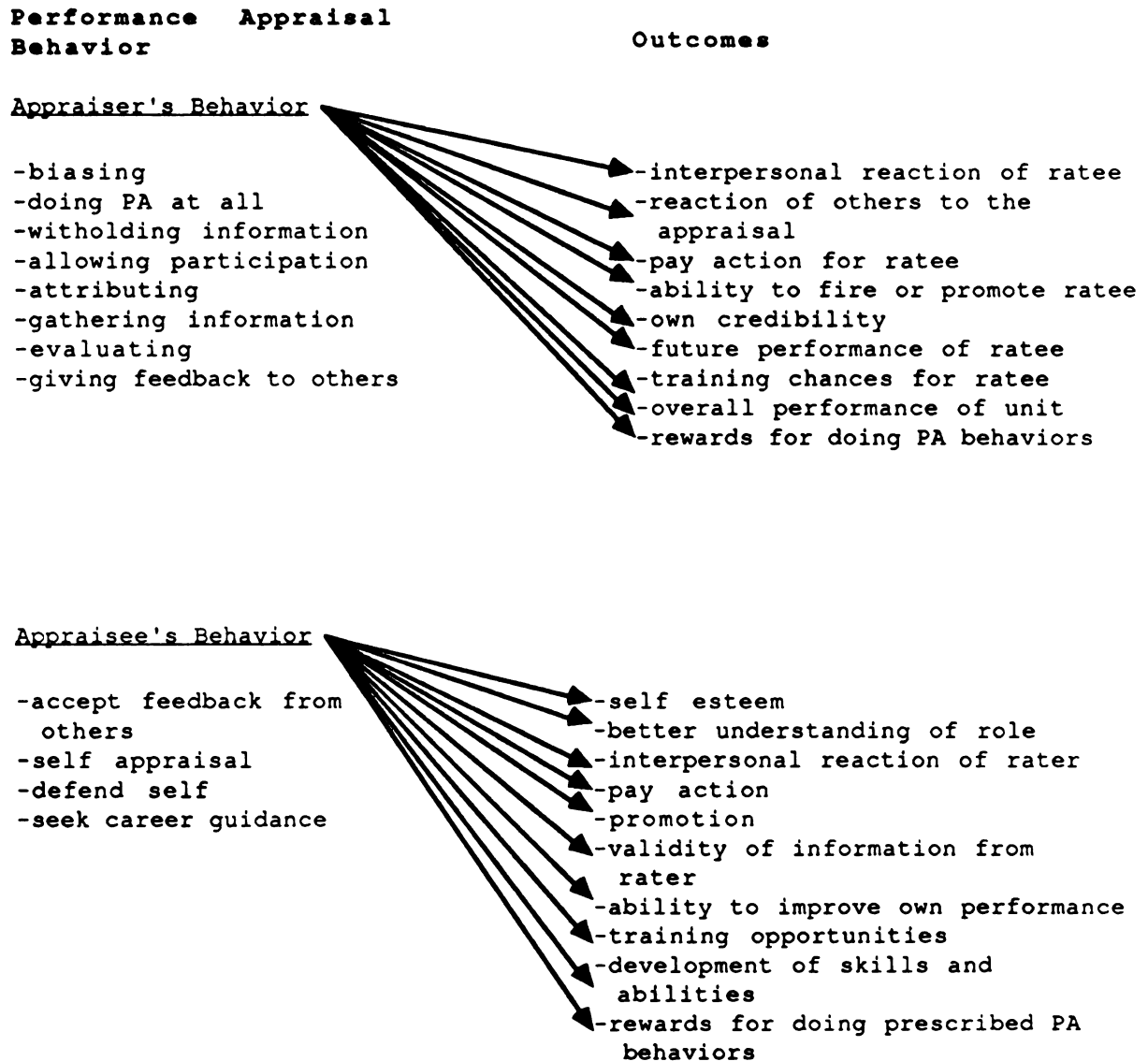
Although research results are somewhat inconsistent, the most common finding is that ratings for research purposes are less lenient and more accurate than ratings for personnel decisions (Sharon & Bartlett, 1969). Within the category of personnel decisions, ratings are less lenient and more accurate when they are used for subordinate development than for salary, promotion or termination decisions (Bernardin et al., 1981; Meyer, Kay & French, 1965; Zedeck & Cascio, 1982).

Unfortunately, most of the research on appraisal purpose has been limited to examining its effect on either rating accuracy or the occurrence of rating errors. Few attempts have been made to understand how and why purpose influences performance ratings. An exception is the DeCotiis & Petit (1978) model of performance appraisal which went beyond simply noting that purpose influenced performance ratings to describe why this relationship might occur. Specifically, they argued that appraisal purpose has an important motivational component because of its inextricable linkage to the consequences of the appraisal for the rater and ratee. The importance of appraisal consequences for rater motivation can be derived from an expectancy theory framework (Mitchell, 1974; Porter & Lawler, 1968; Vroom, 1964). This perspective has been specifically applied to performance appraisals by Mohrman & Lawler (1983) in an attempt to understand the motivations of both raters and ratees in an appraisal situation. However, because the focus of this study was on understanding the actions of the rater, only this aspect of performance appraisal motivation is considered in the discussion below.

According to expectancy theory, an individual's motivation to exert effort toward some behavior is a function of three cognitions, the expectation that effort will result in the desired behavior, the perceived outcomes of those behaviors, and the attractiveness of those outcomes (Porter & Lawler, 1968). In a performance appraisal situation the relevant behavior is doing an accurate performance evaluation. Therefore, it follows that a rater's motivation to evaluate performance accurately should be influenced by the extent to which the rater believes he/she is able to evaluate performance accurately, the perceived consequences of doing an accurate appraisal and the attractiveness of those consequences. The latter two cognitions have the greatest relevance for this discussion. Important appraisal consequences for the rater include both what happens to the rater directly as a result of the evaluation and what happens to the ratee because of the evaluation. Ratee consequences (e.g. the size of salary increases, the likelihood of promotion, effects on self-esteem) represent important concerns for the rater because of the potential ramifications for his/her day-to-day interactions and future relationship with the subordinate (Dayal, 1969; Decotiis & Petit, 1978; McCall & DeVries, 1977).

Some of the possible consequences of appraisals that might occur for raters and ratees are described in Figure 3. While some of the outcomes for the desired appraisal behaviors are positive, there is the potential for many negative outcomes to result from accurately recording performance evaluations as well. In fact, it could be argued that many more of the probable consequences for the rater will

Figure 3: Performance Appraisal Behaviors and Possible Outcomes of these Behaviors for Raters and Ratees (from Mohrman and Lawler, 1983)



be negative (e.g. getting an undesirable pay or promotion action for the employee, having the employee react defensively to the appraisal, having superiors in the organization reject the appraisal, damaging his/her relationship with the employee etc.). Thus, it should not be surprising that, in many cases, the motivation to provide accurate appraisals is low.

There is also some empirical evidence that the anticipated consequences of the appraisal for the rater and ratee are important influences on the public performance ratings given by raters. In the only empirical study specifically examining motivational issues in performance appraisal (Longenecker, Gioia & Sims, 1987), the evaluation process was viewed as a political process where actors (i.e., raters) were motivated to enhance or protect their own interests. This study involved in-depth semi-structured interviews with 60 executives employed in seven large organizations. The methodology employed was primarily inductive in that no hypotheses were tested (although some a priori "probes" were used during the interview). Rather, executives were encouraged to freely and subjectively describe how they perceived their performance evaluation processes. All interviews were tape-recorded and then transcribed onto notecards, with each card containing one directly quoted idea or thought from one executive. Notecards were then classified into categories representing the various political/motivational issues that emerged during the interviews. Only issues that were raised by 72% or more of the executives were reported.

Perhaps the most important finding from this study was the open recognition and admission by managers that performance appraisal was a

political process and that it was not uncommon for them to intentionally modify their performance ratings of an employee (most typically by inflating the rating, but, in a few circumstances, by deflating it) if this resulted in more positive outcomes for either the employee or themselves. Some of the reasons given by the managers interviewed in the study for intentionally inflating performance ratings included: (1) a desire to maximize the merit increases an employee would be able to receive; (2) to protect or encourage an employee whose performance was suffering for personal reasons; (3) to avoid letting people outside the department know about problems within the department; (4) to avoid creating a written record of poor performance that would become a permanent part of the employee's personnel file; (5) to avoid confronting a problem employee; (6) to give an employee whose performance had improved a break; and (7) to promote out of the department an employee who was a trouble-maker or who didn't fit in.

On the other side of the coin, consciously deflating performance ratings, while much less frequent, occurred when the manager wanted to: (1) shock an employee back to high performance; (2) teach a rebellious employee who was in charge; (3) let an employee know that they should consider leaving the organization; and (4) begin to build a documented case that would facilitate the process of terminating the employee.

The study by Longenecker and his colleagues discussed above represents a significant first step in identifying some of the important influences on rater motivation. The study employed real

managers from a wide variety of different organizations (although currently employed in only seven different companies, collectively they had been involved in the appraisal processes from 197 organizations over the course of their careers), and thus, has a high degree of the external validity. On the other hand, the study suffers from several limitations. First, although having managers freely describe their own evaluation process reduces the potential for priming effects, where the questions asked during the interview create feelings or opinions not otherwise present (Salancik & Pfeffer, 1978), this methodology makes the data inherently more subjective and less rigorous since no a priori hypotheses could be tested. Second, the study contained no actual behavioral measure of rating inflation or deflation except the verbal reports of the managers who were interviewed. Thus, there is no direct evidence that rating distortion, such as that reported by the managers, actually occurred, nor is there information about the magnitude of the distortions. Finally, while the study does demonstrate the pervasiveness of motivational influences on performance appraisals, it consists primarily of a listing of potential motivational constructs. No attempt is made to develop these constructs into an integrated model of rater motivation. As a result, it is less helpful in directing future research on the motivation to rate performance accurately. The study described in this paper attempts to eliminate some of these deficiencies.

One specific potential negative consequence of the appraisal that has received some attention in the literature concerns the extent to which raters feel able to confront employees about their performance

(Bernardin & Beatty, 1984; Bernardin & Buckley, 1981; Dayal, 1969).

Research suggests that raters are more likely to provide lenient ratings when they expect to have to provide employees with feedback on their performance (Fisher, 1979; Sharon & Barlett, 1969). While no research on the psychological processes mediating this relationship exists, a likely explanation is that having to openly discuss and justify their evaluations with employees is an unpleasant and difficult situation that many raters would prefer to avoid. Since inflating performance ratings is one way to avoid a confrontation of this nature, particularly with those employees not performing at the highest level where some negative feedback is required, this practice is not surprising.

To deal with this problem, Bernardin and his colleagues (e.g. Bernardin & Buckley, 1981) suggest training raters on how to be critical. Utilizing a social learning perspective (Bandura, 1977), they argue that training should focus on increasing a rater's efficacy expectations, or the belief that he/she can successfully execute some behavior, in this case, meeting with the employee and discussing the performance evaluation.

Trust in the Appraisal Process

The final motivational issue that has been discussed in the literature is the rater's trust in the appraisal process (Bernardin & Beatty, 1984). Trust in the appraisal process is defined as, "the extent to which both raters and ratees perceive that the appraisal data will be (or has been) rated accurately and the extent to which they perceive that the appraisal data will be (or has been) used

fairly and objectively for pertinent personnel decisions" (Bernardin & Beatty, 1984, p. 268). Although to a large degree, trust in the appraisal process may be indicative of the overall organizational climate, it reflects more specifically the organizational climate with regard to performance evaluations.

Trust in the appraisal process is important because it seems to correlate with the degree of leniency in ratings (Bernardin, Orban & Carlyle, 1981). In the Bernardin, Orban and Carlyle (1981) study, performance appraisal systems were going to be developed for two agencies, neither of which had been doing performance evaluations for several years. In one agency, the appraisal was only to be used for employee feedback and development while in the other agency, it was to be used for both employee development and administrative decision-making (e.g. promotion and salary decisions). Before actually implementing the new appraisal systems, managers in both agencies completed a questionnaire designed to measure their expected trust in the appraisal process (called the TAPS questionnaire). Then confidential practice performance ratings were collected and their actual trust in the appraisal process was assessed. A week later performance ratings were collected again.

Several interesting findings emerged. First, from time 1 to time 2, trust in the appraisal process decreased while performance ratings increased (relative to the initial rating) in both agencies. It is interesting to note that both changes were greater in the agency that intended to use appraisals for administrative decision-making. Second, in both agencies a negative correlation was found between trust in the appraisal process and scores on the rating scale,

indicating that as trust decreased, ratings became more lenient. Again, the relationship was stronger in the agency where appraisals were also used for administrative purposes which suggests that appraisal purpose may moderate the relationship between trust in the appraisal process and rating level.

Although this study makes a contribution to our understanding of performance rating processes by suggesting that trust in the appraisal process may be an important determinant of rating level, it has several weaknesses which complicate interpretation of the results. First, the initial set of performance ratings collected were practice ratings that were kept confidential from everyone in the organization (i.e., they were for research purposes only), while the second set of ratings were not confidential. Previous research indicates that ratings for research purposes only are less lenient than ratings that are used by the organization in some way (e.g. Sharon & Bartlett, 1969). As a result, the increase in leniency observed could have resulted from the change in how ratings were used rather than from changes in the amount of trust in the appraisal process, as suggested by the authors. In addition, the study treats the relationship between trust and performance ratings as a "black box", in that it does not identify the psychological mediators of this relationship or how trust impacts rater motivation.

To some extent, issues similar to trust in the appraisal process appeared in the study by Longenecker and his colleagues (Longenecker, Gioia and Sims, 1987) described above, providing further support for the importance of this construct. For example, managers reported

greater likelihood of political behavior (and therefore, probably less trust in the appraisal process) in situations where top management did not take the appraisal process seriously and only gave "lip-service" to its importance. It was also more likely when upper managers themselves used political tactics when appraising the performance of their subordinates. Finally, political behavior occurred more frequently when there was a lack of openness and trust between managers and employees about performance appraisal and when raters did not personally value performance appraisal as a tool for helping employees grow and develop.

Conclusion

While very little empirical research has been done examining motivational issues in performance appraisal, the small amount that has occurred suggests that motivational influences are likely to have a significant effect on performance ratings in actual organizational settings. Unfortunately, the research on rater motivation so far is fragmented and consists of little more than a listing of some of the organizational and individual level conditions that might reduce rater motivation to record accurate evaluations. What is needed is a more theoretical approach to examining rater motivation that will direct future research on and facilitate understanding of this important influence on performance ratings. One step toward achieving this goal would involve the development of a causal model detailing the way in which these motivational conditions are related to one another. In subsequent sections of this paper, such a model is described and submitted to empirical test.

CHAPTER 2: MODEL AND HYPOTHESES

In this section, a model detailing the relationships between a number of potential motivational influences is presented. Before describing the model, however, several introductory comments should be made. First, this model is not intended to be an exhaustive description of all the constructs relevant to understanding rater motivation. Such an undertaking is beyond the scope of a single study. As a result, a subset of potentially interrelated motivational influences was selected in order to explain a part of the complicated process by which raters are motivated to provide accurate or inaccurate ratings of performance.

Secondly, the model presented is a cognitively based model, in that it relies primarily on a rater's cognitions or beliefs about the performance appraisal situation to explain his/her actions in that situation. This idea is based on the work of a number of theorists in psychology (e.g. Markus & Zajonc, 1985) and sociology (e.g. Ball, 1972; Berger & Luckman, 1966; Silverman, 1971; Thomas, 1928), who have argued that an individual's actions are determined by his/her definition of the situation. According to Ball (1972), an individual's definition of the situation may be seen as "the sum of all recognized information, from the point-of-view of the actor, which is relevant to his locating himself and others, so that he can engage in self-determined lines of action and interaction" (p. 63). The definition of the situation is important because it means that in order to understand social behavior one must look to the meanings of situations as they are experienced by the actors within those situations, rather than to "objective reality" (if, in fact, such a

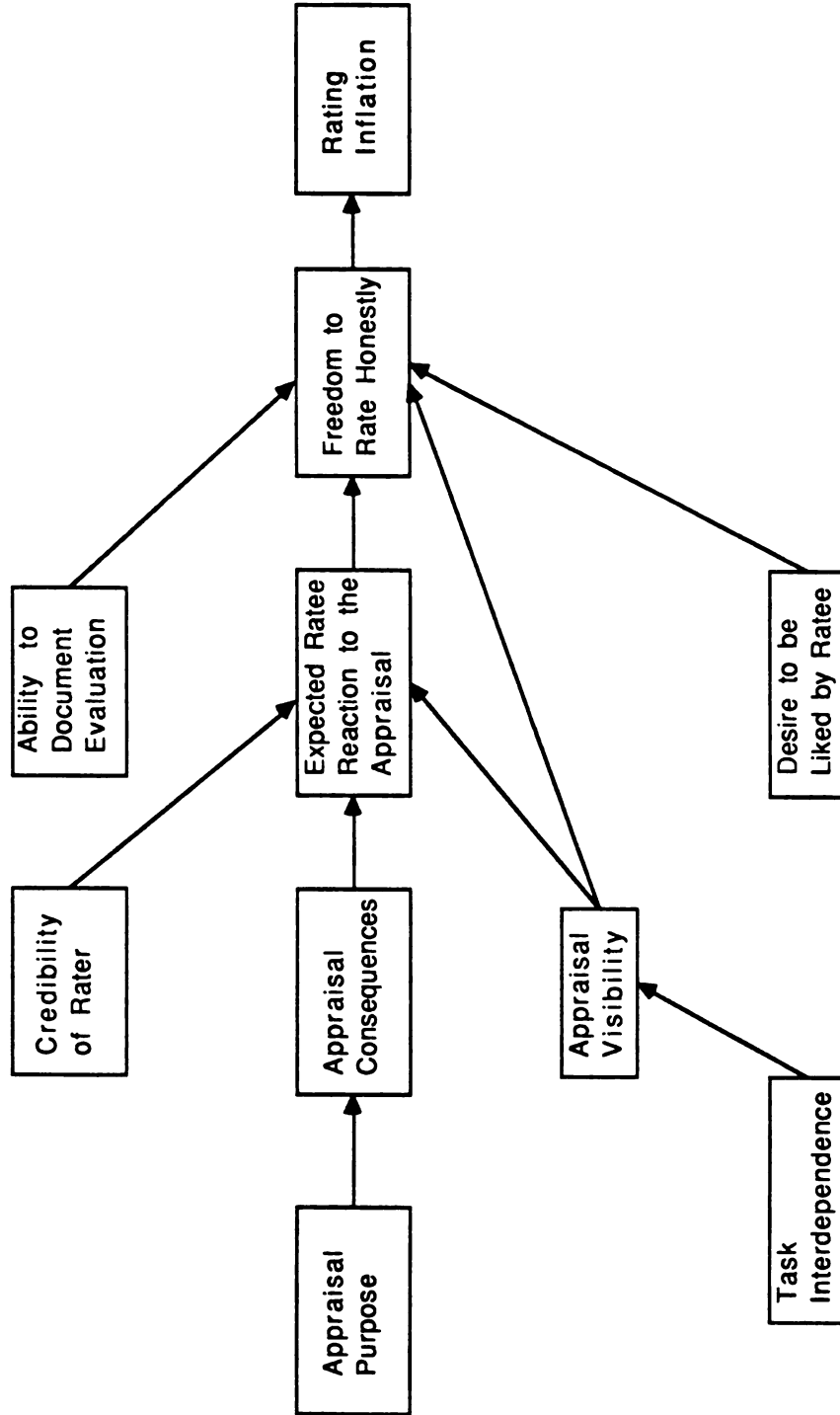
thing really exists) since the former determines how an individual will behave.

Within a performance appraisal context, this suggests that the rater's definition of the performance appraisal situation (i.e. his/her perceptions or cognitions), rather than objective reality, determines whether or not he/she will be motivated to provide an accurate evaluation of the performance of employees. For example, if a manager believes that an employee is likely to react defensively to a negative performance appraisal, this will have implications for his/her motivation to rate accurately and actual rating behaviors. Whether or not the employee would, in fact, react defensively is irrelevant in determining the subsequent actions of the manager. As Thomas (1928) put it, "if men define situations as real, then they are real in their consequences" (p. 572). Because the model presented here is from the perspective of the rater's definition of the performance appraisal situation (a perspective similar to that adopted by Mohrman & Lawler, 1983), all of the constructs in the model involve the rater's perceptions about various elements of the performance appraisal context.

Overview of the Model

The model presented in Figure 4 is an attempt to incorporate some of the motivational influences described by previous researchers as well as some additional ones into an integrated picture of the processes affecting rater motivation. Before describing the rationale for specific linkages in the model, a brief summary of the entire model is presented.

Figure 4: Model of the Factors Influencing Rater Motivation to Provide Accurate Performance Evaluations



One important exogeneous variable in the model is the purpose of the performance appraisal, conceptualized as the extent to which appraisals are used for employee development and/or for salary, promotion and termination decisions. Appraisal purpose is important because it influences raters' perceived consequences of the appraisal for the ratee (e.g. salary or promotion decisions, self esteem). The magnitude and direction of these consequences is hypothesized to mediate the relationship between appraisal purpose and the anticipated employee reaction to the appraisal (e.g. accepts the appraisal, becomes angry and defensive etc.). The more negative the expected consequences of the appraisal for the ratee, the more negatively the ratee is expected to react to the appraisal.

The anticipated reaction of the ratee to the performance appraisal is also likely to be influenced by the extent to which raters believe they are credible to the ratee as a feedback source. The more credible raters feel they are to the ratee, the more likely they are to expect the ratee to accept the appraisal, regardless of its sign and consequences. Finally, the ratee's expected reaction should be affected by the perceived visibility of performance appraisals to coworkers. When raters believe that members of their workgroup will find out how coworkers were evaluated, then they are also likely to expect a more negative employee reaction to the appraisal due to the potential for comparisons and felt inequities. The model suggests that appraisal visibility is enhanced when task interdependence among ratees in the workgroup is greater because of the greater number of opportunities for interaction and discussion among members of the workgroup.

The anticipated reaction of the ratee to the appraisal is expected to influence raters' perceived freedom to be honest and objective when evaluating performance. The more negative the expected reaction, the less raters should feel free to be honest. The perceived freedom to be honest is also hypothesized to be affected by the visibility of performance appraisals to ratees. The more that raters expect employees to find out how coworkers were evaluated, the less freedom they are likely to feel to be honest when doing performance evaluations. Raters may feel unable to allow true performance differences to show up in performance ratings since this might result in conflict among employees in the workgroup. Freedom to be honest should also be lower when raters have a strong desire to be liked by the ratee and when they do not feel they have adequate documentation to support their evaluation.

Finally, raters' perceived freedom to be honest and objective when evaluating ratee performance is hypothesized to be positively related to the occurrence of rendering errors (i.e., differences between public and private evaluations of performance). In the sections which follow, the specific motivational influences and the linkages between them will be described in greater detail. Each endogeneous (i.e., dependent) variable, and its hypothesized causes, is discussed separately.

Expected Consequences of the Appraisal for the Ratee Purpose of the Appraisal

The results of research done so far on appraisal purpose (e.g. Bernardin, Orban & Carlyle, 1981; McIntyre, Smith & Hassett, 1984;

Meyer, Kay & French, 1965; Sharon & Bartlett, 1969; Williams, DeNisi, Blencoe & Cafferty, 1985; Zedeck & Cascio, 1982) are fairly clear in demonstrating that the way in which appraisal information is used does affect performance ratings, probably through its effect on rater ability and rater motivation. However, the mechanism by which appraisal purpose affects performance ratings has not been specified. The effect of appraisal purpose in the model described here is hypothesized to be a motivational one.

One reason that appraisal purpose might be important from a motivational point of view is that it influences the expected consequences of the appraisal both for raters and ratees (Bartlett, 1983; DeCotiis & Petit, 1987). While both rater and ratee consequences are likely to be important influences on rater motivation, in this model, more emphasis is placed on the consequences of the appraisal that raters expect for ratees. When appraisals are used in the organization for either controlling or coaching purposes, this should influence the attractiveness of potential consequences that raters might expect to result to ratees from the performance appraisal. For example, when appraisals are done for research purposes or when they are completed but serve little purpose in the organization (i.e., they are filed and forgotten), they have few consequences of any significance for ratees. In these situations, raters are likely to perceive little need to distort public ratings of performance, which is likely to account for the lower leniency and greater accuracy typically found when ratings are done for research purposes (e.g. Sharon & Bartlett, 1969).

On the other hand, performance ratings that are used in the

organization, either for coaching or controlling purposes, do have important consequences for employees. The most obvious and, perhaps, most significant, consequences of appraisals occur when they are used for purposes of control because then they affect ratees' salaries as well as their likelihood of being promoted, demoted or transferred. Appraisals may also affect whether employees are given opportunities for training. When performance appraisals are used for developmental purposes (i.e., to provide developmental feedback to employees), appraisal outcomes might include such things as development of skills and abilities, a better understanding of the job, the ability to improve their own performance and increased (or decreased, depending on the sign of the feedback) self esteem (Mohrman & Lawler, 1983). Thus, the way in which appraisals are used in an organization implies the existence of potential consequences, each of which will have some valence, either positive or negative, to the ratee. This suggests the following hypothesis:

- H1: The purpose that performance appraisals serve in the organization will be significantly related to the overall attractiveness of the consequences of the appraisal for the ratee.

Reaction of the Ratee to the Appraisal

Expected Consequences of the Appraisal for the Ratee

The overall attractiveness of the appraisal consequences to the ratee is hypothesized to influence his/her reaction to the appraisal. Consistent with the cognitive and perceptual nature of the model, it is suggested that raters attempt to estimate the overall valence of the consequences likely to accrue to a ratee from the appraisal and use this information to predict how he/she will react to the

evaluation. The more negative the overall valence of these consequences, the more likely that raters will expect the ratee to reject the appraisal and become defensive and angry in reaction to it. Since many raters are likely to feel uncomfortable about confronting a potentially hostile employee and furthermore, may lack confidence in their ability to cope with this reaction (Bernardin & Buckley, 1981), they may try to avoid the situation by inflating public ratings of performance. In this way, the ratee is less likely to react defensively to the appraisal (a favorable consequence for the rater) and is likely to receive more positive personal and organizational outcomes (a positive consequence for the ratee that is likely to favorably impact raters' future interactions with the ratee).

Appraisee reactions to an evaluation can occur both at the time they actually receive the evaluation and later. Clearly, both have important implications for rater motivation. A second negative, but more extreme, ratee reaction that would happen some time after they receive the appraisal might be complaining to the rater's superior or, if the organization is unionized, filing a formal grievance with the union. This might occur if the ratee decides that the appraisal was incorrect or unfair. Here again, raters are likely to try to avoid this potential negative reaction.

At the opposite end of the continuum are situations in which raters expect employees to respond favorably to the appraisal (accept the criticism constructively, try to change in response to the feedback etc.). Favorable reactions might be anticipated because raters believe the overall valence of the appraisal consequences for

the ratee will be positive. This suggests the following hypothesis:

- H2: The more positive the expected consequences of the appraisal for the ratee the more positive the anticipated reaction of the ratee to the appraisal.

Credibility of the Rater to the Ratee

The extent to which raters believe they have a high degree of credibility to the ratee as a feedback source is also expected to influence the anticipated reaction of the ratee to the appraisal. This suggestion comes from previous research on performance feedback (Ilgen, Fisher & Taylor, 1979). Ilgen et al. (1979) argued that perhaps the most important factor influencing the extent to which feedback recipients (i.e., ratees) will accept feedback is the degree of credibility that they attribute to the source of the feedback (i.e., the rater). The credibility of the source is a function of a number of factors. For example, the more expertise the source has the greater his/her credibility to the ratee and the more likely that feedback from them will be accepted (Klein, Kraut and Wolfson, 1971; Tuckman & Oliver, 1968). In addition, the more that recipients trust the source the more credible he/she is and the greater the probability that performance feedback will be accepted (Huse, 1967).

Along these lines, source trustworthiness has been found relate positively to ratee perceptions of the atmosphere and helpfulness of feedback sessions and their satisfaction with the session (Ilgen, Peterson, Martin & Boeschen, 1981). Similarly, Wexley & Snell (1987) found that the extent to which managers were attributed with positive power (a composite consisting of French & Raven's (1959) reward power, expert power and referent power) was positively correlated with employee reactions to an appraisal, such as the perceived accuracy of

the feedback and the motivation to improve. Taken together, this research suggests that when raters believe they are respected and trusted by ratees (i.e., have a good working relationship with ratees) they should worry less about ratees responding defensively to the performance appraisal and should expect acceptance of the feedback regardless of whether it is positive or negative. This leads to the third hypothesis:

- H3: The more credible raters believe they are to the ratee as a feedback source, the more positive the expected reaction of the ratee to the performance appraisal.

Appraisal Visibility

The final variable hypothesized to influence the reaction of the ratee to the performance appraisal is the extent to which raters believe that performance appraisals have a high degree of visibility. Performance appraisals are visible to employees when they are able to find out how coworkers were evaluated. This might happen if employees directly discuss this information among themselves or if it gets passed through the grapevine. A high degree of visibility is expected to increase the probability that raters will anticipate negative ratee reactions to their performance appraisals, either during the actual evaluation interview or subsequently, depending on when employees obtain information about the evaluations of coworkers. To some extent, this relationship may depend on the sign of the performance feedback being received, in that if the evaluation is positive, a high degree of visibility might result in a more positive response to the evaluation (i.e., employees might like others to find out that they received a positive evaluation).

Receiving feedback that is basically positive, however, does not ensure that appraisal visibility will lead to a favorable response because of the likelihood that self appraisals and self-other equity comparisons will reduce the attractiveness of the feedback received. Specifically, since most of the research on self appraisals (Holzbach, 1978; Kirchner, 1966; Klimoski & London, 1975; Parker et al., 1959; Thornton, 1968; 1980; Waldman & Thornton, 1978) suggests that they are more lenient than supervisory appraisals, even an employee who receives a fairly positive evaluation might be unhappy about it, and thus, react negatively, if they felt they should have received a more favorable rating. This possibility is further enhanced by the tendency for self-other equity comparisons to occur. According to Adams (1965), employees frequently compare their inputs (e.g. how hard they work, the amount of work they do) and outcomes (in this case, the evaluation they received) with the inputs and outcomes of others. One reason that self appraisals are generally higher than supervisory evaluations may be that employees tend to overestimate their inputs. If this is true (or if managers believe it to be true), it means that felt-negative inequity perceptions are likely to occur frequently (or at least be expected frequently by managers). The more visible that appraisals are to members of the workgroup, the more likely that raters will expect such comparisons to occur and to expect employees to react negatively to appraisals because they believe they should have received a higher evaluation. This suggests the fourth hypothesis to be tested in this study:

- H4: The more raters believe appraisals are visible to members of their workgroup, the more negative the anticipated reaction of the ratee to the appraisal.

Appraisal Visibility

Task Interdependence Among Employees

Several researchers have noted the potentially important influence of task interdependence on the behavior of people in organizations (e.g. Cheng, 1983; Kane & Lawler, 1979; Kiggundu, 1983; Liden & Mitchell, 1983; Mitchell, 1983). Task interdependence among employees in a workgroup exists when the nature of the tasks performed requires them to work together and interact on a regular basis in order to achieve high performance. According to Cheng (1983), when tasks are highly interdependent, no one work role can be performed effectively unless all or most other work roles are carried out properly.

Task interdependence is likely to have a direct and positive effect on appraisal visibility. Specifically, the more interdependent that tasks are in the rater's workgroup, the more opportunities there are for employees to discuss their evaluations with each other. When a manager's employees rarely see one another (an extreme example of task independence) it is less likely that they will find out how each other were evaluated. Therefore, it is hypothesized that:

- H5: The greater the degree of task interdependence among members of a workgroup, the greater the degree of perceived appraisal visibility.

Perceived Freedom to be Honest

A rater's perceived freedom to be honest is seen as an important attitudinal indicator of whether or not raters are motivated to evaluate performance accurately. It is seen as playing a role in this model that is similar to that played by "turnover intentions" in models of

turnover (e.g. Mobley, Horner & Hollingsworth, 1978.). Perceived freedom to be honest has several hypothesized causes in the model of rater motivation that was presented in Figure 4.

Reaction of the Ratee to the Appraisal

The expected reaction of the ratee to his/her performance appraisal is hypothesized to have a direct effect on raters' perceived freedom to be honest and objective in rating performance. Specifically, the more negative the expected reaction the less free raters are likely to feel to be honest. Bernardin and his colleagues (Bernardin & Beatty, 1984; Bernardin & Buckley, 1981) have argued that the tendency of raters to be lenient in providing performance evaluations is probably a defensive behavior aimed at avoiding the potential negative reactions of employees to harsh ratings. This defensiveness may occur because many raters feel they lack the ability to cope effectively with the ratee's anger. This leads to the following hypothesis:

- H6: The more negative the anticipated ratee reaction to the performance appraisal, the lower the perceived freedom of raters to be honest when providing performance evaluations.

Rater's Desire to be Liked by the Ratee

A second hypothesized cause of the perceived freedom of the rater to be honest when evaluating performance is the extent to which the rater wants to be liked by the ratee. Raters having a strong desire to be liked by a particular ratee are likely to do things they believe will make the ratee like them (e.g. giving the ratee higher ratings than they feel are warranted) and to avoid behaviors which might threaten their relationship with the employee (e.g. giving them low ratings, even if they are deserved).

This notion is supported by research concerning the need for affiliation which suggests that individuals possessing a strong desire for companionship and friendly interpersonal relationships may dispense rewards (e.g. positive performance evaluations and thus, the potential for larger salary increases and/or likelihood of being promoted) as a way of winning or keeping friends (McClelland & Burnham, 1976). As a result, individuals with a strong desire to be liked by their employees may feel less freedom to be honest when evaluating performance since an honest, but at least partially negative, evaluation could create tension between them and the ratee. Therefore, it is hypothesized that:

- H7: The stronger a rater's desire to be liked by a ratee, the lower his/her perceived freedom to be honest when providing performance evaluations.

Ability to Document the Appraisal

Another hypothesized cause of the freedom of raters to be honest is their ability to document the performance evaluation. When raters believe they are able to document and support their evaluation of a ratee with critical incidents of good and poor performance their perceived freedom to be honest when evaluating performance should be greater. This is based on the belief that feedback is more likely to be accepted when it is supported by specific documentation (Ilgen, Fisher & Taylor, 1978; Leskovec, 1967).

In a similar vein, several researchers (e.g. Bernardin & Buckley, 1981) have advocated diary-keeping as way of increasing the ability of raters to make accurate ratings of performance by improving their observation of behavior. For example, results of one study on diary-keeping (Bernardin & Walter, 1977) found that those raters who

regularly recorded critical incidents of employee performance had significantly lower leniency and halo and greater interrater agreement than those who did not keep such diaries. It is suggested here that diary-keeping may also positively impact the motivation to rate accurately. Raters who feel they have concrete information and examples to support their evaluations are likely to have greater confidence in their appraisals and thus, feel less apprehensive about providing negative feedback to ratees. This should then increase their perceived freedom to be honest. Therefore, the eighth hypothesis to be tested is:

- H8: The more that raters feel they have adequate documentation to support their performance evaluations, the greater their perceived freedom to be honest when providing performance evaluations.

Appraisal Visibility

Finally, the perceived visibility of the appraisal is hypothesized to influence the freedom of raters to be honest when doing performance appraisals. Appraisal visibility is expected to have a direct and negative effect on the freedom of raters to be honest because it increases the potential for conflict among employees. If an employee finds out how coworkers were evaluated and, as a result, comes to believe he/she has been treated unfairly, not only may the employee react negatively during the appraisal, but he/she may also argue with coworkers. The possibility of conflict occurring among employees may cause raters to feel less free to differentially evaluate (and reward) all employees based on their actual performance level. This suggests the following hypothesis:

- H9: The greater the degree of appraisal visibility, the lower the perceived freedom of raters to be honest when doing performance evaluations.

The Occurrence of Rendering Errors

As described earlier, rendering errors occur when there is a difference between private evaluations of performance and the actual ratings publically recorded on an appraisal form. This is considered to be a behavioral indicator of the extent of rater motivation since the motivation to rate accurately must be low if private and public evaluations of performance differ.

Perceived Freedom to be Honest

The only hypothesized cause of the occurrence of rendering errors in the model is the perceived freedom of raters to be honest when evaluating performance. The occurrence of rendering errors represents an external and behavioral (i.e., nonattitudinal) measure of the outcome of the perceived freedom to be honest. In other words, it indicates the extent to which this attitude is translated into actual behavior. Thus, the role that the occurrence of rendering errors plays is comparable to that of measures of actual turnover in turnover models. The less free that raters feel to be honest, the more likely they are to record public evaluations that differ from their private evaluations. Therefore, it is hypothesized that:

- H10: The lower the perceived freedom of raters to be honest when doing performance evaluations the greater the difference between their public and private evaluations of performance.

Summary

In the previous section a model describing the relationship between several motivational influences on performance appraisal was

described. Previous researchers have recognized appraisal purpose as an important motivational variable but thus far, have failed to describe the process by which purpose influences performance ratings. The model described above attempts to remedy this deficiency by elaborating some of the important motivational constructs intervening between the purpose that appraisals serve in organizations and actual performance ratings. Thus, the model represents a first step in understanding the process by which raters are motivated to provide accurate or inaccurate evaluations of performance. In the study described below, this model was submitted to empirical test.

CHAPTER 3: METHOD

Overview of Methodology

Based on a review of relevant literature a model was developed describing the relationship between a number of constructs thought to influence rater motivation to evaluate performance accurately. A questionnaire was then designed to assess these motivational influences. The questionnaire was piloted on a small group of managers to assess its psychometric adequacy and was then completed by a group of full-time employed managers. In addition to completing the questionnaire, these managers participated in a short interview during which they were asked to provide researchers with their most accurate assessment of the performance of a subordinate in their workgroup. This private evaluation was then compared to the most recent public evaluation which the manager provided for that employee (collected from organizational records) to develop a measure of the occurrence of rendering errors. The extent to which the model of rater motivation fit the data was then assessed using latent variable structural equation analysis.

Participants

Two groups of people participated in the study. Forty-seven managers and 54 students participated in the three phases of the pilot study (described below) and 124 managers were involved in the primary data collection for the study. Recruitment of participants for both the pilot study and primary study took place in several stages. The personnel office of the organization was contacted about participation in the study. When consent was given, the researcher was provided with the names of personnel representatives in various units of the

university. The representatives were contacted and asked to provide the researcher with the names of managers in their units. These managers were then approached by the researcher and asked to participate in the study. It is worth noting that of all the managers called by the researcher, for either the pilot study or the primary data collection, only 3 were unwilling to participate in the study.

As recommended by Cohen (1969), an a priori power analysis was conducted to determine the sample size needed to have adequate power to detect a significant effect. Based on an overall effect size of .20 (considered by Cohen to be a small effect size), 12 predictors, an alpha of .05, and power of .80, it was determined that 100 participants would provide an adequate amount of power. One hundred and twenty four managers agreed to participate in the primary study.

The managers participating in the primary data collection were employed full-time by a large midwestern university. Although there may be external validity problems with collecting data from a single organization, it was considered desirable that the subject population be drawn from the same organization to ensure consistency in the performance evaluation forms used across participants. This eliminated the need to standardize performance ratings before doing statistical analyses.

Furthermore, in many ways, a large university setting offered a partial solution to the generalizability problem. This is because a large university consists of many relatively autonomous units that are involved in very different types of work activities. Therefore, it allowed, in a single setting, the collection of data from both skilled and unskilled, and white collar and blue collar workers, as well as

greater variance on educational level. The managers participating in the primary study were employed in a wide variety of areas of the university. These included, but were not limited to: grounds maintenance, security, housing and food service, clerical services, library, administration (e.g. accounting, payroll, recruitment etc.), personnel services, university health center, university computers and information services, and public relations and funds development.

The primary sample consisted of 55 males and 69 females, ranging in age from 26 to 65. Ninety-four percent were white, 5% were black and 1% were Asian. It should be noted that all managers involved in the study had been employed by the participating organization for at least a year, had been employed in their current position for at least 6 months, and had held a supervisory position for at least a year. This is important because it demonstrates that all participants in the study had some exposure to the performance evaluation process as implemented in this organization (and their specific department) and had actually conducted performance evaluations on several occasions. Although the number of employees that participants evaluated varied to some extent depending on the area in which they worked, 95% of the participants evaluated between 2 and 15 employees.

Procedure

The Pilot Study

Three separate sets of people participated in the pilot study: two groups of managers and one group of students. All managers were employed by the same organization as used in the primary data collection but were not members of the sample for the main study. The

purpose of the pilot study was twofold: (1) to make a preliminary assessment of the technical adequacy of the questionnaire and (2) to determine the extent of variance in the measure of rendering errors. The former was important because all of the constructs measured with the questionnaire were exploratory in nature and, thus, not measured with previously used and tested scales. The latter was necessary because the potentially sensitive nature of the information being assessed (i.e., the extent to which raters intentionally provide inaccurate performance ratings) made it possible that managers would hesitate to respond honestly and thus, reduce the variance on this measure.

The first phase of the pilot study assessed the technical adequacy of the questionnaire by checking the clarity of the items on the questionnaire and making preliminary assessments of scale reliabilities. Five managers completed the questionnaire and then met with the researcher to discuss it. They provided input which was used to edit unclear or ambiguous items, eliminate irrelevant items and reduce possible misinterpretations.

Thirty managers completed the revised questionnaire so that initial assessments of scale reliability could be made. Although Nunnally (1978) suggests that reliabilities above .60 are adequate for exploratory work, it was desired that these preliminary reliabilities be above .70 if possible. Results from these managers revealed that all of the scales had reliabilities above .70 with the exception of the termination scale ($\alpha = .63$) and the interdependence scale ($\alpha = .54$). In order to improve these scales, the items in them were re-written. In addition, two of the other scales were each

modified by eliminating a single item that substantially improved the reliability of the overall scale.

The four scales which had been modified were cross-validated by administering them to a group of 54 part-time employed undergraduate students in two organizational behavior classes. These students were asked to fill out the questionnaire by responding in terms of the organization where they currently worked. Reliabilities for the two scales which had been rewritten were found to be above .70 while those for the two scales which had been modified were similar to the reliabilities found in the original pilot sample (after eliminating the bad item). The results of these analyses suggested that the questionnaire scales were adequate for the primary data collection. The actual items and scales included on the questionnaire are described in detail later.

Twelve managers participated in the piloting of the dependent measure. These managers completed the final version of the questionnaire and then participated in the interview during which performance evaluation information was collected. The actual procedures followed were the same as those for the primary data collection and are described in more detail below. Results from this pilot demonstrated enough variance on the dependent measure to proceed with the primary study.

The Primary Study

All managers were personally contacted by telephone and asked to participate in the study. During the initial contact, the study was described to participants as one with the purpose of learning how

managers conduct performance appraisals and some of the factors affecting this process. Participants were told that the study would consist of filling out a questionnaire and meeting with the researcher for a short interview and that the total time involved would be about one hour.

After managers agreed to participate in the study, the researcher noted that portions of the questionnaire would require them to answer questions in relation to a particular employee in their workgroup (called the "focal ratee"). Managers were told that the focal ratee should be the individual on whom they had most recently completed a formal performance evaluation subject to the constraint that the evaluation had been done at least 2 weeks ago. This procedure for selecting the focal ratee approximates random selection since which individual was most recently evaluated depended only on the ratee's employment date (evaluations were completed annually during the month that employment in the organization commenced). It was important that specific procedures for selecting the focal ratee be given to managers to ensure that they did not use nonrandom selection criteria (e.g. selecting a subordinate whom they personally liked or who was a good performer).

Participants were contacted about a week after mailing the questionnaire to ensure that it was received and to set a date for the interview portion of the study (which lasted approximately 30-40 minutes). Questionnaires were collected from participants at the beginning of the interview. It should be noted that because of the procedures used to collect data in this study, the response rate for the questionnaire was very high (over 95%). Only those managers who

were unable to continue their participation in the study due to unanticipated time constraints (5 managers), or illness (1 manager) did not return the questionnaire.

The primary purpose of the interview was to collect from participants the information needed to determine the extent of difference between the private evaluation and the public evaluation (i.e., the measure of rendering errors, described in greater detail later). The public rating was the evaluation of the focal ratee most recently completed for the organization (obtained from the employee's personnel file) while the private rating was the rater's actual opinion of the ratee.

In order for participants to feel free to provide the private evaluation it was necessary to create a climate in which they would feel comfortable providing an honest assessment of the focal ratee's performance. It was felt that this could best be accomplished by giving participants the opportunity to talk informally with the researcher for 25-30 minutes so that rapport could be developed. This should also have increased the willingness of managers to provide the researcher with a copy of the employee's most recent evaluation from organizational records. Thus, the interview portion of the study was considered to be crucial for obtaining the data.

The interview proceeded by asking managers to talk about some of their experiences in doing performance evaluations. The interview followed a semi-structured format, in which there were both standard questions asked of all participants and questions which flowed naturally from the comments which participants made¹. After the

informal discussion ended, participants were asked to provide the researcher with their private evaluation of the focal ratee's performance. The private rating was done on the same evaluation form normally used by the manager when providing performance evaluations for organizational purposes.

Although during the interview participants completed their private rating prior to obtaining the public rating from the employee's file, the public rating actually occurred in time before the private rating (i.e., it took place before managers were asked to participate in the study). It was important, however, that the private rating be collected from participants first so that the public evaluation would not be particularly salient to the manager when completing the private evaluation. Thus, priming and consistency effects (Salancik & Pfeffer, 1978) should not have caused managers to intentionally or unintentionally reduce the difference between public and private ratings. This was further reinforced by making sure that there was at least three weeks between completing the public evaluation and the date of the interview, when the private evaluation was done.

To increase the likelihood that participants would be honest when providing this evaluation, the researcher informed them that the evaluation was for research purposes only and thus, would not be seen by anyone in the organization. This is because previous research (e.g. Sharon & Bartlett, 1969) suggests that evaluations are less lenient, and therefore, may be more accurate, when completed for research purposes only. The instructions for providing this evaluation were as follows:

"Now what I'd like you to do is take a few minutes to think about the performance of the focal ratee and provide me with the most accurate evaluation of the performance of this employee that you can. In doing this evaluation, think only of how well the employee does his/her job. The reason I say this is because sometimes when managers evaluate an employee's performance they may think about things other than just how the employee performs on the job. If this is the case for you, in this situation, please ignore any of the other factors that might affect your evaluation when you do it for the organization and think only of the employee's performance. Keep in mind that this evaluation is being done only for purposes of this research and will not be seen by anybody in the organization. Also, do not write the name of the employee on the evaluation form so that there will be no way of identifying whose performance is being evaluated."

It was important that it be clear to participants that the evaluation they were providing here (i.e., the private evaluation) need not be the same as the last evaluation they completed for the employee while in no way suggesting that the two evaluations ought to be different. Informal observations of managers during these instructions indicated that they often seemed confused about what they were to do until the researcher noted that sometimes managers considered factors other than performance when doing evaluations for organizational purposes. After this point, most managers had no difficulty in doing the task, and, in fact, often volunteered the information that their evaluations for the organization did not always reflect their true opinion of an employee's performance.

Next managers were asked to provide the researcher with a copy of the most recent evaluation that they had done for the focal ratee. In making this copy, it was stressed that the manager should black out the employee's name, the signatures at the bottom of the form, and any comments written about the employee that might identify him/her. In this way, the employee's identity was protected.

To ensure that the manager did not believe the performance of the focal ratee had changed significantly since the date of the public evaluation (and thus, that differences between the two ratings were not due to true performance differences) several safeguards were taken. First, all focal ratees included had been in their current position for at least six months so that the majority of the initial learning would have taken place (and 91% of the ratees had been in their job for a year or more). In addition, it was important that the time between the two evaluations was not great enough that the ratee's performance was likely to have changed. Thus, for 90% of the ratees, the time between the two evaluations was between 3 weeks and 6.5 months. Finally, after obtaining a copy of the public evaluation, managers were directly asked if they believed the focal ratee's performance had changed significantly since the date of this evaluation. Three managers gave an affirmative response to this question, and thus, were eliminated from the study. This resulted in a final sample of 115 managers.

Variables

Rendering Errors

The primary criterion in this study was the extent to which raters commit rendering errors. Following the ideas of several researchers (Banks & Murphy, 1985; Mohrman & Lawler, 1983), rendering errors occur when there is a discrepancy between a rater's actual opinion of the ratee's performance (i.e., the private rating) and what she/he marks on an evaluation form (i.e., the public rating). It is suggested that the extent to which there is a difference between these

two evaluations is a behavioral indication of a rater's motivation to rate accurately. Specifically, the relationship between rendering errors and rater motivation should be negative, so that the greater the difference between private and public evaluations of performance the lower a rater's motivation to rate accurately. This seems appropriate since motivation to rate accurately must be low when raters intentionally evaluate subordinates differently than they feel they really ought to be rated.

The occurrence of rendering errors was measured using a difference score. The algebraic difference (rather than absolute difference) between the public and private evaluations was used in this measure. While both intentional under-evaluations (i.e., deflated ratings) and over-evaluations (i.e., inflated ratings) are equally indicative of low motivation to rate accurately, only differences in the positive direction (where public evaluations are higher than private evaluations) were included in the measure of rendering errors. This is because inflated ratings have been found to occur more frequently than deflated ratings (Longenecker, Gioia & Sims, 1987) and thus, only motivational influences likely to result in positive rendering errors were incorporated into the model. Therefore, the dependent measure was really a measure of rating inflation.

The participating organization utilized two primary evaluation forms, one for employees in administrative and/or management positions and one for nonmanagerial employees². The forms are presented in Appendix A. These evaluation forms were the basis for determining the extent to which raters made rendering errors. The forms were very

similar in that both consisted of a list of general job traits or dimensions (e.g. Job Knowledge, Dependability, Attitude and Cooperation etc.) that were evaluated on a 5-point rating scale, with anchors ranging from "outstanding" to "unsatisfactory."

The primary difference between the forms was in the number of job dimensions on the form and what the actual dimensions were. The nonmanagerial form consisted of seven performance dimensions while the administrative/managerial form contained nine performance dimensions. There was some overlap in the dimensions included on the two forms but the administrative/managerial form contained two dimensions that were only appropriate for managers (Ability to Develop Subordinates and Supervision) as well as two dimensions that were only relevant for employees in certain types of administrative/managerial positions (Cost Control and Affirmative Action). Because there was some variability in the number of performance dimensions that were actually used by participants in evaluating their subordinates, it was necessary to take this into account when computing the measure of rating inflation.

The actual measure of rating inflation was calculated by subtracting the private evaluation from the public evaluation on each performance dimension. Positive differences (i.e., where the public evaluation was higher than the private evaluation) on any performance dimension were then summed and divided by the maximum difference possible given the number of dimensions used on the evaluation form. For example, with 5-point rating scales, the maximum difference between public and private ratings possible on any performance

dimension would be 4 ($5 - 1 = 4$). If evaluations were provided on seven performance dimensions, then the maximum difference possible would be 28 ($4 \times 7 = 28$). Calculated in this way, the measure of rating inflation indicates the proportion of all differences possible that were in the positive direction. It is worth noting that 35% of the instances of differences between public and private evaluations were deflations (i.e., the public evaluation on some performance dimension was lower than the private evaluation for that dimension) and, as indicated earlier, deflations were ignored when calculating the measure of rating inflation for each participant.

Motivational Influences

The motivational influences were measured using a questionnaire developed by the researcher for this purpose. With the exception of Performance Appraisal Consequences, they were all measured using 5-point Likert scales (ranging from "strongly agree" to "strongly disagree") where respondents were asked to indicate the extent to which they agreed with each statement. All of the motivational influences reflected the rater's perceptions of various aspects of the performance appraisal situation. This stems from the underlying assumption behind the model that whether or not raters are motivated to rate accurately is determined by their definition of the appraisal situation. However, it creates a potential percept-percept (or common method variance) measurement problem (Campbell & Fiske, 1959). The problem with having all measures provided by one source is that indicators of relationships between variables may be inflated. While this problem cannot be eliminated in the present study, several factors combine to lessen the negative effect of the problem.

First, although all measures were provided by the rater, the two components of the measure of rating inflation (i.e., the public and private evaluations) were obtained at different times and both were obtained at a different time than the questionnaire measures of the motivational influences. As noted above, the public evaluation was collected from organizational records while the private evaluation was obtained about two weeks after the questionnaire measures during the interview with the participant. This temporal separation reduces the potential for inflated relationships resulting from obtaining measures of variables from the same source.

Secondly, when testing the model of rater motivation, the pattern of relationships between variables is more important than the actual magnitude of the relationships in determining the fit of the data to the model. Thus, although the magnitudes could be inflated due to percept-percept bias, this should not affect the pattern of relationships between the variables and hence, should not bias a test of the overall fit of the model.

Each of the motivational influences is described below. Two types of motivational influences were measured. Some of the motivational influences deal with a particular employee. Items assessing these motivational influences were answered in relation to the "focal ratee" (described above). The other motivational influences deal with either the organization in general or characteristics of the participant's department or workgroup. Thus, these items were answered independently of any particular employee. Motivational influences of the former type were: (1) expected

consequences of the appraisal for the ratee; (2) credibility of the rater to the ratee; (3) rater's desire to be liked by the ratee; and (4) reaction of the ratee to the appraisal. Motivational influences of the latter type were: (1) appraisal purpose; (2) task interdependence among employees; (3) ability to document the appraisal; (4) appraisal visibility; and (5) perceived freedom to be honest. Examples of items to measure each of the variables are provided below. A complete copy of the questionnaire is included in Appendix B. The procedures for measuring the expected consequences of the appraisal for the ratee are provided in Appendix C while a list of the questionnaire items included in each of the other scales is included in Appendix D.

Consequences of the Appraisal to the Ratee. This scale was used to determine the overall perceived attractiveness of the appraisal consequences to the focal ratee. Participants were provided with a list of eleven potential outcomes that a subordinate could obtain from a performance appraisal (e.g. a large salary increase, a promotion, a transfer, development of skills and abilities etc.) and asked to indicate two things for each outcome. First, participants indicated the likelihood that each of the outcomes would occur given the subordinate's actual performance level. This was measured as a probability, ranging from "0" (will definitely not occur) to "1" (will definitely occur). Second, managers were asked to indicate how attractive they thought each outcome was to the subordinate. This was measured with a 5-point scale ranging from "would like receiving this outcome very much" to "would dislike receiving this outcome very much." The perceived instrumentality of each outcome was multiplied

by the valence of that outcome and summed across all outcomes to yield an overall indication of the valence of the appraisal consequences for the ratee. High scores on this variable indicated that the rater believed that, given the ratee's true performance level, the consequences of the appraisal for the ratee would be positive.

Credibility of the Rater to the Ratee. This scale measured the extent to which raters believed they were trusted and respected by the focal ratee. Sample items included: "This individual trusts my judgment on work-related matters" and "This employee does not think very highly of me as a supervisor" (reverse scored). A high score on this scale indicated that raters believed they had a high degree of credibility to the ratee. There were six items in this scale.

Desire to be Liked by the Ratee. This scale assessed the degree to which the rater wanted to be liked by the focal ratee and to have a good relationship with him/her. Sample items included, "In order to be satisfied with my work, I need to have a good working relationship with this employee" and "I would not go out of my way to try to get this person to like me" (reverse scored). High scores on this scale indicated that it was important to the rater to be liked by the focal ratee. Five items were included in this scale.

Reaction of the Ratee to the Appraisal. This scale measured the extent to which raters felt that the focal ratee was likely to react defensively to performance feedback. Sample items included, "This person is able to respond constructively to feedback on his/her performance" and "It is not uncommon for this individual to feel that I am attacking him/her personally if he/she receives less than the

highest performance ratings" (reverse scored). High scores on this scale indicated that the rater believed the ratee would respond positively to the performance evaluation. There were seven items included in this scale.

Purpose of the appraisal consisted for four subscales, employee development, salary decisions, promotion decisions, and termination decisions, each of which were measured separately. Each subscale was included in the causal model tested in this study as a separate exogenous variable. Each subscale is described below.

Purpose of the Appraisal: Employee Development. This scale indicated the extent to which the rater believed performance appraisal information was used to help employees grow and develop on the job. Sample items included, "Formal performance appraisals provide a means for me to get together with each of the individuals in my department to discuss how to help them become better employees" and "In this organization, performance appraisals are rarely used to show individuals areas of their performance where improvement is needed" (reverse scored). A high score on this scale indicated that raters believed subordinate development was an important purpose for performance appraisals. Five items were included in this scale.

Purpose of the Appraisal: Salary Decisions. This scale measured the extent to which managers believed performance appraisal information was used in making salary decisions. Sample items included: "In this organization the best way to ensure receiving a large wage/salary increase is to receive a good performance appraisal rating" and "Most of the raises that the people in my unit receive are based very little upon merit" (reverse scored). High scores on this

scale meant that raters believed performance appraisals had a great deal of impact on salary decisions. There were six items included in this scale.

Purpose of the Appraisal: Promotion Decisions. This scale assessed whether or not managers believed performance appraisal information was used in making promotion decisions. Sample items included: "Only people who receive high performance evaluations will be promoted in this organization" and "Promotions are based on who you know rather than how well you perform" (reverse scored). High scores indicated that managers perceived promotion decisions to be based upon performance appraisal information. There were four items in this scale.

Purpose of the Appraisal: Termination Decisions. This set of items indicated the extent to which managers thought termination decisions depended upon performance appraisal data. Sample items included, "Termination decisions are made only after consulting an employee's performance appraisal records" and "A person's performance on the job is not a major factor considered by those who make termination decisions" (reverse scored). High scores indicated that managers believed performance appraisal information was used in making termination decisions. Five items were included in this scale.

Ability to Document the Evaluation. Items in this scale measured the extent to which raters felt they were able to support their performance evaluations of employees with specific behavioral examples. Sample items in this scale included, "I am generally able to support my evaluations of individuals working in my unit with

specific incidents of good and poor performance" and "I should keep better records on the performance of people in my department than I do" (reverse scored). A high score on this scale meant that raters believed they were typically able to document their performance evaluations of subordinates. There were four items in this scale.

Task Interdependence Among Employees. This scale measured the extent to which the jobs supervised by the rater required a great deal of interaction among employees in order to be completed effectively. Sample items included, "The people that I supervise often need to coordinate their work activities with each other" and "The jobs which I supervise don't require much interaction among employees" (reverse scored). High scores on this scale were indicative of a high degree of task interdependence among subordinates. Six items were included in this scale.

Appraisal Visibility. This scale assessed whether raters believed that members of their workgroup would find out how each other were evaluated. Sample items in this scale included, "People in my workgroup often compare their performance ratings" and "It would be very unusual for individuals in my unit to mention their performance appraisal ratings to each other" (reverse scored). High scores on this scale meant that managers believed performance appraisals had a high degree of visibility. Four items were included in this scale.

Perceived Freedom to be Honest. This scale measured the extent to which raters felt free to rate employee performance honestly and openly. Sample items included, "I would rarely hesitate to tell an employee my true assessment of his/her performance" and "If there was some way I could avoid having to approach my employees about a problem

with their performance I would do it" (reverse scored). High scores on this scale indicated that raters felt free to be honest when evaluating an employee's performance. There were four items in this scale.

Data Analysis

Overview of Linear Structural Equation Analysis

The primary data analytic strategy used in this study was the analysis of linear structural equations. The analysis of linear structural equations was accomplished using LISREL VI (Joreskog and Sorbom, 1984), a procedure that derives parameter estimates for the unknown coefficients in a set of linear structural equations. Parameter estimates can be derived using either a maximum likelihood or an unweighted least squares solution.

In this study, a latent variable structural model with multiple manifest indicators was utilized. A latent variable is an unobserved variable presumed to exist within a structural model but which can't be measured directly; in other words, it is a hypothetical or theoretical construct (James, Mulaik & Brett, 1982). The primary reason for using latent variable models with multiple indicators, rather than manifest variable models where each latent construct is represented by only one manifest variable, is that they offer a solution to the problem of working with variables that are not measured with perfect reliability (Bentler, 1980). Unreliability of variables is a problem because it results in biased estimates of structural parameters and path coefficients linking latent variables (James et al., 1982). In addition the use of latent variable models

allows testing the a priori measurement model to determine whether the manifest indicators are, in fact, related to the latent variables with the hypothesized structure. Essentially, this is a test of the construct validity of the measurement instrument (James, et al., 1982).

The testing of latent variable structural models with LISREL proceeds through a two step process. The first step involves testing the measurement model, which details the relationships between each latent variable (or cause) and the manifest, or measured, variables (the effects) that serve as indicators of that cause. The measurement model is tested using confirmatory factor analysis to determine whether the items on the questionnaire form the clusters intended a priori to exist. The second step involves an assessment of the adequacy of the hypothesized structural model, which specifies the causal relationships among the latent variables. The goodness-of-fit for both models is determined by the extent to which the observed correlation matrix is similar to the reproduced correlation matrix based on the parameter estimates derived from the hypothesized model. The more similar the reproduced and observed correlation matrices are, the better the degree of fit.

Assessment of Fit

There are a number of ways to assess the degree of fit for a model (i.e., the extent to which the hypothesized model is consistent with the data). According to Joreskog & Sorbom (1984), unreasonable values for parameter estimates (e.g. correlations greater than 1.00), squared multiple correlations or coefficients of determination that are negative or large standard errors are all indications that the

model does not fit the data very well.

In addition, there are several specific measures which indicate the overall goodness of fit for both the measurement and the structural model. The Chi-Square (χ^2) and its associated degrees of freedom and probability level provides one overall measure of fit (for maximum likelihood estimation procedures only). Although the χ^2 measure can theoretically be shown to be the likelihood ratio test statistic for testing the hypothesized model against the alternative that the model is unconstrained (in which case perfect fit would result), Joreskog and Sorbom (1984) do not recommend using it in this way since the assumptions underlying this usage are rarely met in practice. Rather, they suggest that the χ^2 be used as an overall index of fit, where large values correspond to poor fit and small values correspond to good fit. The degrees of freedom in the model serve as the standard for determining whether the χ^2 is large or small. According to the authors, a ratio of χ^2 to degrees of freedom of 3:1 or less reflects a good degree of fit.

A second overall way to assess fit, the goodness of fit index (GFI), is a measure of the relative amount of variances and covariances jointly accounted for by the model (Joreskog & Sorbom, 1984). This goodness of fit index can also be adjusted for the number of degrees of freedom in the model (called the adjusted goodness of fit index, or AGFI). Both of these measures should be between zero and one, with values closer to one indicating better fit. Finally, the root mean square residual can be used to assess overall fit. The root mean square residual (RMSR) is a measure of the average of the

residual variances and covariances. The smaller the value of the RMSR, the less the difference between the observed and reproduced matrices and thus, the better the degree of fit. Formulas for all fit indices are provided in Joreskog and Sorbom (1984).

Assumptions Underlying the Use of Structural Equation Analysis

The use of structural equation analysis involves a number of assumptions about the data and the model which should be true when data are collected at one point in time in order to make strong causal statements about the relationships among the variables in the model (see Bentler, 1980 and James et al., 1982 for a more complete discussion of the conditions and assumptions underlying the use of causal analysis). For example, it is assumed that causal effects have occurred rapidly, that the system of relationships among the variables has reached an "equilibrium-type condition" at the time of data collection (i.e., are relatively stable and constant), and that the structural model, as originally hypothesized, is specified correctly. The latter condition implies (1) that the paths hypothesized to have nonzero structural parameters are actually significantly different from zero and (2) that unspecified paths hypothesized to have structural parameters equal to zero do have parameters not differing from zero. Additionally, it is assumed that all variables are measured on interval scales and with a high degree of reliability and that relationships among variables within the model are linear.

A further assumption inherent in all forms of causal analysis is that the causes for a dependent endogenous variable are uncorrelated with the residual (or error term) of the causal equation for that endogenous variable and with the residual for any endogenous variable

occurring later in the causal ordering of the model (James, 1980). This assumption also implies that the error terms for the path equations of each endogenous variable are uncorrelated (Duncan, 1975; James, 1980; James et al., 1982). To the extent that this assumption is violated, it indicates that there are relevant unmeasured causes in the model. An unmeasured cause is considered to be relevant (or important) when it is stable, has a nontrivial direct influence on an effect, is related to at least one other cause in the model and makes a unique contribution to the model (James et al., 1982). Relevant unmeasured variables are a problem because they result in biased estimates of path coefficients for the variables that are included in the model.

When using structural equation analysis, an important concern is the extent to which the hypothesized model is identified. Identification concerns whether or not enough information is available to obtain unique mathematical solutions for the structural parameters (James et al., 1982). In order for a model to be tested it must be overidentified, which means, loosely speaking, that there are more data points (correlations) than there are parameters to estimate. Models which are underidentified or just identified cannot be tested because the one-to-one correspondance between data and parameters means that these models cannot be rejected (Bentler, 1980).

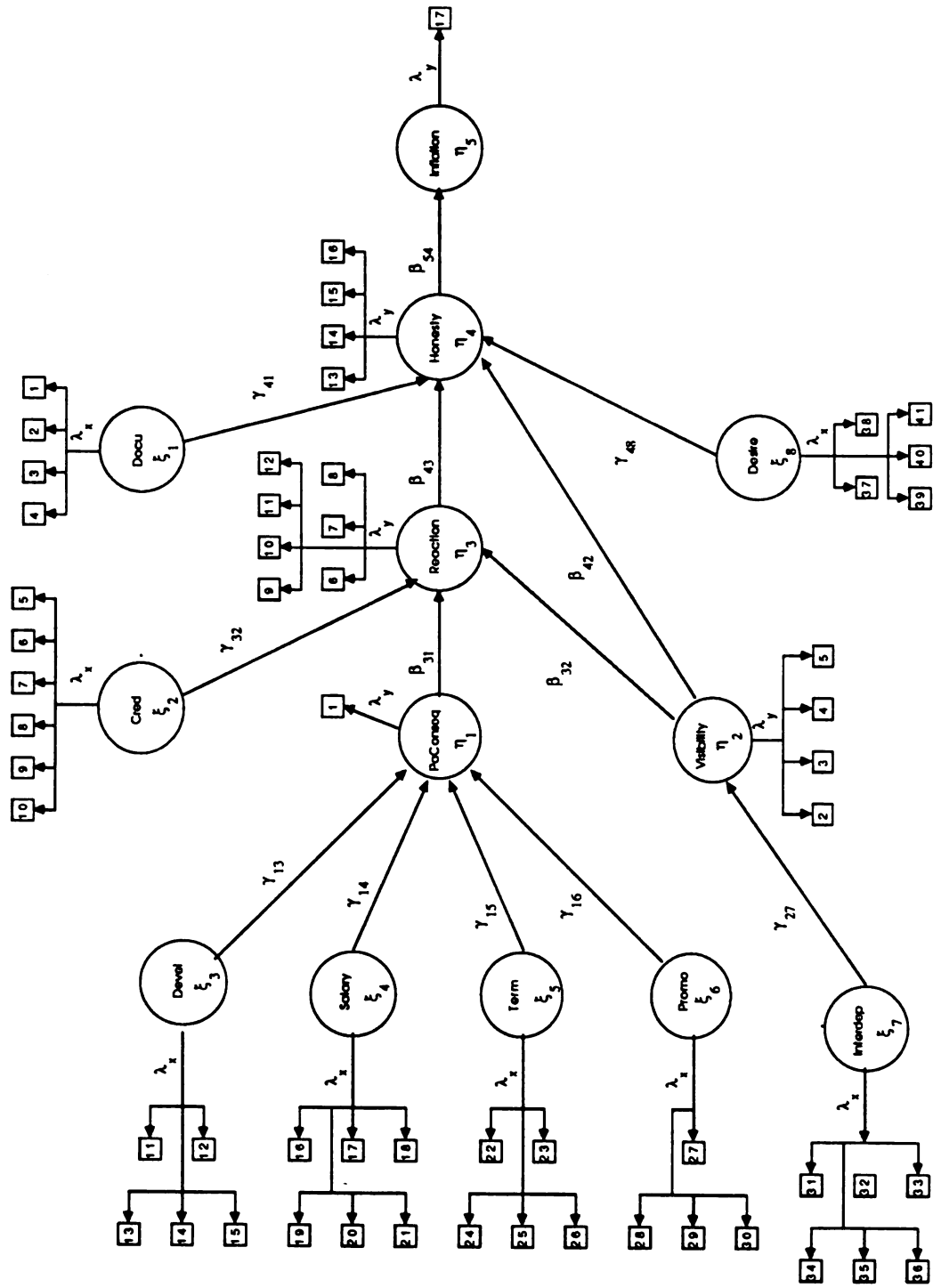
Although determination of the identification status of any model is extremely complex, some guidelines are available. Specifically, James, Mulaik and Brett (1982) suggest that each latent variable should be represented by at least four manifest indicators. When this

is the case, the measurement submodels relating a set of manifest indicators to their respective latent variable will be overidentified and it should be possible to test the underlying measurement model. In the present study, all latent variables had at least four indicators. Due to the complexity of the identification issue, it should be noted that the LISREL VI procedures check the identification status of any model before computing parameter estimates. This check has been found to be nearly 100% reliable (Joreskog & Sorbom, 1984). For a more thorough discussion of identification see James et al. (1982) or Kenny (1975).

Description of Diagram Depicting the Measurement and Structural Models

A pictorial representation of the measurement model and structural model tested in this study is provided in Figure 5³. Following the conventions in the structural modeling literature, observed variables (i.e., the manifest indicators) are enclosed in squares and denoted with Arabic letters ("x" for the manifest indicators of the exogenous variables and "y" for the indicators of the endogenous variables). Due to space limitations in the figure, x-variables are numbered consecutively from "1" to "41" (rather than "x₁" to "x₄₁"), beginning with "docu" (ability to document the appraisal) and ending with "desire" (desire to be liked by the ratee). The y-variables are numbered consecutively from "1" to "17", beginning with "paconseq" (performance appraisal consequences) and ending with "inflation" (rating inflation). The actual questionnaire items corresponding to each of the manifest indicators in Figure 5 are presented in Appendix D. Latent variables are enclosed in circles and labeled with Greek letters (κ , ξ , for the eight exogenous variables

Figure 5: Hypothesized Measurement and Structural Model^a of Rater Motivation



^a See footnote on next page

Footnote for Figure 5

A. Variables Names

1. Docu - Ability to Document the Appraisal
2. Cred - Credibility of the Rater to the Ratee
3. Devel - Purpose of the Appraisal: Employee Development
4. Salary - Purpose of the Appraisal: Salary Decisions
5. Term - Purpose of the Appraisal: Termination Decisions
6. Promo - Purpose of the Appraisal: Promotion Decisions
7. Interdep - Task Interdependence Among Employees
8. Desire - Rater's Desire to be Liked by the Ratee
9. Paconseq - Expected Consequences of the Appraisal for the Ratee
10. Visibility - Appraisal Visibility
11. Reaction - Reaction of the Ratee to the Appraisal
12. Honesty - Perceived Freedom to be Honest
13. Inflation - Rating Inflation

B. Symbols

1. ξ_i - latent exogenous variables (enclosed in a circle)
2. η_i - latent endogenous variables (enclosed in a circle)
3. x_i - manifest indicators (i.e., questionnaire items) for each latent variable (enclosed in a square)
4. $\lambda_{xi j}$ - path from ξ_j to x_i
5. $\lambda_{yi j}$ - path from η_j to y_i
6. γ_{ij} - path from ξ_j to η_i
7. β_{ij} - path from η_j to η_i

and eta, η , for the five endogenous variables).

Parameters to be estimated are labeled with the appropriate Greek letters. Paths from each of the ξ -variables to the appropriate x-variables are denoted lambda-x (λ_x) while those from the η -variables to the appropriate y-variables are denoted lambda-y (λ_y). Even though there were multiple λ_x or λ_y paths for each exogenous and endogenous variable, respectively, space limitations required that only one of these paths in Figure 5 be labeled for each of the latent variables. Paths between an endogenous and exogenous variable are labeled by gammas (γ) and those between two endogenous variables are labeled by betas (β). Again, following the conventions in causal analysis, each path coefficient has two subscripts, the first being the subscript of the variable that the arrow is pointing to (i.e., the effect) and the second being the subscript of the variable that the arrow is coming from (i.e., the cause). Thus, the paths from the latent variable "docu" (ξ_1) to its four manifest indicators are, respectively, λ_{x11} , λ_{x21} , λ_{x31} , and λ_{x41} , while the path from "paconseq" (η_1) to "reaction" (η_3) is labeled β_{31} .

CHAPTER 4: RESULTS

The results of this study are discussed in three sections. The first section describes the findings from the confirmatory factor analysis used to assess the construct validity of the proposed measurement model. Secondly, the results from the confirmatory analysis of linear structural equations for the hypothesized structural model are discussed. The final section presents the findings from the subsequent exploratory analysis done on the data.

Assessment of the Measurement Model

The measurement model was assessed by confirmatory factor analysis using LISREL VI (Joreskog & Sorbom, 1984). As described previously, the measurement model examined in this study is depicted in Figure 5. The initial confirmatory factor analysis was done using all of the items on the questionnaire. The GFI for this model was .803 the AGFI was .788 and the RMSR was .096. These indices reflect a moderate degree of fit between the data and the model. However, examination of the factor loadings obtained for each item on the relevant latent construct revealed that some of the items were not good indicators of the latent construct they were intended to measure. Items with factor loadings below .30 were dropped from the measurement model. Eliminating these items served both to improve the fit of the model and to increase the stability of the parameter estimates by increasing the ratio of people to items. In most cases, dropping the items with low factor loadings from the scales also resulted in an increase in the reliability estimate (assessed with coefficient alpha) for the scale. The confirmatory factor analysis was then repeated with the smaller set of items. It should be noted that the final set

of parameter estimates for the measurement model could be due to capitalization on chance and thus, should be replicated to increase confidence in their validity.

The means, standard deviations and reliabilities for each of the motivational influences (based on the final set of items used to measure each latent construct) are presented in Table 1. This table also shows the intercorrelations (based on the raw data) between the scales assessing the motivational influences. Table 2 contains the final set of factor loadings for each latent construct (i.e., the lambda matrix in LISREL terminology). The indices of fit for the final measurement model showed a sizeable improvement. Although it is clear that the imposed structure did not account for all of the covariance between the items, the GFI (.880), the AGFI (.864) and the RMSR (.087), taken together, suggest that the measurement model does adequately account for the observed data.

Table 3 presents the correlations between the latent constructs (the phi matrix). The phi matrix was used as the input into LISREL for the assessment of the structural model.

Assessment of the Structural Model

Figure 5 depicted the structural model (in combination with the measurement model) that was tested in this study while Figure 6 shows the structural model with the obtained structural parameters. T-values are calculated to assess the significance of the structural parameters. The t-value for a parameter estimate is calculated by dividing the parameter estimate by its standard error. T-values larger than two are judged to be significantly different from zero at the $\alpha = .05$ level. The structural coefficients for the original

Table 1: Means, Standard Deviations, Reliabilities and Correlations between Scales Measuring the Motivational Influences

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Documentation	4.24	.44	(.61) ²												
2. Credibility	4.00	.51	.17	(.79)											
3. Development	4.05	.48	.32	.20	(.66)										
4. Salary	2.81	1.07	.12	.05	.27	(.90)									
5. Termination	3.11	.76	.12	.09	.19	.51	(.71)								
6. Promotion	3.44	.72	.06	.14	.22	.31	.38	(.70)							
7. Interdeap	4.09	.67	.22	.21	.50	.19	.25	.21	(.77)						
8. Desire	3.62	.67	.06	.62	.13	-.01	-.11	-.02	.13	(.80)					
9. PA Con-sequences ¹	1.71	.66	.01	.31	.31	.12	.03	.05	.20	.37	—				
10. Visibility	2.84	.76	-.06	-.29	.01	-.11	-.09	.01	.09	-.17	-.13	(.81)			
11. Reaction	3.64	.87	.10	.72	.22	.13	.16	.18	.27	.47	.42	-.32	(.92)		
12. Honesty	3.45	.70	.33	.31	.38	.15	.28	.17	.35	.23	.19	-.19	-.26	(.65)	
13. Inflation ¹	.06	.08	-.04	-.19	-.12	-.13	-.17	-.19	-.15	-.04	-.27	.05	-.38	-.23	—

¹These two scales only contain a single item so no reliability estimates are available.

²Coefficient Alpha Reliabilities are in parentheses on the main diagonal.

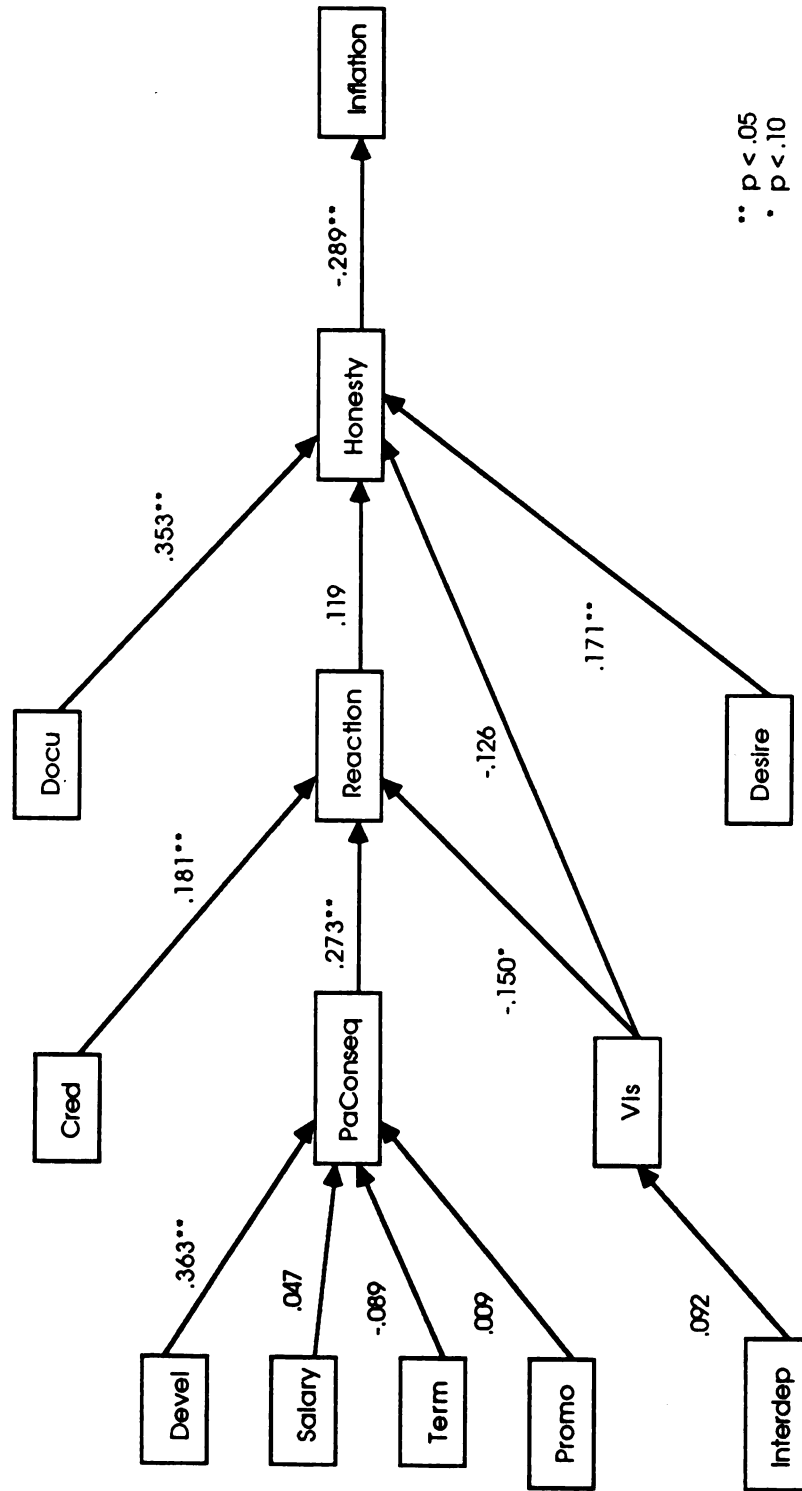
Table 2: Factor Loadings for Confirmatory Factor Analysis -
Lambda Matrix

<u>Item</u>	<u>Factor</u>	<u>Item</u>	<u>Factor</u>
<u>PA Consequences</u>		<u>Development</u>	
	1.000	B16	.350
<u>Visibility</u>		B26	.514
B3	.852	B35	.828
B22	.775	B41	.412
B32	.696	B56	.703
B50	.726	<u>Salary</u>	
<u>Reaction</u>		B2	.741
E2	.629	B9	.647
E3	.712	B14	.782
E9	.696	B31	.848
E11	.589	B55	.862
E17	.553	B60	.845
E24	.600	<u>Termination</u>	
E27	.675	B4	.570
<u>Honesty</u>		B24	.762
B20	.612	B30	.593
B29	.782	B40	.758
B43	.585	B59	.520
B47	.499	<u>Promotion</u>	
<u>Inflation</u>		B7	.727
	1.000	B10	.589
<u>Documentation</u>		B39	.478
B8	.606	B54	.785
B19	.574	<u>Interdependence</u>	
B42	.604	B11	.584
B48	.664	B15	.692
<u>Credibility</u>		B21	.641
E1	.659	B36	.761
E8	.542	B53	.691
E10	.559	B58	.552
E16	.380	<u>Desire</u>	
E22	.401	E6	.859
E26	.602	E7	.464
		E15	.757
		E23	.538
		E26	.466

Table 3: Intercorrelations between Latent Variables - the Phi Matrix

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Documenta- tion	1.00												
2. Credibility	.18	1.00											
3. Development	.37	.21	1.00										
4. Salary	.13	.05	.35	1.00									
5. Termination	.18	.10	.26	.55	1.00								
6. Promotion	.08	.15	.30	.37	.45	1.00							
7. Interdependence	.24	.19	.56	.20	.26	.24	1.00						
8. Desire	.07	.60	.19	.02	-.07	-.01	.17	1.00					
9. PA Con- sequences	-.01	.33	.36	.13	.03	.10	.21	.47	1.00				
10. Visibility	-.07	-.29	.01	-.10	-.10	-.01	.09	-.20	-.15	1.00			
11. Reaction	.11	.32	.22	-.05	.15	.19	.18	.37	.36	-.24	1.00		
12. Honesty	.39	.31	.44	.19	.33	.23	.41	.27	.22	-.22	.25	1.00	
13. Inflation	-.06	-.21	-.17	-.14	-.19	-.26	-.18	-.07	-.27	.06	-.34	-.29	1.00

Figure 6: Structural Parameters for Hypothesized Model of Rater Motivation



model and their associated t-values are presented in Table 4.

The χ^2 for this model was 110.16 with 37 degrees of freedom ($p < .01$), indicating that the observed and reproduced matrices differed significantly from one another. While this is typically considered disconfirming evidence, Joreskog (1978) and others (e.g. Bentler & Bonett, 1980; Tucker & Lewis, 1973) have noted that the χ^2 test is very powerful, particularly with large samples, and thus, tends to reject the model even when differences between the observed and reproduced matrices are small. Thus, as noted earlier, the preferred use of the χ^2 is in comparison to the degrees of freedom, with the recommended ratio being 3:1 or less. In this case, the ratio of χ^2 to degrees of freedom was slightly under 3:1, which suggests a reasonably good fit of the model to the data. Further support for this conclusion comes from examining the other indices of fit. Specifically, the GFI (.883), the AGFI (.711) and the RMSR (.118) all indicate that the hypothesized model accounted for the observed data reasonably well. The specific hypotheses concerning the relationships between the motivational influences are discussed below.

Hypothesis 1

The first hypothesis stated that the purpose of the performance appraisal would be significantly related to the overall magnitude and attractiveness of the consequences of the appraisal for the ratee. Four potential purposes for performance appraisals were included in the model: subordinate development, salary decisions, promotion decisions and termination decisions. This hypothesis was only supported for subordinate development. Subordinate development had a

Table 4: Structural Coefficients and T-Values for the Originally Hypothesized Model

Parameter	Structural Coefficient	T-Value
β_{31}	.273	3.119**
β_{32}	-.150	-1.716*
β_{42}	-.126	-1.464
β_{43}	.119	1.330
β_{54}	-.289	-3.052**
γ_{13}	.363	3.694**
γ_{14}	.047	.413
γ_{15}	-.089	-.781
γ_{16}	.009	.087
γ_{27}	.092	.951
γ_{32}	.181	2.068**
γ_{41}	.353	4.144**
γ_{48}	.171	2.000**

** $p < .05$

* $p < .10$

significant and positive structural coefficient with performance appraisal consequences ($\gamma_{13} = .363$; $t = 3.694$), indicating that the more performance appraisals are used for purposes of subordinate development the more positive the consequences of the appraisal for subordinates. The structural coefficients for salary decisions ($\gamma_{14} = .047$; $t = .413$), termination decisions ($\gamma_{15} = -.089$; $t = -.781$) and promotion decisions ($\gamma_{16} = .009$; $t = .087$) were not significant.

Hypothesis 2

The second hypothesis was that the more attractive the expected consequences of the appraisal for the ratee, the more positive the anticipated subordinate reaction to the appraisal. This hypothesis was supported, as indicated by the significant and positive structural coefficient between performance appraisal consequences and subordinate reaction to the appraisal ($\beta_{31} = .273$; $t = 3.119$).

Hypothesis 3

This hypothesis stated that the more credible raters believe they are to subordinates as feedback sources, the more they will expect subordinates to respond positively to the evaluation. The significant positive structural coefficient between rater credibility and anticipated subordinate reaction ($\gamma_{32} = .181$; $t = 2.068$) demonstrates that this hypothesis was supported.

Hypothesis 4

The fourth hypothesis was that appraisal visibility would be negatively related to the anticipated reaction of the subordinate to the appraisal. There was partial support for this hypothesis. Although the structural coefficient ($\beta_{32} = -.150$) indicated that the relationship between appraisal visibility and anticipated subordinate

reaction was in the hypothesized direction, the coefficient was only marginally significant ($t = -1.716$; $p < .10$).

Hypothesis 5

The fifth hypothesis was that the greater the degree of task interdependence between subordinates in a workgroup the greater the degree of appraisal visibility. The structural coefficient between task interdependence and appraisal visibility was not significant ($\gamma_{27} = .092$; $t = .951$), indicating that this hypothesis was not supported.

Hypothesis 6

This hypothesis suggested that the expected reaction of the subordinate to the appraisal would be positively related to the perceived freedom of the rater to be honest when evaluating performance. This hypothesis was not supported. While the direction of the relationship was as hypothesized ($\beta_{43} = .119$), the path coefficient was not significantly different from zero ($t = 1.330$).

Hypothesis 7

The seventh hypothesis stated that the stronger a rater's desire to be liked by the ratee, the lower his/her perceived freedom to be honest when evaluating performance. Examination of the structural coefficient between desire to be liked and freedom to be honest ($\gamma_{48} = .171$) indicates that although the coefficient was significant ($t = 2.000$), the direction of the relationship was the opposite of that hypothesized. Specifically, the stronger a rater's desire to be liked, the greater his/her perceived freedom to be honest.

Hypothesis 8

The next hypothesis posited a positive relationship between the ability of the rater to document the performance evaluation and his/her perceived freedom to be honest. The structural coefficient for this relationship ($\gamma_{41} = .353$; $t = 4.114$) was significantly different from zero and in the direction hypothesized, indicating that the more raters believe they are able to document their evaluations, the more they feel free to be honest when evaluating performance.

Hypothesis 9

The final cause hypothesized for perceived freedom to be honest was appraisal visibility. Specifically, it was hypothesized that the greater the degree of appraisal visibility, the lower the perceived freedom of the rater to be honest. This hypothesis was not supported. Although the structural coefficient ($\beta_{42} = -.126$) was in the right direction, the coefficient was not significant ($t = -1.464$).

Hypothesis 10

The last hypothesis was that the lower the perceived freedom of the rater to be honest the greater the occurrence of rendering errors (i.e., differences between public and private evaluations). The significant negative structural coefficient between freedom to be honest and rating inflation ($\beta_{54} = -.289$; $t = -3.052$) indicates that this hypothesis was supported.

Exploratory Analysis

The last analyses to be described are the results from an exploratory analysis designed to improve the fit of the data to the model and thus, to provide suggestions for future research. As noted above, although the originally hypothesized model fit the data

reasonably well, examination of the results revealed several possible changes in the model that might improve its fit to the data. The modification indices provided by LISREL for each fixed parameter are useful in identifying possible ways to change the model by relaxing parameters previously fixed to zero. While the specific computation for the modification indices is complicated, it can be shown that the modification index for a given path equals the expected decrease in χ^2 if this particular constraint is relaxed and all other estimated parameters are held fixed at their estimated values (Joreskog & Sorbom, 1984).

Clearly, utilizing the modification indices to suggest changes in the model to improve fit constitutes an exploratory analysis of the data and could result in capitalizing on chance relationships that might be present in the data. Thus, it is important to note that any significant structural coefficient resulting from changes made in the model would need to be cross-validated with another sample.

Joreskog and Sorbom (1984) provide several guidelines for using the modification indices to make changes in the model. First, they recommend making changes sequentially, which means that one parameter should be relaxed at a time. Specifically, they suggest relaxing the fixed parameter with the largest modification index, as long as it is greater than 5.00, and then reassessing the fit of the model. A reduction in χ^2 that is large relative to the change in the degrees of freedom represents a real improvement in the model. In contrast, a drop in χ^2 equal to or smaller than the change in the degrees of freedom probably indicates that the improvement in fit was obtained by

capitalizing on chance. The second recommendation made by Joreskog and Sorbom (1984) is to only make changes that have substantive meaning and which result in parameters that can be interpreted.

Following these procedures, several changes were made in the originally hypothesized model. The structural model summarizing these changes and the resulting structural coefficients are presented in Figure 7. Table 5 presents the structural coefficients and their associated t-values for this modified model⁴. The indices of fit obtained after each sequential change in the model are contained in Table 6. Specific changes made in the model are discussed below.

The first change involved the addition of a path from task interdependence to perceived freedom to be honest (based on a modification index of 13.342). The resulting structural coefficient ($\gamma_{47} = .323$) was significantly different from zero ($t = 3.882$), indicating that a high degree of task interdependence resulted in greater perceived freedom to be honest. Furthermore, the decrease in χ^2 of 14.7 compared to a decrease of 1 in the degrees of freedom was large enough to suggest that the change probably represented a real improvement in the model.

A second change in the model was adding a path from anticipated subordinate reaction to the appraisal to rating inflation (based on a modification index of 9.939). The structural coefficient for this path was negative ($\beta_{53} = -.290$) and significantly different from zero ($t = -3.124$), indicating that when subordinates were expected to react negatively to the appraisal, the amount of rating inflation increased. Again, the ratio of the decrease in χ^2 to the decrease in degrees of freedom (10:1), suggests a substantial improvement in the model.

Figure 7: Structural Parameters for Modified Model of Rater Motivation

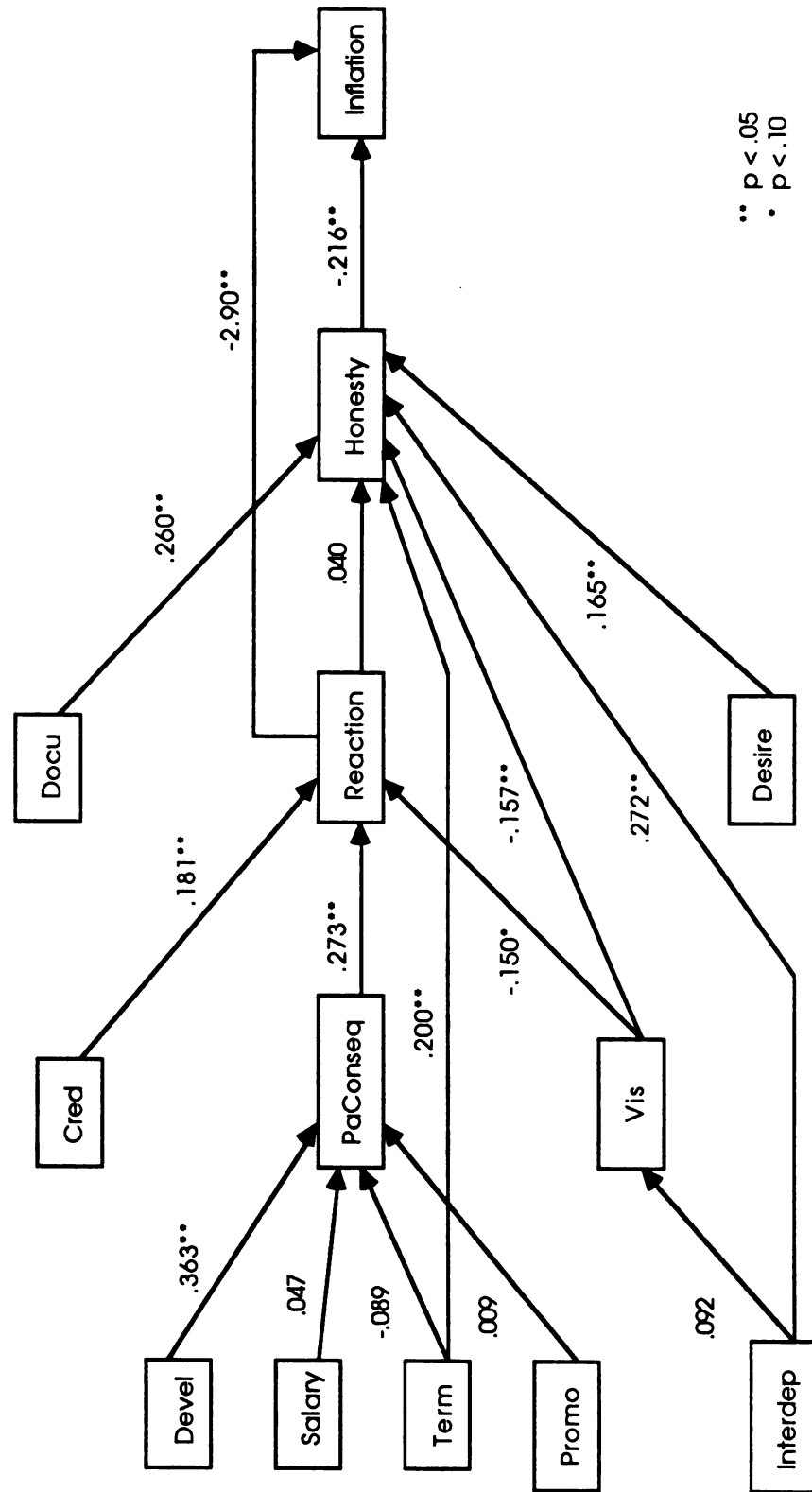


Table 5: Structural Coefficients and T-Values for the Modified Model

Parameter	Structural Coefficient	T-Value
β_{31}	.273	3.119**
β_{32}	-.150	-1.716*
β_{42}	-.157	-1.997**
β_{43}	.044	.542
β_{54}	-.216	-2.362**
β_{53}	-.290	-3.135**
γ_{13}	.363	3.694**
γ_{14}	.047	.413
γ_{15}	-.089	-.781
γ_{16}	.009	.087
γ_{27}	.092	.951
γ_{32}	.181	2.068**
γ_{41}	.260	3.230**
γ_{48}	.165	2.074**
γ_{45}	.200	2.453**
γ_{47}	.272	3.251**

** $p < .05$

* $p < .10$

Table 6: Goodness of Fit Indices for Original Model and Sequential Modifications

Model	Chi-Square	Change in Chi-Square	df	GFI	AGFI	RMSR
Original Model	110.16	---	37	.883	.711	.118
1st Modification	95.46	14.70	36	.895	.735	.105
2nd Modification	85.46	10.00	35	.904	.751	.098
3rd Modification	79.18	6.28	34	.910	.760	.093
Final Model	81.11	---	39	.908	.786	.090

The third change in the model was the addition of a path from using appraisals for termination decisions to perceived freedom to be honest (based on a modification index of 5.695). The structural coefficient for the path from termination decisions was positive and significantly different from zero ($\gamma_{45} = .200$; $t = 2.453$). This indicates that the more appraisals are used for termination decisions, the greater the perceived freedom of the rater to be honest. Finally, the decrease in χ^2 (6.28) relative to the decrease in the degrees of freedom (1) again suggested a sizeable improvement in the model.

The final change in the model involved eliminating all paths that were not significantly different from zero. This model is presented in Figure 8. The structural coefficients and t-values are presented in Table 7. The overall assessment of the degree of fit of the final model to the data indicated some improvement over the originally hypothesized model. Specifically, the χ^2 for the final model was 81.11, with 39 degrees of freedom, a ratio of about 2:1. Furthermore, the GFI (.908) and the AGFI (.786) were both higher while the RMSR (.090) was lower than in the initial model.

Figure 8: Structural Parameters for Final Model of Rater Motivation

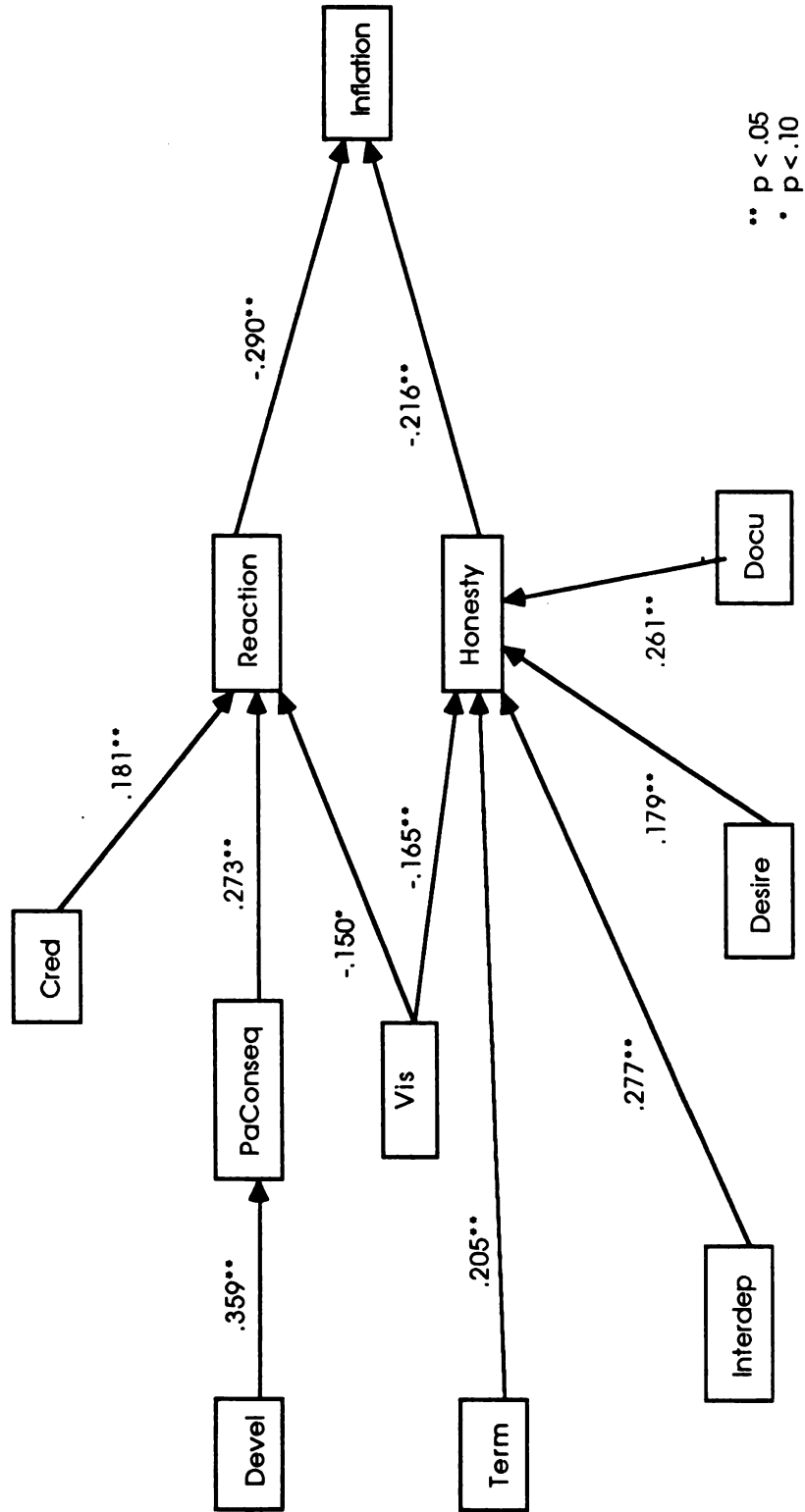


Table 7: Structural Coefficients and T-Values for the Final Model

Parameter	Structural Coefficient	T-Value
β_{31}	.273	3.119**
β_{32}	-.150	-1.717*
β_{42}	-.165	-2.127**
β_{54}	-.216	-2.397**
β_{53}	-.290	-3.151**
γ_{13}	.359	3.960**
γ_{32}	.181	2.067**
γ_{41}	.261	3.246**
γ_{48}	.179	2.265**
γ_{45}	.205	2.513**
γ_{47}	.277	3.320**

** $p < .05$

* $p < .10$

CHAPTER 5: DISCUSSION

Summary and Implications of Findings

The purpose of this study was to gain a better understanding of some of the factors that can influence the accuracy of performance ratings. It has been suggested here and by others (e.g. Banks & Murphy, 1985; Bernardin & Beatty, 1984; DeCotiis & Petit, 1978; Mohrman & Lawler, 1983) that performance rating accuracy has two primary determinants, rater ability and rater motivation. Given the large body of previous research that has examined influences on rater ability (cf. Landy & Farr, 1980), the focus of the present study was on rater motivation. More specifically, a cognitive process model depicting the relationships between a number of potential motivational influences was proposed and submitted to empirical test. Before discussing the results of the analysis of this model, however, several brief, informal observations about the data are presented.

Informal Observations

The first observation concerns the extent to which managers appear to intentionally provide public ratings of performance that are not accurate. Over 70% of the participants in the study provided public evaluations that were higher on one or more performance dimensions than their actual opinions about the employee's performance. At the same time, they indicated their belief that the employee's performance had not changed since the date of the public evaluation, suggesting that true performance change did not account for the difference between the two evaluations. While this was an indirect and unobtrusive measure of rater motivation, many managers frankly admitted during the interview with the researcher that they

intentionally distorted evaluations of employees when they felt there was a "good" reason for doing so. This is consistent with the findings of Longenecker, Gioia & Sims (1987), who also found widespread admission by managers that political considerations and intentional rating distortions frequently entered into performance evaluation processes.

One interesting contrast between the present study and the Longenecker et al. study was in which aspects of the evaluation form were most subject to distortion. Longenecker and his colleagues reported that managers were more likely to distort the overall evaluation of performance than they were their evaluation of any of the specific dimensions on the form. This apparently occurred because the overall rating was believed to be the most important to employees and because this was the evaluation used for administrative decision-making. In the present study, the opposite was found, in that distortions appeared to be more likely on individual dimensions than in the overall evaluation. Nearly 70% of managers did not distort the overall evaluation even though many of them manipulated ratings on specific dimensions.

A possible reason for this difference is that the organization from which the data were collected in this study tended not to use performance evaluations for administrative decision-making so there was less reason to distort the overall evaluation (which would probably be the basis for these decisions). Thus, managers could make an employee feel better by inflating some of the dimension ratings while at the same time maintain a reasonably accurate overall evaluation. Another potential explanation for this finding was the

fact that in the present organization, giving an employee an overall evaluation of "outstanding" (the highest score on the rating form) required attaching a separate written explanation supporting the evaluation. During the interview, many managers stated that they were reluctant to give overall "outstanding" ratings for this reason. On the other hand, "outstanding" on one or more individual dimensions did not require any documentation, making inflation of dimension evaluations less "costly" and thus, more likely.

Another interesting informal observation was that the particular subordinate being evaluated seemed to influence whether or not such distortion took place and thus, distortion was not a general phenomenon that occurred for all the employees evaluated by managers. One assumption underlying the model tested in this study was that managers make decisions about how accurately to rate performance based, at least partially, on characteristics of the specific person being evaluated and their relationship with this person. In conversations with the researcher, a number of managers noted that they were more likely to distort the evaluations of some subordinates than of others. While the reasons for this varied, it is significant that managers considered a number of factors relevant to the specific recipient of the evaluation before determining the extent of distortion of the public rating.

Formal Analyses: Supported Hypotheses

More sophisticated examination of the hypothesized model using linear structural equation analysis showed that the data were generally consistent with the overall model depicting rater cognitive

processes as well as with a number of the specific linkages hypothesized to exist. Results relating to specific linkages in the model are discussed next.

The finding that using appraisals for employee development positively affected the attractiveness of the consequences of the appraisal to the ratee is consistent with discussions of previous authors (e.g. DeCotiis & Petit, 1978; Mohrman & Lawler, 1983; Sharon & Bartlett, 1969) on the effect of appraisal purpose on performance ratings. These researchers have noted that evaluations done for purposes of development are typically less lenient and more accurate than evaluations used for administrative decision-making. To a large extent, this may be due to the fact that evaluations used for developmental purposes are more likely to have positive consequences for ratees, as found in this study. When suggestions for employee performance improvement are made from the perspective of helping the person develop into a more competent and valuable employee they are less likely to have a negative effect on his/her self esteem and, in fact, may even increase self esteem. Furthermore, employees who are concerned about doing a good job are likely to value the opportunity for training and for the development of job skills and abilities as well as the chance to gain a better understanding of their job and role in the organization.

As hypothesized, the attractiveness of the consequences of the appraisal seemed to be important, at least partially because of its relationship with the expected reaction of the ratee to the appraisal. When raters expected the consequences of the appraisal for the ratee to be negative, they were more likely to expect the ratee to react

defensively and nonconstructively to the performance evaluation. A negative reaction was also expected when managers believed that employees would find out how their coworkers were evaluated (i.e., appraisal visibility was high) and when they did not believe they were credible to employees as a feedback source. Furthermore, results from the exploratory analysis revealed a significant negative relationship between expected ratee reaction and the occurrence of rendering errors. Specifically, an anticipated negative reaction resulted in greater positive differences between public and private evaluations (i.e., inflated public evaluations).

The anticipated reaction of the ratee to the appraisal seems to be a pivotal influence on rater motivation to evaluate performance accurately. There are several possible reasons for this. First, how the ratee reacts to the appraisal has long term implications for future interactions between the rater and the ratee (DeCotiis & Petit, 1978; McCall & DeVries, 1977). Raters may (justifiably, perhaps) hesitate to provide negative (but honest) feedback to ratees if they believe the ratee won't accept the feedback, will get hostile or defensive, or possibly even file a grievance, particularly when they know that in the future they will have to work with the employee and keep them motivated to do their job.

Another possible reason for the importance of the anticipated reaction of the ratee to the appraisal stems from the perceived ability of the rater to effectively handle the feedback situation (Bernardin & Beatty, 1984; Bernardin & Buckley, 1981). Intentionally inflating performance ratings may be a defensive strategy for raters

designed to avoid having to cope with the anticipated negative reaction of an employee to the evaluation. Bernardin and his colleagues discuss this tendency within the context of Bandura's (1977) social learning theory. Social learning theory suggests two critical cognitions that could influence a rater's motivation to evaluate performance accurately: (1) an efficacy expectation, which is the conviction that one can successfully execute a behavior in order to produce a particular outcome and (2) an outcome expectation, which concerns the extent to which a person believes that some outcome will result from the behavior. Even if the manager believes that something positive will result from confronting the employee about problems with his/her performance (e.g. the employee will be motivated to improve), if the manager does not believe that he/she would be successful in dealing with the situation (an example of low efficacy expectations), then he/she would probably not be very motivated to rate the employee accurately. As noted by Bandura, Adams, and Beyer (1977):

Strength of convictions in one's own effectiveness determines whether coping behavior will be attempted in the first place. People fear and avoid threatening situations they believe exceed their coping abilities, whereas they behave assuredly when they judge themselves capable of managing situations that otherwise intimidate them (p. 126).

This description of managerial behavior is consistent with the expectancy theory perspective on rater motivation discussed earlier. Low efficacy expectations in social learning theory would be comparable to a low effort---performance expectancy in expectancy theory (Porter & Lawler, 1968).

From a practical point of view the importance of the anticipated

reaction of the employee to the appraisal suggests the need to increase the rater's expectation of personal efficacy for dealing with this reaction (Bernardin & Beatty, 1984). Only when raters believe they are capable of effectively handling this difficult situation will a potential negative reaction by the ratee not result in lower motivation to rate accurately. To this end, Bandura (1977) suggests several sources of information that should serve to increase efficacy expectations. These include performance accomplishments, vicarious experience, verbal persuasion and emotional arousal.

Performance accomplishments are considered to be the most effective way to increase personal efficacy since they are based on experiences of personal mastery. However, vicarious experience can also be useful. In order to increase rater motivation, these two sources of information could be utilized within a typical behavioral modeling training program (e.g. Goldstein & Sorcher, 1974; Latham, Wexley & Pursell, 1975; Spool, 1978). Such a training program might involve having managers view videotapes of people successfully dealing with a difficult ratee during a performance appraisal session, along with a discussion of several key learning points that would help them to execute the appropriate behaviors themselves. Then managers would be given the opportunity to actually practice these behaviors and receive feedback on their effectiveness. This form of training has been found to be successful in teaching managers interpersonal skills (e.g. Carroll, Paine & Ivancevich, 1972) so it also appears to have potential for reducing the occurrence of inflated performance ratings resulting from a rater's low efficacy expectations.

Another practical strategy that might have utility for

eliminating rendering errors resulting from an anticipated defensive or hostile employee reaction might be oriented toward teaching ratees how to receive and deal with negative performance feedback (Bernardin & Beatty, 1984). A behavior modeling program similar to that described above might be helpful in this regard. In addition, developing an appraisal system that employees trust and believe is fair and useful should also be effective in reducing the potential for a negative employee reaction since employees should have more confidence in the evaluations they receive (Bernardin & Beatty, 1984).

The perceived freedom of the rater to be honest when evaluating performance was found to be an important attitudinal precursor of the likelihood of rendering errors occurring. The less that raters felt they could be honest when evaluating an employee's performance, the less honest they actually were, as exemplified by the difference between their public and private evaluations. A rater's perceptions concerning how honest they were able to be was influenced by both the visibility of performance appraisals and the ability of the rater to document and support his/her evaluations with concrete behavioral examples. The former indicates that the more employees in the workgroup find out how each other were evaluated, the lower the perceived freedom of the rater to provide honest evaluations.

A likely explanation for this is that managers believe comparisons among members of the workgroup concerning their evaluations will lead to dissatisfaction, anger and/or perceptions of inequity if they find out another coworker received a higher evaluation than they did. In other words, managers appear to believe

that employees are not able to distinguish good and poor performance and thus, if an honest (and at least partially negative) appraisal is given, employees will believe it is unfair. This belief is consistent with research examining self appraisals, which indicates that they tend to be more lenient than ratings from supervisors (e.g. Kirchner, 1965; Parker, Taylor, Barrett & Martens, 1959).

Nevertheless, one practical implication of this concern is that there is a need for explicit and unambiguous definitions of both the dimensions upon which performance will be evaluated and the standards that will be used in identifying various levels of performance effectiveness. The more that raters and ratees share a common understanding of what constitutes effective performance and the more that raters feel they can apply the standards consistently, the less likely that raters will be concerned about employees feeling they have been evaluated unfairly (even if they find out a coworker received a higher evaluation than they did).

It is interesting to note, in support of this suggestion, that the organization where the data for this study were collected used an evaluation form that consisted of seven to nine general performance dimensions. When asked by the researcher if they felt the form was adequate, most managers reported that it was not and that they disliked it because the dimensions were too vague and the standards (e.g. what constitutes "very good" performance on some dimension) unclear. If the managers using the form felt it was ambiguous then it would be surprising if employees didn't also feel the same way, thereby opening the door for misunderstandings that most managers would probably prefer to avoid (and hence, the lower perceived freedom

to be honest).

The ability of the rater to document and support his/her evaluations of performance was another determinant of the perceived freedom of the rater to be honest when evaluating performance. The more that raters felt they were able to provide concrete behavioral examples to back-up their ratings, the more willing they were to provide an honest performance appraisal. A number of researchers (e.g. Bernardin & Buckley, 1981; Borman, 1979a) have recommended diary-keeping of critical incidents of work performance as a way of improving rater observational skills and thus, rating accuracy. People who have been trained to record critical incidents have been found to provide ratings with less leniency and halo and greater interrater agreement (e.g. Buckley & Bernardin, 1980; Bernardin & Walters, 1977).

The implicit assumption of this research is that diary-keeping resulted in improved rater ability to evaluate performance accurately through better observational skills. The results of the present study, however, suggest an alternative explanation. Specifically, since diary-keeping is likely to increase the extent to which raters feel they are able to document their evaluations this should give them greater confidence in their evaluation and thus, greater perceived freedom to provide an honest assessment of performance. Since perceived freedom to be honest was found reduce the occurrence of rendering errors it appears that being able to document evaluations has an indirect and positive effect on rater motivation and the accuracy of public ratings of performance.

Formal Analyses: Unsupported Hypotheses

In spite of the overall fit for the proposed model, there were several hypothesized linkages in the model that did not receive support from the data. Most notable perhaps, given the large amount of research that has been done on appraisal purpose, was the lack of any relationship between the administrative purpose of performance appraisals (e.g. salary, promotion and termination decisions) and the attractiveness of performance appraisal consequences. One probable explanation for this is that in the organization from which data were collected, performance appraisals were only tangentially related to administrative decision-making in most units.

For example, the organization is unionized at most levels and, thus, union contract, rather than an employee's performance level, determines salary increases for most employees. Further, both promotion and termination decisions also bear little direct relationship to employee performance. Promotions in this organization occur through a somewhat unusual process, which differs substantially from that used in most organizations, because of a highly complicated job classification system. Except in a few units of the organization, it is fairly uncommon for there to be a standard career path in the department or administrative unit into which management selects individuals based on their performance. Rather, promotions in this organizations typically occur through one of two processes: (1) the employee decides he/she wants a job at a higher classification level and applies for such a position in the organization when one is posted or (2) the employee has his/her current job reclassified at a higher level by demonstrating that the duties involved in the position

correspond more closely to those duties typically part of jobs at the higher level. Similarly, managers in the study reported that it was extremely rare for employees to be fired, regardless of their performance level. Given the strong probability of range restriction on appraisal purpose it is likely that the hypotheses involving these variables did not receive an adequate test in this study.

Two other hypotheses, both involving the perceived freedom of the rater to be honest, were also not supported. First, it was hypothesized that the expected reaction of the ratee to the appraisal would have a direct positive effect on the perceived freedom of the rater to be honest. Although this relationship was in the hypothesized direction, it did not reach statistical significance. Furthermore, when the direct linkage between expected ratee reaction and the occurrence of rendering errors was added into the model (during the exploratory analysis), the relationship between ratee reaction and honesty became trivial in magnitude. This suggests that the relationship originally observed between reaction and honesty could have been spurious (i.e., it may have only existed because both reaction and honesty were correlated with the occurrence of rendering errors). In addition, the desire of the rater to be liked by the ratee was hypothesized to be negatively related to the rater's freedom to be honest. However, the exact opposite relationship was found.

At first glance these results seem surprising. However, an examination of the items contained in the original honesty scale suggests a possible explanation for these findings. Specifically, it appears as though two somewhat distinct subscales existed among the

items. Four of the items seemed to be related to the general feelings that raters have about doing performance appraisals (e.g. "I feel uncomfortable telling an employee he/she is not performing well") while three of the items appeared to measure whether or not raters believed it was important to tell the truth when evaluating performance (e.g. "When evaluating an employee's performance, I don't feel that complete honesty is always the best policy"). After doing the confirmatory factor analysis, the items remaining in the scale were primarily those of the former type that dealt with the manager's general feelings concerning performance appraisals. Given this, it is not surprising that the desire of the rater to be liked was positively related to perceived freedom to be honest - when raters want very much to be liked by subordinates they are more likely to feel uncomfortable about doing performance appraisals because of the fear that providing negative feedback will cause employees not to like them.

This would also be consistent with the findings from the exploratory analysis that the extent to which appraisals were used for termination decisions and degree of task interdependence were both significantly and positively related to perceived freedom to be honest. Concerning the former relationship, several managers participating in the study had gone through the process of terminating an employee because of poor performance and all of them agreed that it was a very difficult and time-consuming process that generally created many negative feelings between the manager and his/her employees (due to the involvement of the union in the process). Given the difficulty involved in terminating an employee, it is not surprising that the potential for this kind of situation would make managers feel very

uncomfortable about providing honest performance appraisals. Similarly, when a high degree of task interdependence exists among employees in the workgroup managers appear to feel more uncomfortable about providing honest appraisals, perhaps because of the fact that such honesty might result in conflict among members of the workgroup, which could then lower the overall performance level of the group. While the tentative nature of these findings should be recognized (since they resulted from an exploratory examination of the data), they are consistent with other relationships observed. However, cross-validation with another sample is necessary to have greater confidence in the validity of these conclusions.

On the other hand, the expected reaction of the ratee to the performance appraisal appears to be more strongly related to the importance managers place on being honest and on the extent to which they actually are honest than to their general feelings about doing performance appraisals. When honesty was assessed as the manager's general feelings about performance appraisals, the relationship between expected reaction and honesty was positive but not significant. However, when a separate structural equation analysis was done using only the items assessing the manager's belief that telling the truth is important (the items not used in the original analysis), the expected reaction of the ratee was found to be significantly related to the perceived freedom of the rater to be honest, as hypothesized.

In addition, the exploratory analysis provided results consistent with this. Specifically, the addition of the direct linkage between

reaction to the appraisal and the occurrence of rendering errors improved the overall fit of the model. Thus, anticipated reaction to the appraisal appears to have a stronger effect on the actual behavior of the rater (or measures that are more closely related to actual behavior) than it does on his/her general feelings about doing performance evaluations. Why this occurs is somewhat unclear although it may be related to the amount of variability among raters on their general feelings toward doing performance appraisals. While most people probably feel uncomfortable doing appraisals and providing negative feedback to employees, there may be more variability across people in the extent to which these feelings actually translate into distorted performance ratings.

The final unsupported hypothesis concerned the relationship between task interdependence and appraisal visibility. It was expected that task interdependence would increase appraisal visibility due to more opportunities to discuss or hear about the evaluations of coworkers. While there was a slight tendency for this to be true, the relationship was not significant. It appears that that other factors, besides just opportunity, influence whether or not employees find out how their coworkers were evaluated. For example, it is likely that there are informal norms in workgroups about the extent to which evaluations are discussed and become "public knowledge." In addition, whether or not coworkers have personal friendships with one another is likely to influence whether they talk to one another about evaluations. Thus, while opportunity may be a necessary condition for a high degree of appraisal visibility it does not appear to be a sufficient condition.

Limitations in the Study

In spite of the reasonably strong degree of empirical support found for the model of rater motivation hypothesized in this study, it is necessary to recognize some of the limitations present in this study. The first limitation concerns the somewhat small sample size of the study ($n = 115$). Although this sample size was determined, a priori, to result in an adequate amount of power to detect significant effects, the difficulty with the sample size stems from the fact that parameter estimates are less stable when they are based on a smaller sample size. The potential instability of the parameter estimates indicates a greater need for cross-validating the findings from this study with another sample of people.

Another limitation in the present study concerns the extent to which there are unmeasured causes for any of the endogeneous variables in the model (James, 1980; James et al., 1982). This is typically referred to as the "unmeasured variables problem." Specifically, to the extent that there are relevant unmeasured causes for any of the endogeneous variables included in the structural model, biased estimates of structural coefficients may result. James (1980), has argued that the unmeasured variables problem is unavoidable and thus, the relevant question is not whether or not there is an unmeasured variables problem but rather, to what extent does the problem exist.

James (1980) presents several decision rules for determining the seriousness of an unmeasured variables problem. These involve determining if there are unmeasured causes for any of the endogeneous variables in the model and then assessing the extent to which these

cause are expected to make a unique and nontrivial contribution to one of the effects in the model. Unmeasured causes that are expected to have only small effects and which are linearly dependent on other causes that are measured are not relevant and thus, are not likely to bias parameter estimates. Unfortunately, given the lack of previous empirical research on factors influencing rater motivation to provide accurate performance ratings, it is extremely difficult to determine the seriousness of the unmeasured variables problem in this study.

While clearly only speculative, other potential causes of the anticipated reaction of the ratee to the appraisal might include (1) the extent to which the ratee believes the evaluation was a fair assessment of his/her performance or (2) the extent to which the ratee values personal growth and development. For perceived freedom to be honest, possible unmeasured causes might be: (1) the extent to which raters believe their evaluations will be reviewed by superiors and (2) the extent to which raters want to create a favorable impression with superiors.

For the first cause (i.e., the anticipated reaction of the ratee to the appraisal), the extent to which the ratee believes the evaluation is fair is likely to be at least moderately related to the credibility of the rater as a feedback source. For the other potential unmeasured causes of the two endogeneous variables discussed above, the lack of previous research makes the assessment of linear dependence with other causes only speculative. The same would be true for potential unmeasured causes for the other endogeneous variables in the model. Overall, the seriousness of the unmeasured variables problem in the model tested in this study is indeterminable. Thus,

the magnitude of the structural coefficients found in this study should be accepted with some caution. Future research on other possible causes is clearly needed to resolve this issue.

The issue of the external validity of the findings from this study also deserves mention. The data were all obtained from a single organization that appeared to be unique in some respects (e.g. in the degree to which performance appraisals were used for making administrative decisions) and thus, potentially unrepresentative of other organizations. On the other hand, as noted earlier, the organization selected was large enough and diverse enough to allow collecting data from many semi-autonomous units and a wide variety of types of employees (e.g. skilled, unskilled, educated, uneducated etc.). Since the findings appeared to be consistent across such divergent organizational settings and types of employees it is likely that the phenomena observed in this study are fairly representative of the behavior of people in a variety of situations.

Nevertheless, the extent to which the findings from this study actually do generalize to other types of settings with other groups of people is an important issue that remains for future research to demonstrate. The only way to be truly confident about the generalizability of the results of a study is to empirically, and conceptually (i.e., using different research procedures), replicate the study (Cook & Campbell, 1976).

Suggestions for Future Research

This study examined the relationships between a subset of motivational influences in order to begin to develop an understanding of the complicated process by which individuals are motivated to provide accurate or inaccurate ratings of performance. However, while this study has advanced our understanding of this important aspect of the performance appraisal process, it is clear that there is much yet that we do not know. The component model of performance rating (Landy & Farr, 1980) used earlier to summarize the research examining rater ability suggests some areas for future research on rater motivation.

With respect to the rating instrument, it would be helpful to learn whether or not the appraisal form itself influences the motivation of raters to provide accurate ratings. DeCotiis and Petit (1978) suggested that when raters understand how to use the appraisal instrument and when they perceive it as being adequate (e.g. includes all relevant aspects of job performance, does not include irrelevant job dimensions) and appropriate for its purpose, then they will be more motivated to use the appraisal form accurately. Informal comments by managers participating in the present study would appear to support this as a possible motivational influence, but it remains to be tested empirically.

It might also be interesting to determine if different types of evaluation forms are more subject to intentional distortions. For example, because of the ambiguity of the dimensions contained on a typical trait-oriented rating scale, intentional distortions might be more likely to occur because of the difficulty of detecting them. On the other hand, a behavior or results oriented rating system might be

less subject to this type of distortion since these formats are less ambiguous, require less interpretation, and are easier to verify.

Several rater characteristics might also influence motivation to rate accurately. For example, it is possible that some traditional individual difference variables, such as rater self esteem, locus of control, or need for affiliation, might influence whether or not rating distortion occurs. Similarly, raters who generally have positive beliefs about the nature of other people (Wexley & Youtz, 1985) or who are very people-oriented might be more likely to inflate ratings because they don't want to make employees feel bad by giving them a negative evaluation or because they feel sorry for employees who are having problems. On a more specific level, raters who personally value performance appraisals and believe they are worthwhile might be less likely to intentionally inflate the performance ratings of their employees (as suggested by Longenecker et al., 1987).

It is also plausible that ratee demographic characteristics might influence the extent to which raters are motivated to rate performance accurately. Intentional distortion of performance ratings might be more likely for females or blacks. Along similar lines, dyadic characteristics might also be relevant. For example, rating distortion might be more prevalent in mixed sex or mixed race dyads than in dyads where the manager is evaluating someone of the same race or sex. Research on these ratee characteristics could shed some light onto the reasons for race and sex discrimination in performance evaluations.

There are also a number of potential contextual influences on the motivation to rate accurately. The culture of the organization is one such influence. To the extent that top management in the organization believes in the appraisal process and values employee growth and development then the motivation of managers to rate accurately should be higher. This is similar to the notions of the "political culture" of the organization (Longenecker et al., 1987) and trust in the appraisal process (Bernardin & Beatty, 1984) discussed by others. Other contextual factors might include the extent to which superiors scrutinize and evaluate the performance appraisals of their employees (Kane & Lawler, 1979; Longenecker et al., 1987) and the amount of time pressure managers are under to complete evaluations. Future research is needed to examine these and other potential influences on rater motivation.

From a more practical point of view, research examining the effectiveness of training raters to increase confidence in their ability to deal with defensive or hostile employees during appraisal sessions is also needed. The utility of training ratees to respond more appropriately to negative feedback is also worthy of investigation (Bernardin & Beatty, 1984). As noted earlier, a behavioral modeling approach to training might be an effective method for strengthening rater and ratee interpersonal skills in this type of situation. Training in diary-keeping procedures (e.g. Bernardin & Buckley, 1981) could also be an effective way to increase the perceived ability of raters to document their performance evaluations.

Finally, future research should address reasons for intentional deflation of performance ratings. In this study, about 35% of the

incidences of differences between public and private ratings involved deflations, where public ratings were lower than private ratings, which indicates that deflation is a phenomenon worth some attention. While no attempt was made to identify correlates of deflation in this study, Longenecker et al. (1987) suggested several possible reasons for deflation. These included shocking an employee back to high performance or sending a message to an employee that they should consider quitting their job. However, these and other explanations for deflation need to be examined more systematically.

Conclusion

The present study represents a different focus for performance appraisal research. Until recently, most research on performance appraisal has ignored the impact of the social context in which ratings occur on the accuracy of those ratings. This study demonstrates that such an omission has resulted in a serious gap in our understanding of the performance appraisal process as it occurs in organizational settings. Clearly, rater motivation is an important influence on performance rating accuracy. While the present study is only a preliminary investigation of some of the possible motivational influences, it is a first step toward gaining an understanding of this important phenomenon. Future research needs to focus on the motivational determinants of performance ratings if the goal of accurate ratings is to be achieved.

APPENDICES

APPENDIX A

Evaluation Forms Used by the Organization

NAME
 DEPARTMENT.....
 CLASSIFICATION:.....

FORM SENT
 DATE EMPLOYED.....
 EVALUATION DATE.....

The supervisor's opinion of the employee's performance should be indicated on the scale as objectively as possible. The evaluation must be reviewed and discussed with the employee.

Rating Factors: Consider each factor separately and independently. Base your rating on observable and proven performance.

Outstanding: (O) Indicates an extremely exceptional level of job performance.

Very Good: (V) Performance is beyond normal requirements and competence.

Satisfactory: (S) Fulfills the normal job requirements with some strong points.

Needs Improvement: (N) Performance is below job requirements, but improvement is anticipated.

Unsatisfactory: (U) Job performance level shows a significant limitation that must be improved substantially to be acceptable.

When appropriate, write in "comments" section(s), "No opportunity to observe".

READ ENTIRE REVERSE SIDE BEFORE USING THIS FORM

QUANTITY OF WORK: Consider achievements resulting from personal effort. Also completion of assignments.

O V S N U

Comments: _____

QUALITY OF WORK: Consider accuracy, thoroughness, usability, and dependability of results.

O V S N U

Comments: _____

JOB KNOWLEDGE: Understanding of objectives, duties and responsibilities gained through education, training and experience.

O V S N U

Comments: _____

INITIATIVE AND ORGANIZATION: Ability to be self-starting, efficient, resourceful and creative toward job objectives, duties and responsibilities.

O V S N U

Comments: _____

ATTITUDE AND COOPERATION: Ability and willingness to cooperate with supervisors, co-workers and others, follow directions and rules, accept constructive criticism and exhibit good judgement.

O V S N U

Comments: _____

DEPENDABILITY: Consider regularity of attendance, punctuality, and attention to use of rest periods. Also, meets deadlines.

O V S N U

Comments: _____

CAPACITY TO DEVELOP: Consider the potential to develop skills, improve job performance and assume more responsibility.

O V S N U

Comments: _____

OVERALL EVALUATION: AN OVERALL RATING OF "OUTSTANDING," "NEEDS IMPROVEMENT," OR "UNSATISFACTORY" REQUIRES WRITTEN DOCUMENTATION TO BE INCLUDED WITH THIS EVALUATION.

Consider the employee's total job performance. If a major factor not rated above is considered, please explain.

O V S N U

Explain & Document U, N or O →

A FOLLOW-UP EVALUATION FOR EMPLOYEES RATED "NEEDS IMPROVEMENT" OR "UNSATISFACTORY" IS NORMALLY REQUIRED WITHIN 90 DAYS. THE FOLLOW-UP REVIEW SHOULD BE SCHEDULED FOR _____

SUPERVISOR'S COMMENTS: _____

EMPLOYEE'S COMMENTS: _____

I certify that this evaluation was reviewed with me by my supervisor. My signature does not necessarily indicate my concurrence.

EMPLOYEE

DATE

DEPARTMENT ADMINISTRATOR

DATE

NAME
 DEPARTMENT
 CLASSIFICATION:
 FORM SENT
 DATE EMPLOYED
 EVALUATION PERIOD from to

The performance of the employee should be indicated on the scale as objectively as possible. The evaluation is to be reviewed and discussed with the employee.

READ ENTIRE REVERSE SIDE BEFORE USING THIS FORM

ABILITY TO DEVELOP SUBORDINATES: Extent to which employee utilizes abilities of subordinates and develops them for more responsible duties. O V S N U
 Comments: _____

AFFIRMATIVE ACTION: Significance of goals relative to availability data; achievement of goals. O V S N U
 Comments: _____

ATTITUDE: Initiative, interest and enthusiasm demonstrated. Extent to which employee is a "self-starter" in attaining position objectives. O V S N U
 Comments: _____

COST CONTROL: Extent to which employee maintains cost controls with effective production and operations. O V S N U
 Comments: _____

DEPENDABILITY: Extent to which employee can be counted upon to carry out instructions, have good attendance and fulfill position responsibilities. O V S N U
 Comments: _____

HUMAN RELATIONS: Ability and willingness to effectively work with people. O V S N U
 Comments: _____

JUDGEMENT: Effectiveness of thought, vision, creativity, analysis, sound reasoning and results. O V S N U
 Comments: _____

KNOWLEDGE OF POSITION: Understanding of objectives, duties and responsibilities. O V S N U
 Comments: _____

SUPERVISION: Effectiveness in planning activities of group supervised, and resultant productivity. O V S N U
 Comments: _____

OVERALL EVALUATION: CONSIDER THE EMPLOYEE'S TOTAL JOB PERFORMANCE. IF A MAJOR FACTOR NOT RATED ABOVE IS CONSIDERED, PLEASE EXPLAIN _____

O V S N U
 Explain O, N or U

A FOLLOW-UP EVALUATION FOR EMPLOYEES RATED "NEEDS IMPROVEMENT" OR "UNSATISFACTORY" IS REQUIRED NORMALLY WITHIN 90 DAYS.

THE FOLLOW-UP REVIEW IS SCHEDULED FOR _____

COMMENTS, EVALUATOR: _____

COMMENTS, EMPLOYEE: _____

My signature indicates that this evaluation has been reviewed with me.

EMPLOYEE _____ DATE _____ DEPARTMENT ADMINISTRATOR _____ DATE _____

EVALUATOR _____ DATE _____ DIRECTOR OF PERSONNEL ADMINISTRATION _____ DATE _____

QUALITY OF WORK:	Needs Improvement (errors are frequent) 1 2		Meets Requirements (errors are few) 3 4		Exceeds Requirements (errors are rare) 5 6	
QUANTITY OF WORK:	Insufficient Work 1 2		Completes Required Volume of Work 3 4		Highly Productive 5 6	
JOB KNOWLEDGE:	Limited Knowledge 1 2		Understands Job Duties and Responsibilities 3 4		Excellent Comprehension 5 6	
ADAPTABILITY:	Resists Change 1 2		Adapts Well 3 4		Extremely Flexible 5 6	
DEPENDABILITY:	Unreliable (needs constant supervision) 1 2		Consistent Performance (needs general supervision) 3 4		Highly Reliable (needs minimum supervision) 5 6	
COOPERATION:	Has Difficulty Working With Others 1 2		Generally Cooperative 3 4		Works Well With Others 5 6	
SELF MOTIVATION:	Indifferent, Little Effort to Achieve 1 2		Does Routine Work Without Awaiting Directions 3 4		Seeks Out Work 5 6	
COMMUNICATION:	Poor Communicating Abilities 1 2		Clearly Expresses Self and Understands Others 3 4		Excellent Communication Abilities 5 6	
SAFETY:	Not Safety Conscious 1 2		Generally Observes Safety Rules 3 4		Always Safety Minded 5 6	
CARE OF EQUIPMENT:	Neglects Care of Equipment 1 2		Alert to Condition of Equipment 3 4		Keeps Equipment Clean & in Good Operating Order 5 6	
OVERALL EVALUATION:	Unsatisfactory 1 2		Meets Expectations 3 4		Highly Productive 5 6	

APPENDIX B

Questionnaire Completed by Study Participants

Thank you for your interest in participating in this study. The purpose of the study is to gain a better understanding of how managers such as yourself make performance appraisal ratings. The study is being conducted by Margaret Y. Padgett, a graduate student in the Department of Management, as part of her dissertation and is under the direction of Professor Daniel R. Ilgen, also from Michigan State University. Your participation in the study will consist of two things: (1) completing a questionnaire (this should take approximately 20-30 minutes) and (2) meeting with the researcher for an interview (approximately 30 minutes).

The questionnaire will ask you to provide some background information about yourself and about a randomly selected person working in your unit (the "focal ratee"). THE FOCAL RATEE SHOULD BE THE INDIVIDUAL ON WHOM YOU MOST RECENTLY COMPLETED A PERFORMANCE EVALUATION. Do not provide the full name of this individual. However, as a reminder to yourself, you might find it helpful to write his/her initials in the space provided on the questionnaire. In addition, you will be asked to respond to a number of opinion items about yourself, the focal ratee, and your perceptions of some characteristics of your organization. Keep in mind that we are only interested in your opinion; there are no right or wrong answers to the items. The purpose of the interview will be to give you the opportunity to discuss in more detail some of your personal experiences when conducting performance appraisals. During the interview, you will also be asked to provide an evaluation of the focal ratee. The identity of the focal ratee will, of course, be protected by having the evaluations done anonymously.

All of the information that you provide on the questionnaire and in the interview, including the performance evaluation of the focal ratee, will be kept in strict confidence and will only be seen by the researchers directly involved in the project. At the completion of the project a report will be prepared for Personnel and Employee Relations at Michigan State University. All data in this report, as well as the dissertation report, will be provided in ways that maintain the anonymity of respondents and focal ratees.

To participate in the study, please read the consent statement below and sign and date the form. Be sure to return this form with the questionnaire. Again, thank you in advance for your time and interest in the study.

Consent to Participate

"I agree to participate in this project as described above. I understand that my participation will consist of completing a questionnaire and meeting with the researcher for an interview, with the total time commitment being approximately 60 minutes. I understand that the researchers agree to keep any data that I provide completely confidential. I further recognize that I am free to discontinue my participation in this study at any time without recrimination."

Signature of Participant

Date

PART I

Background Information about Yourself

1. Please indicate your approximate age using the following scale (circle one).

a. 20-25	b. 26-30	c. 31-35	d. 36-40	e. 41-45
f. 46-50	g. 51-55	h. 56-60	i. 61-65	j. 66-70
2. Sex (circle the appropriate response): Male Female
3. Race (circle the appropriate response):

a. Caucasian	b. Black	c. Indian	d. Asian
e. Other (please specify): _____			
4. Length of employment with Michigan State University (in years):
_____ years
5. Time in your current position (in years): _____ years
6. Length of time in a supervisory position (in years): _____ years
7. Number of individuals on whom you currently complete performance evaluations: _____
8. Type of work currently supervised (please circle all that are relevant):

a. clerical	b. technical	c. administrative
d. supervisory	e. operating engineers	f. crafts
g. laborers	h. police	
i. other (please specify) _____		

The remainder of this part of the questionnaire consists of a number of opinion items concerning yourself and your organization. There are two things that you should keep in mind as you are working on the questionnaire. First, for all items on the questionnaire, we are interested in your opinion about what actually exists in the organization rather than how you think things ought to be or should be. Secondly, when the term "workgroup" is used, we are referring to those people that you supervise and on whom you regularly complete performance evaluations. Hence, when the term "coworkers" is used in relation to a person in your unit, it refers to the other people that are also directly supervised by you.

Please read each statement carefully and indicate whether or not you agree with it. There are no right or wrong answers so please respond as honestly as possible. When responding to each item, please use the scale which follows. Place the number corresponding to your opinion for each item in the blank space to the left of each statement. For your convenience, the scale will be reprinted at the top of each page of the questionnaire.

5 - Strongly Agree

4 - Agree

3 - Undecided

2 - Disagree

1 - Strongly Disagree

- _____ 1. After filling out performance evaluations on employees in my department, I am expected to meet with them to discuss their evaluation.
- _____ 2. The performance of individuals, as indicated by their performance appraisal, has little influence on the size of raise that they receive.
- _____ 3. Individuals in my department rarely talk about their performance appraisals with each other.
- _____ 4. In this organization, even individuals who receive low performance ratings are unlikely to be fired.
- _____ 5. Individuals in my workgroup often ask me how they were evaluated compared to their coworkers.
- _____ 6. In this organization, performance appraisals are rarely used to show individuals areas of their performance where improvement is needed.
- _____ 7. Performance appraisal data is given a lot of weight in making promotion decisions.

Please continue to use the scale which follows when responding to each item:

- | | | | | |
|-------------------|-------|-----------|----------|----------------------|
| 5 | 4 | 3 | 2 | 1 |
| Strongly
Agree | Agree | Undecided | Disagree | Strongly
Disagree |
-
- _____ 8. I generally can provide specific examples of things which individuals in my department did during the appraisal period if they ever question my evaluation of their performance.
- _____ 9. Most individuals receive about the same pay increase regardless of their performance level.
- _____ 10. Only people who receive high performance evaluations will be promoted in this organization.
- _____ 11. The jobs which I supervise don't require much interaction among employees.
- _____ 12. Employees in my workgroup typically find out how their coworkers were evaluated by me.
- _____ 13. I have very little trouble being open to my subordinates about their performance.
- _____ 14. Most raises that the people in my unit receive are based very little upon merit.
- _____ 15. The people that I supervise often need to coordinate their work activities with each other.
- _____ 16. Formal performance appraisals provide a means for me to get together with each of the individuals in my department to discuss how to help them become better employees.
- _____ 17. In this organization, wage/salary decisions are based on seniority, such that employees with greater tenure receive higher raises.
- _____ 18. When making decisions about who to terminate, performance appraisal information is rarely examined.
- _____ 19. I often do not feel that I could explain to my employees why I evaluated them as I did.
- _____ 20. I feel uncomfortable telling an employee that he/she is not performing well.
- _____ 21. Individuals in my workgroup need to interact with one another a great deal in performing their jobs.

Please continue to use the scale which follows when responding to each item:

- | | | | | |
|-------------------|-------|-----------|----------|----------------------|
| 5 | 4 | 3 | 2 | 1 |
| Strongly
Agree | Agree | Undecided | Disagree | Strongly
Disagree |
-
- _____ 22. People in my workgroup often compare their performance ratings.
- _____ 23. Sometimes organizational "politics" is a more important factor in determining who gets fired than is a person's job performance.
- _____ 24. It is rare for individuals to be terminated in this organization, regardless of how they perform.
- _____ 25. Promotions are based on who you know rather than on how well you perform.
- _____ 26. In this organization, performance appraisals are not used to provide feedback to employees.
- _____ 27. Performance appraisals are one of the major means by which employees learn how to improve their performance on the job.
- _____ 28. I should keep better records on the performance of people in my department than I do.
- _____ 29. When evaluating an employee's performance, I don't feel that complete honesty is always the best policy.
- _____ 30. When people are terminated in this organization, it is typically those who have been on the job less time, rather than those who perform less well.
- _____ 31. In this organization, the best way to ensure receiving a large wage/salary increase is to receive a good performance appraisal rating.
- _____ 32. Most of the people on whom I do performance appraisals are not very interested in learning how their coworkers were evaluated or rewarded.
- _____ 33. I often base my evaluations of employees on general impressions of their performance rather than on concrete behaviors which I have observed.
- _____ 34. The employees in my department are often not aware of when I do performance appraisals on their coworkers.

Please continue to use the scale which follows when responding to each item:

- | | | | | |
|-------------------|-------|-----------|----------|----------------------|
| 5 | 4 | 3 | 2 | 1 |
| Strongly
Agree | Agree | Undecided | Disagree | Strongly
Disagree |
-
- _____ 35. One of the reasons that we do performance appraisals is to help employees develop their job-related skills and abilities.
- _____ 36. The large amount of interaction needed between members of my department in doing their jobs requires that interpersonal conflicts be dealt with immediately.
- _____ 37. I don't really think it is necessary to discuss my evaluation of an employee's performance with him/her.
- _____ 38. When doing performance evaluations, I feel that it is better for people to know the truth, even if this is unpleasant for either the employee or myself.
- _____ 39. If someone receives several low performance ratings, they are unlikely to ever get promoted to a better position.
- _____ 40. A person's performance on the job is not a major factor considered by those who make termination decisions.
- _____ 41. After completing a performance evaluation on an individual, I turn it in to the appropriate personnel and then forget about it.
- _____ 42. I am generally able to support my evaluations of individuals working in my unit with specific incidents of good and poor performance.
- _____ 43. If there was some way that I could avoid having to approach my employees about problem with their performance I would do it.
- _____ 44. Individuals in my department are aware of the wage/salary increases that their coworkers receive.
- _____ 45. My department often has assignments that require several members of the group to work together in order to complete the project.
- _____ 46. Performance appraisal data is checked regularly by those who make decisions on salary increases.
- _____ 47. It is not difficult for me to discuss the performance of my employees with them.

Please continue to use the scale which follows when responding to each item:

- | | | | | |
|-------------------|-------|-----------|----------|----------------------|
| 5 | 4 | 3 | 2 | 1 |
| Strongly
Agree | Agree | Undecided | Disagree | Strongly
Disagree |
- _____ 48. I am always prepared to back up the performance appraisals of the individuals in my department.
- _____ 49. When an individual in my department is out of the office for a day, his/her absence would make it difficult for others to complete their normal work assignments.
- _____ 50. It would be very unusual for individuals in my unit to mention their performance appraisal ratings to each other.
- _____ 51. I typically keep a file on what each person in my unit has done during the year to help me when I do his/her annual performance appraisal.
- _____ 52. People do not get fired in this organization unless they receive a number of low performance ratings.
- _____ 53. The people in my department do not require much information or assistance from coworkers in order to do their individual jobs effectively.
- _____ 54. People who do not perform well cannot expect to be promoted in this organization.
- _____ 55. Wage/salary decisions are made independently of information about a person's performance evaluations.
- _____ 56. Performance appraisals are used to help employees perform better in the future.
- _____ 57. I would rarely hesitate to tell an employee my true assessment of his/her performance.
- _____ 58. The work areas of individuals in my department are located close together.
- _____ 59. Termination decisions are made only after consulting an employee's performance appraisal records.
- _____ 60. Individuals who receive favorable performance appraisal ratings are likely to be given larger salary increases than those who perform less well.

PART II

This section of the questionnaire deals primarily with the individual from your department selected as the "focal ratee." All of the remaining items on the questionnaire should be answered in relation to this person. Recall that the focal ratee should be the individual on whom you most recently completed a performance evaluation. Be sure NOT to identify the individual by his/her name. However, as a reminder to yourself, you might find it helpful to write his/her initials in the space provided below. First, I would like you to provide some background information about the focal ratee.

Initials of Focal Ratee: _____

1. Sex (circle the appropriate response): Male Female
2. Race (circle the appropriate letter):
 a. Caucasian b. Black c. Indian d. Asian
 e. other (please specify): _____
3. Please indicate the approximate age of the focal ratee using the following scale (circle one).
 a. 20-25 b. 26-30 c. 31-35 d. 36-40 e. 41-45
 f. 46-50 g. 51-55 h. 56-60 i. 61-65 j. 66-70
4. Length of this individual's employment at Michigan State University (in years): _____
5. Amount of time this individual has been in his/her in current position (in years): _____ years
6. Date of his/her most recent performance evaluation:

Now I would like you to respond to several questions about the types of outcomes which you believe the focal ratee might receive as a result of your evaluation of his/her performance. Below is a list of several potential outcomes that might result for the focal ratee because of how you evaluated his/her performance. After each outcome are two blank spaces. Please use them to answer the following two questions about each outcome (see next page).

- (1) GIVEN THIS INDIVIDUAL'S ACTUAL PERFORMANCE LEVEL, HOW LIKELY IS IT THAT EACH OF THESE POTENTIAL OUTCOMES WOULD OCCUR FOR THAT INDIVIDUAL?

Your responses to this item should range from "0%" - will definitely not occur to "100%" - will definitely occur. You may use any percentage between 0% and 100% in your response to this question.

- (2) IN YOUR OPINION, HOW MUCH WOULD THIS INDIVIDUAL LIKE OR DISLIKE RECEIVING EACH OF THESE POTENTIAL OUTCOMES? IN OTHER WORDS, HOW ATTRACTIVE WOULD EACH OUTCOME BE TO THIS PERSON?

Your responses to this item should be made using the following scale:

- 5 - Would like receiving this outcome very much; receiving it is necessary in order for this person to be satisfied with his/her job
- 4 - Would like receiving this outcome but it is not necessary in order for this person to be satisfied with his/her job
- 3 - Would be neutral about receiving this outcome
- 2 - Would dislike receiving this outcome but receiving it wouldn't make this person dissatisfied with his/her job
- 1 - Would dislike receiving this outcome very much; receiving it would make this person dissatisfied with his/her job

EXAMPLE

	(1)	(2)
<u>Outcomes</u>	<u>Likelihood of Outcome Occurring</u>	<u>Attractiveness of Outcome</u>
1. Promotion within the next three years	<u>75%</u>	<u>5</u>
<p>(1) If you believe that, given this person's performance, there is a 75% chance that he/she will be promoted to a higher job level within the next three years, then you would write "75%" in the first blank space to the right of the outcome "promotion within the next three years."</p> <p>(2) If you believe that getting promoted would be necessary in order for this person to be satisfied with his/her job, then you would place a "5" in the second blank space to the right of the outcome "promotion within the next three years."</p>		

Please respond to each of the outcomes in the list which follows in a similar manner.

<u>Outcomes</u>	(1) Likelihood of Outcome Occurring	(2) Attractiveness of Outcome
1. Salary increase	_____	_____
2. Promotion within the next three years	_____	_____
3. Termination of Employment	_____	_____
4. Transfer to an equal but different position (i.e. lateral transfer)	_____	_____
5. Receive remedial training	_____	_____
6. Opportunities for training to prepare for potential advancement	_____	_____
7. Demotion	_____	_____
8. Opportunity to develop job- related skills and abilities	_____	_____
9. Improved self-esteem	_____	_____
10. Lowered self-esteem	_____	_____
11. Better understanding of how to do his/her job	_____	_____

The last section of the questionnaire also concerns your beliefs about the particular person in your workgroup selected to be the focal ratee so your responses should be made with ONLY this person in mind. The items will ask you to indicate the extent to which you feel each statement is true for this individual. As before, the items ask for your opinion so there are no right or wrong answers. Please respond to the following items as honestly as possible using the same scale as you used above. Place the number corresponding to your opinion on the blank space provided to the left of the statement. For your convenience, the rating scale is reprinted below and again at the top of each page.

5 - Strongly Agree

4 - Agree

3 - Undecided

2 - Disagree

1 - Strongly Disagree

- _____ 1. This individual trusts my judgment on work-related matters.
- _____ 2. This person is able to respond constructively to feedback on his/her performance.
- _____ 3. I don't worry about discussing this employee's performance evaluation with him/her because he/she is usually open to any suggestions that I make for improvement.
- _____ 4. It is not important to me that I be liked by this employee.
- _____ 5. This person rarely seeks my help in doing his/her job.
- _____ 6. I really like being around and working with this employee.
- _____ 7. In order to be satisfied with my work, I need to have a good working relationship with this employee.
- _____ 8. In general, I think that this individual values my opinion on most subjects.
- _____ 9. This individual is receptive to receiving feedback on his/her performance even if it is negative.
- _____ 10. I sometimes feel that this individual does not have much respect for my ideas and opinions.
- _____ 11. It is not uncommon for this individual to feel that I am attacking him/her personally if he/she receives less than the highest performance ratings.

Please continue to use the scale which follows when responding to each item:

- | 5 | 4 | 3 | 2 | 1 |
|-------------------|---|-----------|----------|----------------------|
| Strongly
Agree | Agree | Undecided | Disagree | Strongly
Disagree |
| _____ 12. | I would be very surprised if this person ever complained to my superior about a performance appraisal received from me. | | | |
| _____ 13. | I would not go out of my way to try to get this person to like me. | | | |
| _____ 14. | This employee has asked for my advice on nonwork-related issues. | | | |
| _____ 15. | I value the admiration and respect of this person. | | | |
| _____ 16. | I would be very surprised if this person ever followed any advice that I gave him/her. | | | |
| _____ 17. | This employee usually does not have difficulty admitting that he/she has areas of performance on which improvement is needed. | | | |
| _____ 18. | This individual tends to react defensively to negative performance feedback. | | | |
| _____ 19. | This person relies on my advice when making decisions. | | | |
| _____ 20. | This individual is likely to file a grievance against me if unhappy with the performance appraisal received. | | | |
| _____ 21. | I would dislike work if I didn't get along well with this person. | | | |
| _____ 22. | It is not uncommon for this employee to ask my opinion about important work issues. | | | |
| _____ 23. | It wouldn't bother me if this individual didn't like me very much. | | | |
| _____ 24. | This employee values performance feedback as a means for becoming a better performer. | | | |
| _____ 25. | This person's opinion of me as a person or as a manager makes very little difference to me. | | | |
| _____ 26. | This employee does not think very highly of me as a supervisor. | | | |
| _____ 27. | This individual seems to feel threatened by criticism no matter how constructively it is given. | | | |

APPENDIX C

Procedures for Measuring Expected Consequences of the Performance Appraisal for the Ratee

Now I would like you to respond to several questions about the types of outcomes which you believe the focal ratee might receive as a result of your evaluation of his/her performance. Below is a list of several potential outcomes that might result for the focal ratee because of how you evaluated his/her performance. After each outcome are two blank spaces. Please use them to answer the following two questions about each outcome.

- (1) GIVEN THIS INDIVIDUAL'S ACTUAL PERFORMANCE LEVEL, HOW LIKELY IS IT THAT EACH OF THESE POTENTIAL OUTCOMES WOULD OCCUR FOR THAT INDIVIDUAL?

Your responses to this item should range from "0%" - will definitely not occur to "100%" - will definitely occur. You may use any percentage between 0% and 100% in your response to this question.

- (2) IN YOUR OPINION, HOW MUCH WOULD THIS INDIVIDUAL LIKE OR DISLIKE RECEIVING EACH OF THESE POTENTIAL OUTCOMES? IN OTHER WORDS, HOW ATTRACTIVE WOULD EACH OUTCOME BE TO THIS PERSON?

Your responses to this item should be made using the following scale:

- 5 - Would like receiving this outcome very much; receiving it is necessary in order for this person to be satisfied with his/her job
- 4 - Would like receiving this outcome but it is not necessary in order for this person to be satisfied with his/her job
- 3 - Would be neutral about receiving this outcome
- 2 - Would dislike receiving this outcome but receiving it wouldn't make this person dissatisfied with his/her job
- 1 - Would dislike receiving this outcome very much; receiving it would make this person dissatisfied with his/her job

<u>Outcomes</u>	(1)	(2)
<u>Likelihood of Outcome Occurring</u>	<u>Attractiveness of Outcome</u>	
1. Salary increase	_____	_____
2. Promotion within the next three years	_____	_____
3. Termination of Employment	_____	_____
4. Transfer to an equal but different position (i.e. lateral transfer)	_____	_____
5. Receive remedial training	_____	_____
6. Opportunities for training to prepare for potential advancement	_____	_____
7. Demotion	_____	_____
8. Opportunity to develop job- related skills and abilities	_____	_____
9. Improved self-esteem	_____	_____
10. Lowered self-esteem	_____	_____
11. Better understanding of how to do his/her job	_____	_____

APPENDIX D

Questionnaire Items Measuring Each Motivational Influence

Credibility of the Rater to the Ratee

1. This individual trusts my judgment on work-related matters. (Item E1)
- *2. This person rarely seeks my help in doing his/her job. (Item E5)
3. In general, I think that this individual values my opinion on most subjects. (Item E8)
4. I sometimes feel that this individual does not have much respect for my ideas and opinions. (Item E10)
- *5. This employee has asked for my advice on nonwork-related issues. (Item E14)
6. I would be very surprised if this person ever followed any advice that I gave him/her. (Item E16)
- *7. This person relies on my advice when making decisions. (Item E19)
8. It is not uncommon for this employee to ask my opinion about important work issues. (Item E22)
9. This employee does not think very highly of me as a supervisor. (Item E26)

Desire to be Liked by the Ratee

- *1. It is not important to me that I be liked by this employee. (Item E4)
2. I really like being around and working with this employee. (Item E6)
3. In order to be satisfied with my work, I need to have a good working relationship with this employee. (Item E7)
- *4. I would not go out of my way to try to get this person to like me. (Item E13)
5. I value the admiration and respect of this person. (Item E15)
- *6. I would dislike work if I didn't get along well with this person. (Item E21)

7. It wouldn't bother me if this individual didn't like me very much. (Item E23)
8. This person's opinion of me as a person or as a manager makes very little difference to me. (Item E25)

Reaction of the Ratee to the Appraisal

1. This person is able to respond constructively to feedback on his/her performance. (Item E2)
2. I don't worry about discussing this employee's performance evaluation with him/her because he/she is usually open to any suggests that I make for improvement. (Item E3)
3. This individual is receptive to receiving feedback on his/her performance even if it is negative. (Item E9)
4. It is not uncommon for this individual to feel that I am attacking him/her personally if he/she receives less than the highest performance ratings. (Item E11)
- *5. I would be very surprised if this person ever complained to my superior about a performance appraisal received from me. (Item E12)
6. This employee usually does not have difficulty admitting that he/she has areas of performance on which improvement is needed. (Item E17)
- *7. This individual tends to react defensively to negative performance feedback. (Item E18)
- *8. This individual is likely to file a grievance against me if unhappy with the performance appraisal received. (Item E20)
9. This employee values performance feedback as a means for becoming a better performer. (Item E24)
10. This individual seems to feel threatened by criticism no matter how constructively it is given. (Item E27)

Purpose of the Appraisal: Employee Development

- *1. After filling out performance evaluations on employees in my department, I am expected to meet with them to discuss their evaluation. (Item B1)
- *2. In this organization, performance appraisals are rarely used to show individuals areas of their performance where improvement is needed. (Item B6)

3. Formal performance appraisals provide a means for me to get together with each of the individuals in my department to discuss how to help them become better employees. (Item B16)
4. In the organization, performance appraisals are not used to provide feedback to employees. (Item B26)
- *5. Performance appraisals are one of the major means by which employees learn how to improve their performance on the job. (Item B27)
6. One of the reasons we do performance appraisals is to help employees develop their job-related skills and abilities. (Item B35)
- *7. I don't really think it is necessary to discuss my evaluation of an employee's performance with him/her. (Item B37)
8. After completing a performance evaluation on an individual, I turn it in to the appropriate personnel and then forget about it. (Item B41)
9. Performance appraisals are used to help employees perform better in the future. (Item B56)

Purpose of the Appraisal: Salary Decisions

1. The performance of individuals, as indicated by their performance appraisal, has little influence on the size of raise that they receive. (Item B2)
2. Most individuals receive about the same pay increase regardless of their performance level. (Item B9)
3. Most raises that the people in my unit receive are based very little upon merit. (Item B14)
- *4. In this organization, wage/salary decisions are based on seniority, such that employees with greater tenure receive higher raises. (Item B17)
5. In this organizations, the best way to ensure receiving a large wage/salary increase is to receive a good performance appraisal rating. (Item B31)
- *6. Performance appraisal data is checked regularly by those who make decisions on salary increases. (Item B46)
7. Wage/salary decisions are made independently of information about a person's performance evaluations. (Item B55)

8. Individuals who receive favorable performance appraisal ratings are likely to be given larger salary increases than those who perform less well. (Item B60)

Purpose of the Appraisal: Promotion Decisions

1. Performance appraisal data is given a lot of weight in making promotion decisions. (Item B7)
2. Only people who receive high performance evaluations will be promoted in this organization. (Item B10)
- *3. Promotions are based on who you know rather than on how well you perform. (Item B25)
4. If someone receives several low performance ratings, they are unlikely to ever get promoted to a better position. (Item B39)
5. People who do not perform well cannot expect to be promoted in this organization. (Item B54)

Purpose of the Appraisal: Termination Decisions

1. In this organization, even individuals who receive low performance ratings are unlikely to be fired. (Item B4)
- *2. When making decisions about who to terminate, performance appraisal information is rarely examined. (Item B18)
- *3. Sometimes organizational "politics" is a more important factor in determining who gets fired than is a person's job performance. (Item B23)
4. It is rare for individuals to be terminated in this organization, regardless of how they perform. (Item B24)
5. When people are terminated in this organization, it is typically those who have been on the job less time, rather than those who perform less well. (Item B30)
6. A person's performance on the job is not a major factor considered by those who make termination decisions. (Item B40)
- *7. People do not get fired in this organization unless they receive a number of low performance ratings. (Item B52)
8. Termination decisions are made only after consulting an employee's performance appraisal records. (Item B59)

Ability to Document Evaluation

1. I generally can provide specific examples of things which individuals in my department did during the appraisal period if they ever question my evaluation of their performance. (Item B8)
2. I often do not feel that I could explain to my employees why I evaluated them as I did. (Item B19)
- *3. I should keep better records on the performance of people in my department than I do. (Item B28)
- *4. I often base my evaluations of employees on general impressions of their performance rather than on concrete behaviors which I have observed. (Item B33)
5. I am generally able to support my evaluations of individuals working in my unit with specific incidents of good and poor performance. (Item B42)
6. I am always prepared to back up the performance appraisals of the individuals in my department. (Item B48)
- *7. I typically keep a file on what each person in my unit has done during the year to help me when I do his/her annual performance appraisal. (Item B51)

Task Interdependence Among Employees

1. The jobs which I supervise don't require much interaction among employees. (Item B11)
2. The people that I supervise often need to coordinate their work activities with each other. (Item B15)
3. Individuals in my workgroup need to interact with one another a great deal in performing their jobs. (Item B21)
4. The large amount of interaction needed between members of my department in doing their jobs requires that interpersonal conflicts be dealt with immediately. (Item B36)
- *5. My department often has assignments that require several members of the group to work together in order to complete the project. (Item B45)
- *6. When an individual in my department is out of the office for a day, his/her absence would make it difficult for others to complete their normal work assignments. (Item B49)
7. The people in my department do not require much information or assistance from coworkers in order to do their individual jobs effectively. (Item B53)

8. The work areas of individuals in my department are located close together. (Item B58)

Appraisal Visibility

1. Individuals in my department rarely talk about their performance appraisals with each other. (Item B3)
- *2. Individuals in my workgroup often ask me how they were evaluated compared to their coworkers. (Item B5)
- *3. Employees in my workgroup typically find out how their coworkers were evaluated by me. (Item B12)
4. People in my workgroup often compare their performance ratings. (Item B22)
5. Most of the people on whom I do performance appraisals are not very interested in learning how their coworkers were evaluated or rewarded. (Item B32)
- *6. The employees in my department are often not aware of when I do performance appraisals on their coworkers. (Item B34)
- *7. Individuals in my department are aware of the wage/salary increases that their coworkers receive. (Item B44)
8. It would be very unusual for individuals in my unit to mention their performance appraisal ratings to each other. (Item B50)

Perceived Freedom to be Honest

- *1. I have very little trouble being open to my subordinates about their performance. (Item B13)
2. I feel uncomfortable telling an employee that he/she is not performing well. (Item B20)
3. When evaluating an employee's performance, I don't feel that complete honesty is always the best policy. (Item B29)
- *4. When doing performance evaluations, I feel that it is better for people to know the truth, even if this is unpleasant for either the employee or myself. (Item B38)
5. If there was some way that I could avoid having to approach my employees about a problem with their performance I would do it. (Item B43)
6. It is not difficult for me to discuss the performance of my employees with them. (Item B47)

*7. I would rarely hesitate to tell an employee my true assessment of his/her performance. (Item B57)

*indicates item was eliminated from the scales when used in the analyses

FOOTNOTES

FOOTNOTES

¹The standard interview questions asked of all participants are given below:

1. How are performance appraisals done in this organization?
2. What types of information do you look for or consider important when evaluating someone's performance?
3. To what extent are the things you look for determined by the evaluation form used?
4. What sorts of problems or difficulties have you had in doing performance evaluations?
5. What kinds of reactions to the evaluation do you typically get from subordinates?
6. How do you feel about doing performance evaluations? Do you like them, dislike them, or feel indifferent to them?
7. Do you think the evaluation form used by this organization is adequate? Does it cover all the relevant aspects of an employee's performance? If you are unsatisfied with it, how would you change it?
8. Do you think performance evaluations are worthwhile? Do you think your subordinates find them to be worthwhile?

²Although most units of the university use the standard two university appraisal forms, a few units had developed their own forms. Four managers from one such unit participated in this study. The form developed by this unit was similar to the university form, except that it contained ten general dimensions (over half of which coincided with dimensions on the university form) evaluated on a 6-point scale. To make these ratings comparable in standard deviation to the university form, all ratings were converted to their equivalents on a 5-point scale before computing the measure of rendering errors.

³To reduce confusion in presentation, the measurement model depicted in Figure 5 only shows the final number of manifest indicators for each latent construct (based on the results of the initial confirmatory factor analysis) rather than all of the items included on the questionnaire. For the same reason, the error terms for each manifest and latent variable are also excluded from the diagram.

⁴It should be noted that the structural coefficients presented in Figure 7 and Table 5 will differ somewhat from those described in the text. This is due to the fact that the table and figure present the coefficients for a model that includes all three of the additional paths (i.e., the overall modified model) while the text lists the coefficients for each path as it was sequentially added to the model. These two sets of structural coefficients will differ because each time a change in the model is made, other coefficients in the model may also be altered.

LIST OF REFERENCES

LIST OF REFERENCES

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. 2, New York: Academic Press, 267-300.
- Ball, W. J. (1972). The definition of situation: Some theoretical and methodological consequences of taking W. I. Thomas seriously. Journal for the Theory of Social Behaviour, 2, 61-82.
- Bandura, A. (1977). Social Learning Theory. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A., Adams, N. E., and Beyer, J. (1977). Cognitive processes mediating behavioral change. Journal of Personality and Social Psychology, 35, 125-139.
- Banks, C. G. and Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. Personnel Psychology, 38, 335-345.
- Barrett, R. S., Taylor, E. K, Parker, J. W. and Martens, L. (1958). Rating scale content: I. Scale information and supervisory ratings. Personnel Psychology, 11, 333-346.
- Bartlett, C. J. (1983). Would you know a properly motivated performance appraisal if you saw one? In F. Landy, S. Zedeck, and J. Cleveland (Eds.), Performance measurement and theory. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, Vol. 31, 419-456.
- Bentler, P. M. and Bonet, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.
- Berger, P. and Luckman, T. (1966). The social construction of reality. Garden City: Doubleday.
- Bernardin, H. J. (1978). Effects of rater training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 63, 301-308.

- Bernardin, H. J., Alvares, K. M., and Cranny, C. J. (1976). A recomparison of behavioral expectation scales to summated scales. Journal of Applied Psychology, 61, 564-570.
- Bernardin, H. J. and Beatty, R. W. (1984). Performance appraisal: Assessing human behavior at work. Boston, MA: Kent Publishing Co.
- Bernardin, H. J. and Buckley, M. R. (1981). Strategies in rater training. Academy of Management Review, 6, 205-212.
- Bernardin, H. J., Orban, J. A., and Carlyle, J. J. (1981). Performance ratings as a function of trust in appraisal, purpose for appraisal, and rater individual differences. Proceedings of the Academy of Management, 311-315.
- Bernardin, H. J. and Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-65.
- Bernardin, H. J. and Walters, C. S. (1977). The effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 62, 64-69.
- Borman, W. C. (1974). The rating of individuals in organizations: An alternate approach. Organizational Behavior and Human Performance, 12, 105-124.
- Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.
- Borman, W. C. (1979a). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.
- Borman, W. C. (1979b). Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 3, 103-115.
- Borman, W. C. and Dunnette, M. D. (1975). Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 60, 561-565.
- Borman, W. C., Hough, L.M. and Dunnette, M. D. (1976). Development of behaviorally based rating scales for the performance of U. S. Navy recruiters. Navy Personnel Research and Development Center Technical Report TR-76-31.
- Bower, G. H. (1981). Mood and memory. American Psychologist, 36, 120-148.

- Buckley, M. R. and Bernardin, H. J. (1980). An assessment of the components of a rater training program. Paper presented at the annual meeting of the Southeastern Psychological Association, Washington, DC.
- Burnaska, R. F. and Hollmann, T. D. (1974). An empirical comparison of the relative effects of rater response biases on three rating scale formats. Journal of Applied Psychology, 59, 307-312.
- Cafferty, T. P., DeNisi, A. S. and Williams, K. J. (1984). Organization of recall and performance evaluation accuracy for multiple targets. Paper presented at the American Psychological Association meeting, Toronto.
- Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Campbell, J. P. and Pritchard, R. D. (1976). Motivation theory in industrial and organizational psychology. In M. Dunnette (Ed.), Handbook of Industrial-Organizational Psychology. New York: John Wiley, 63-130.
- Cardy, R. L. and Dobbins, G. H. (1986). Affect and appraisal: Liking as an integral dimension in evaluating performance. Journal of Applied Psychology, 71, 672-678.
- Carroll, S. J., Paine, F. P., and Ivancevich, J. M. (1972). The relative effectiveness of training methods: expert opinion and research. Personnel Psychology, 25, 495-509.
- Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.
- Cook, T. D. and Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago, IL: Rand McNally.
- Cox, J. A. and Krumboltz, J. D. (1958). Racial bias in peer ratings of basic airmen. Sociometry, 21, 292-299.
- Crockett, W. H., Mahood, S. and Press, A. N. (1975). Impressions of a speaker as a function of variations in the cognitive characteristics of the perceiver and the message. Journal of Personality, 43, 168-178.
- Dayal, I. (1969). Some issues in performance appraisal. Personnel Administration, 32, 27-30.
- DeCotiis, T. and Petit, A. (1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 3, 635-646.

- DeNisi, A. S., Cafferty, T. P. and Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.
- DeNisi, A. S., Williams, K. J., Cafferty, T. P. and Meglino, B. M. (1985). Cognitive processes and performance appraisals: The role of information acquisition and organization. In R. Cardy (Chair), Information processing research in performance appraisal: A consideration of the approach, findings and implications. Symposium presented at the meeting of the Academy of Management in San Diego, CA.
- DeNisi, A. S. and Stevens, G. E. (1981). Profiles of performance, performance evaluations, and personnel decisions. Academy of Management Journal, 24, 592-602.
- Drucker, P. F. (1954). The practice of management. New York: Harper & Row.
- Duncan, O. D. (1975). Introduction to structural equation models. New York: Academic Press.
- Favero, J. L. and Ilgen, D. R. (1983). The effects of rater characteristics on rater performance appraisal behavior. Office of Naval Research, Technical Report 83-5.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.
- Fisher, C. D. (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation. Journal of Applied Psychology, 64, 533-540.
- Ford, J. K., Kraiger, K., and Schechtman, S. L. (1986). Study of race effects in objective indices and subjective evaluations of performance: A meta-analysis of performance criteria. Psychological Bulletin, 99, 330-337.
- French, J. R. and Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), Studies in social power. Ann Arbor, MI: Institute for Social Research.
- Goldstein, A. P. and Sorcher, M. (1974). Changing supervisory behavior. New York: Pergamon.
- Grey, J. and Kipnis, D. (1976). Untangling the performance appraisal dilemma: The influence of perceived organizational context on evaluative processes. Journal of Applied Psychology, 61, 329-335.

- Hamner, W. C., Kim, J. S., Baird, L., and Bigoness, N. J. (1974). Race and sex as determinants of ratings by potential employers in a simulated work sampling task. Journal of Applied Psychology, 59, 705-711.
- Heilman, M. E. and Guzzo, R. A. (1978). The perceived cause of work success as a mediator of sex discrimination in organizations. Organizational Behavior and Human Performance, 21, 346-357.
- Henemen, R. L. and Wexley, K. N. (1983). The effects of time delay in rating and amount of information observed on performance rating accuracy. Academy of Management Journal, 26, 677-686.
- Holzbach, R. L. (1978). Rater bias in performance ratings: Supervisor, self and peer ratings. Journal of Applied Psychology, 63, 579-588.
- Huse, E. F. (1967). Performance appraisal: A new look. Personnel Administration, 30, 3-5, 16-18.
- Ilgen, D. R. and Feldman, J. M. (1983). Performance appraisal: A process approach. In L. L. Cummings and B. M. Staw (Eds.), Research in Organizational Behavior, Vol. 5, 141-197.
- Ilgen, D. R., Fisher, C. D. and Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64, 349-371.
- Ilgen, D. R., Peterson, R. B., Martin, B. A. and Boeschen, D. A. (1981). Supervisor and subordinate reactions to performance appraisal feedback. Organizational Behavior and Human Decision Processes, 28, 311-330.
- Jacobson, M. B. and Effertz, J. (1974). Sex roles and leadership: Perceptions of the leaders and the led. Organizational Behavior and Human Performance, 12, 383-396.
- James, L. R. (1980). The unmeasured variables problem in path analysis. Journal of Applied Psychology, 65, 415-421.
- James, L. R., Mulaik, S. A. and Brett, J. M. Causal analysis. Beverly Hills: Sage Publications.
- Jeffery, K. M. and Mischel, W. (1979). Effects of purpose on the organization and recall of information in person perception. Journal of Personality, 47, 297-319.
- Joreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.
- Joreskog, K. G. and Sorbom, D. (1984). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood. Mooresville, IN: Scientific Software.

- Kane, J. S. and Lawler, E. E. (1979). Performance appraisal effectiveness: Its assessment and determinants. In B. M. Staw (Ed.), Research in Organizational Behavior, Vol. 1, 425-478.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), Nebraska symposium on motivation. Lincoln: University of Nebraska Press, Vol. 15.
- Kenny, D. A. (1979). Correlation and Causality. New York: John Wiley & Sons.
- Kirchner, W. K. (1965). Relationships between supervisory and subordinate ratings for technical personnel. Journal of Industrial Psychology, 3, 57-60.
- Klein, S. M., Kraut, A. K. and Wolfson, A. (1971). Employee reactions to attitude survey feedback: A study of the impact of structure and process. Administrative Science Quarterly, 16, 497-514.
- Klimoski, R. J. and London, M. (1974). Role of the rater in performance appraisal. Journal of Applied Psychology, 59, 445-451.
- Knowlton, W. A. and Mitchell, T. R. (1980). Effects of causal attributions on a supervisor's evaluation of subordinate performance. Journal of Applied Psychology, 65, 459-466.
- Komacki, J. (1981). Behavioral measurement: Toward solving the criterion-problem. Paper presented at the American Psychological Association Convention, Los Angeles, August, 1981.
- Kraiger, K. and Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. Journal of Applied Psychology, 70, 56-65.
- Lance, C. E. and Woehr, D. J. (1986). Statistical control of halo: Clarification from two cognitive models of the performance appraisal process. Journal of Applied Psychology, 71, 679-685.
- Landy, F. J. and Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.
- Landy, F. J., Farr, J. L., Saal, F. G. and Freytag, W. (1976). Behaviorally anchored scales for rating the performance of police officers. Journal of Applied Psychology, 61, 752-758.
- Latham, G. P. and Wexley, K. N. (1977). Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 30, 255-268.

- Latham, G. P. and Wexley, K. N. (1981). Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley Publishing Co.
- Latham, G. P., Wexley, K. N. and Pursell, E. D. (1975). Training raters to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 60, 550-555.
- Lawler, E. E. (1967). The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 51, 369-381.
- Leskovec, E. (1967). A guide for discussing the performance appraisal. Personnel Journal, 46, 150-152.
- Lewin, K. (1935). A dynamic theory of personality. New York: McGraw-Hill.
- Liden, R. C. and Mitchell, T. R. (1983). The effects of group interdependence on supervisor performance evaluations. Personnel Psychology, 36, 289-299.
- Locher, A. H. and Teel, K. S. (1977). Performance appraisal: A survey of current practices. Personnel Journal, 56, 245-247; 254.
- Longenecker, C. O., Gioia, D. A. and Sims, H. P. (1987). Behind the mask: The politics of employee appraisal. The Academy of Management Executive, 1, 183-193.
- Lord, R. B., Foti, R. J. and Phillips, J. S. (1982). A theory of leadership categorization. In J. S. Hunt, W. Sekaran, & C. Schreisheim (Eds.), Leadership: Beyond established views. Carbondale, IL: Southern Illinois University Press, 104-121.
- Madden, J. M. and Bourdon, R. D. (1964). Effects of variations in rating scale format on judgment. Journal of Applied Psychology, 48, 147-151.
- Matte, W. E. (1982). An experimental investigation of information search in performance appraisal. In A. DeNisi (Chair), Cognitive approaches to the study of performance appraisal: Empirical findings. Symposium presented at the meeting of the American Psychological Association, Washington, D. C.
- McCall, M. W. and DeVries, D. L. (1977). Appraisal in context: Clashing with organizational realities. Technical Report #4, Center for Creative Leadership.
- McClelland, D. C. and Burnham, D. H. (1976). Power is the great motivator. Harvard Business Review, 54, 100-110.

- McIntyre, R. M., Smith, D. E. and Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.
- Meyer, H. H., Kay, E. and French, J. R. (1965). Split roles in performance appraisal. Harvard Business Review, 43, 123-129.
- Mitchell, T. R. (1974). Expectancy models of job satisfaction, occupational preference and effort: A theoretical, methodological, and empirical appraisal. Psychological Bulletin, 81, 1053-1077.
- Mitchell, T. R. (1983). The effects of social, task, and situational factors on motivation, performance and appraisal. In F. Landy, S. Zedeck, and J. Cleveland (Eds.), Performance measurement and theory. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 39-59.
- Mitchell, T. R. and Liden, R. C. (1982). The effects of the social context on performance evaluations. Organizational Behavior and Human Performance, 29, 241-256.
- Mitchell, T. R. and Wood, R. E. (1980). Supervisor's responses to subordinate poor performance: A test of an attributional model. Organizational Behavior and Human Performance, 25, 123-138.
- Mobley, W. H., Horner, S. D., and Hollingsworth, A. T. (1978). An evaluation of precursors of hospital employee turnover. Journal of Applied Psychology, 63, 408-414.
- Mohrman, A. M. and Lawler, E. E. (1983). Motivation and performance appraisal behavior. In F. Landy, S. Zedeck, and J. Cleveland (Eds.), Performance measurement and theory. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Mount, M. K. and Thompson, D. E. (1987). Cognitive categorization and quality of performance ratings. Journal of Applied Psychology, 72, 240-246.
- Murphy, K. R. and Balzer, W. K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. Journal of Applied Psychology, 71, 39-44.
- Murphy, K. R., Martin, C. and Garcia, M. (1982). Do behavioral observation scales measure observation? Journal of Applied Psychology, 67, 562-567.
- Nathan, B. R. and Lord, R. B. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68, 102-114.

- Nieva, V. F. and Gutek, B. A. (1980). Sex effects on evaluation. Academy of Management Review, 5, 267-276.
- Nunnally, J. C. (1978). Psychometric theory. New York: McGraw-Hill.
- Padgett, M. Y. and Ilgen, D. R. (1988). The effect of ratee performance characteristics on alternative measures of rater accuracy. Organizational Behavior and Human Decision Processes. In press.
- Park, O. S., Sims, H. P., and Motowidlo, S. J. (1986). Affect in organizations. In H. Sims, D. Gioia, and Associates (Eds.), The thinking organization. San Francisco: Jossey-Bass Publishers.
- Parker, J. W., Taylor, E. K., Barrett, R. S. and Martens, L. (1959). Rating scale content: 3. Relationship between supervisory and self-rating. Personnel Psychology, 12, 49-63.
- Paterson, D. G. (1922). The Scott Company graphic rating scale. Journal of Personnel Research, 1, 361-176.
- Porter, L. W. and Lawler, E. E. (1968). Managerial attitudes and performance. Homewood, IL: Richard D. Irwin.
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.
- Rosen, B. and Jerdee, T. H. (1973). The influence of sex role stereotypes on evaluations of male and female supervisory behavior. Journal of Applied Psychology, 57, 44-48.
- Rowe, K. H. (1964). An appraisal of appraisals. Journal of Management Studies, Vol. 1.
- Salancik, G. R. and Pfeffer, J. (1978). A social information-processing approach to job attitudes and task design. Administrative Science Quarterly, 23, 224-253.
- Sharon, A. T. and Bartlett, C. J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 22, 251-263.
- Schmidt, F. L. and Johnson, R. H. (1973). Effect of race on peer ratings in an industrial setting. Journal of Applied Psychology, 57, 237-241.
- Schmitt, N. and Hill, T. (1977). Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. Journal of Applied Psychology, 62, 261-264.

- Schmitt, N. and Lappin, M. (1980). Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. Journal of Applied Psychology, 62, 261-264.
- Scott, W. E. and Hamner, W. C. (1975). The influence of variations in performance profiles on the performance evaluation process: An examination of the validity of the criterion. Organizational Behavior and Human Performance, 14, 360-370.
- Silverman, D. (1971). The theory of organisations: A sociological framework. New York: Basic Books, Inc., Publishers.
- Smith, P. C. and Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.
- Spool, M. D. (1978). Training programs for observers of behavior: A review. Personnel Psychology, 31, 853-888.
- Taft, R. (1955). The ability to judge people. Psychological Bulletin, 52, 1-23.
- Terborg, J. R. and Ilgen, D. R. (1975). A theoretical approach to sex discrimination in traditionally masculine occupations. Organizational Behavior and Human Performance, 13, 352-376.
- Thayer, F. C. (1981). Civil service reform and performance appraisal: A policy disaster. Public Personnel Management, 10, 20-28.
- Thomas, W. I. (1928). The child in America. New York: Knopf.
- Thornton, G. C. III (1968). The relationship between supervisor and self appraisals of executive performance. Personnel Psychology, 21, 441-455.
- Thornton, G. C. III (1980). Psychometric properties of self appraisals of job performance. Personnel Psychology, 33, 263-271.
- Tucker, L. R. and Lewis, C. A. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.
- Tuckman, B. W. and Oliver, W. F. (1968). Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 59, 297-301.
- Vroom, V. H. (1964). Work and motivation. New York: John Wiley.
- Walsman, D. A. and Thornton, G. C. III (1978). A comparison of supervisors' self appraisals and their administrators' appraisals. Medical Group Management, 46, 42-46.

- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S. and Rosenbaum, R. M. (1971). Perceiving the causes of success and failure. Morristown, NJ: General Learning Press.
- Wexley, K. N. and Klimoski, R. (1984). Performance appraisal: An update. In K. Rowland & G. Ferris (Eds.), Research in Personnel and Human Resources Management, Vol. 2, Greenwich, CT: JAI Press, Inc., p. 35-79.
- Wexley, K. N. and Snell, S. A. (1987). Managerial power: A neglected aspect of the performance appraisal interview. Journal of Business Research, 15, 45-54.
- Wexley, K. N. and Youtz, M. A. (1985). Rater beliefs about others: Their effect on rating errors and rater accuracy. Journal of Occupational Psychology, 58, 265-275.
- White, M. C., Crino, M. D. and DeSanctis, G. L. (1981). A critical review of female performance, performance training and organizational initiatives designed to aid women in the work-role environment. Personnel Psychology, 34, 227-248.
- Williams, K. J., DeNisi, A. S., Blencoe, A. G. and Cafferty, T. P. (1985). The role of appraisal purpose: Effects of purpose on information acquisition and utilization. Organizational Behavior and Human Decision Processes, 35, 314-339.
- Wyer, R. S., Srull, T. K., Gordon, S. E. and Hartwick, J. (1982). Effects of processing objectives on the recall of prose material. Journal of Personality and Social Psychology, 43, 674-688.
- Zedeck, S. and Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67, 752-758.
- Zedeck, S., Imperato, N., Krausz, M. and Oleno, T. (1974). Development of behaviorally anchored rating scales as a function of organizational level. Journal of Applied Psychology, 59, 249-252.

MICHIGAN STATE UNIV. LIBRARIES



31293000562649