PURPOSE OF APPRAISAL REVISITED: AN EXAMINATION OF THE RELATIONSHIP BETWEEN PURPOSE AND CHARACTERISTICS OF PERFORMANCE RATINGS

by

. JAWAHAR I. MOHAMMED

Honors Post Graduate Diploma
in
Personnel Management and Industrial Relations
Madras School of Social Work
Madras, India
1988

Master of Arts
in
Industrial/Organizational Psychology
University of Tulsa
Tulsa, Oklahoma
1990

Submitted to the faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirement for
the Degree of
DOCTOR OF PHILOSOPHY
May, 1994

OKLAHOMA STATE UNIVERSITY

PURPOSE OF APPRAISAL REVISITED: AN EXAMINATION OF THE RELATIONSHIP BETWEEN PURPOSE AND CHARACTERISTICS OF PERFORMANCE RATINGS

Thesis Approved:

Thesis Adviser

Dean of Graduate College

ACKNOWLEDGEMENTS

Several people deserve special recognition for their role in this dissertation and in my doctoral program at Oklahoma State University. Besides championing my admission to the doctoral program, Professor Thomas H. Stone has mentored my professional and personal development from many capacities including those of chairperson (of both doctoral program and dissertation committee) and co-author. To describe or even list the numerous insights into human behavior as well as research skills I have learnt from Professor Stone would be an impossible task. I am very grateful to Professor Stone for enlightening me in so many ways. I look forward to collaborating with him on many more research projects.

As the dissertation went through its drafts, and there were more than a few, its form and substance were greatly shaped by the advice and counsel of Professors Thomas H. Stone, and dissertation committee members, Dr. Debra Nelson, Dr. Steven H. Barr, Dr. Kenneth Eastman, and Dr. Terry Bristol. I am very grateful to my committee members who were very generous with their time, comments and suggestions. Their comments and suggestions were instrumental in developing a document far better than the original manuscript.

Besides contributing to this dissertation, Dr. Barr and Dr. Nelson enhanced my knowledge and understanding of research methodology and organizational behavior. Dr. Eastman was very supportive and encouraging and deserves much credit for the successful completion of this

dissertation.

I am very appreciative of Professor Wayne A. Meinhart who in his capacity as chairperson of the Department of Management helped me in several ways during the course of my education at Oklahoma State University. I am very indebted to Professor Donald D. Bowen (University of Tulsa) for his interest, guidance, and continued support. I will remain grateful to all those who played a role in this dissertation and in my doctoral program.

Finally, I am very grateful to my parents for teaching me the value of hardwork, integrity and other important virtues. I will always remember the numerous sacrifices they made to provide me with the opportunity to pursue higher education. I intend to make their efforts worthwhile by utilizing my knowledge and skills to contribute to the advancement of science and society.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter One

Introduction

This chapter outlines a dissertation that will investigate the influence of certain important characteristics of the rater, and the appraisal context on the characteristics of performance appraisal ratings and two very important personnel decisions: pay raises and training. This introductory chapter will begin with a summary of the research problem. Following an outline of the dissertation objectives, theoretical and practical implications of the project will be described.

The Research Problem

Organizational researchers and practitioners consider performance appraisal to be an important managerial tool that can be used to enhance effectiveness of individuals, groups, and organizations. A survey by Locher and Teel (1988) reported that organizations in North America continue to attach increasing importance to performance appraisals. The basis for the importance attached to appraisals can be seen in the myriad of purposes for which they are used (Cleveland, Murphy & Williams, 1989; Landy & Farr, 1980; Lawler, 1988). The use of performance appraisals by companies for more than one purpose has increased from 11% in 1977 to 30% in 1988 (Locher & Teel, 1988). This increase reflects widespread recognition that performance appraisals, conducted effectively, can increase employee productivity and decrease the organization's cost (Latham, Skarlicki, Irvine & Siegel, in press).

However, the effectiveness of appraisals depends on the accuracy of appraisal ratings. The accuracy of appraisal ratings are significantly influenced by the appraisal instrument, the ability of the rater to provide accurate ratings, and the motivation of the rater to do so. However, research over the last 60 years or so has focused almost exclusively on the rating instrument and the ability of the rater to provide accurate ratings.

By the late 70s, researchers (e.g. Smith, 1976) noted that variations in the instrument had only a slight impact on the accuracy of performance ratings. After an exhaustive review of the performance appraisal literature, Landy and Farr (1980) concluded that different appraisal instruments account for only 4 to 8% of variance in performance ratings (see also Landy & Farr, 1983) suggesting that the search for the perfect appraisal instrument that would generate accurate ratings is futile.

Research directed toward improving accuracy by enhancing rater's ability to provide accurate ratings has focused on perceptual and cognitive processes of raters. One stream of research has endeavored to increase accuracy through 'rater training' to reduce appraisal errors, improve observation skills, and use decision aids such as behavioral diaries. Research along these lines has not been very fruitful and suggestions based on such research have had very little utility for the practitioner (Balzer, 1986; Bernardin & Pence, 1980; Murphy, Martin & Garcia, 1982). A second stream, predominantly cognitive in orientation, was triggered by Feldman's (1981) seminal model of the rating preocess. This stream of research (Ilgen & Feldman, 1983; DeNisi, Cafferty &

Meglino, 1984; DeNisi & Williams, 1988; Feldman, 1986) has focused on understanding how appraisal judgments are formed and retained for use in appraisals. Although research on cognitive processes in performance appraisal may advance our understanding of human judgmental processes, it has not yet led to significant advances in the practice of performance appraisal. Indeed, very few applications of this approach have been suggested (see DeNisi & Williams, 1988) and fewer yet applied (Banks & Murphy, 1985; see also Bretz, Milkovich & Read, 1992). Furthermore, cognitive processing research has generally focused exclusively upon rater ability, and has neglected the study of motivational considerations.

Since several important organizational decisions/activities such as pay raises, promotions, transfers, performance feedback, evaluation of selection and training programs are based on performance appraisal ratings, it is imperative that these ratings be as accurate as possible. Yet, despite the important role that appraisals play in organizations, it is clear that we still do not know much about the appraisal process. In spite of the voluminous amount of research done on appraisals, little has been done to improve the accuracy of ratings, and the rating problems such as leniency have shown themselves to be extremely resistant to efforts to eliminate them (see Bernardin & Villanova, 1986). Why do these problems persist? Perhaps the answer lies in the approaches that have generally been adopted to deal with them. If one were to examine the performance appraisal research conducted over the past 60 years or so (e.g. Bernardin & Beatty, 1984; Bretz et al, 1992; DeNisi et al, 1984; DeNisi & Williams, 1988; Feldman, 1981, Ilgen &

Feldman, 1983; Landy & Farr, 1980, 1983; Latham et al, in press), it would be clear that the majority of this research has focused on either the development of "better" rating instruments, or the training of raters to help them avoid rating errors, or cognitive distortions affecting stages of the information processing sequence. Unfortunately, as mentioned earlier, the failure of these approaches to lead to any real improvement in rating accuracy has been amply documented (see for example, Landy & Farr, 1980, 1983; Bernardin & Pence, 1980; Fisher, 1989; Banks & Murphy, 1985).

These streams of research have generally failed to consider motivational issues confronting the rater. Contextual factors have the potential to influence ability as well as motivation of raters to provide accurate ratings. Such contextual factors include participation by the ratee, timing and frequency of appraisals, consequences of providing accurate ratings for the rater and purpose(s) for which appraisal ratings will be used. This dissertation focuses on such contextual factors. The extant literature on appraisal purpose has generated contradictory results such that the relationship between appraisal purpose and leniency/severity as well as accuracy of ratings is not clear. It is important to understand and resolve these inconsistencies for several reasons including the following. First, inconsistencies are always of theoretical interest to the researcher. Second, as a boundary variable, appraisal purpose has the potential to limit the external validity of performance appraisal research as performance ratings obtained for research purposes may be more or less accurate and/or lenient (severe) than those obtained for administrative

purposes. Finally, since organizations use appraisal ratings for making several important administrative and personnel decisions, suggestions based on ratings for research purposes are likely to be of little value to the practitioner. Therefore, these inconsistencies have theoretical importance as well as practical relevance and hence need to be addressed.

Dissertation Objectives

As mentioned in the previous section, most research concerned with the accuracy of performance ratings has focused on the rating instrument and the ability of the rater to provide accurate ratings. This dissertation, on the other hand, focuses on motivational issues confronting the raters and draws on relevant theory (DeNisi et al, 1984; Heneman, Moore & Wexley, 1987; DeCotiis & Petit, 1978; Ilgen & Feldman, 1983; Landy & Farr, 1980; Wherry, 1952) and research (e.g. Bernardin, Orban & Carlyle, 1981; Bernardin, Abbott, & Cooper, 1985; Berkshire & Highland, 1953; Gmelch & Glasman, 1977; McIntyre, Smith & Hassett, 1984; Sharon & Bartlett, 1969; Taylor & Wherry, 1951; Williams, DeNisi, Blencoe & Cafferty, 1985; Williams, DeNisi, Meglino & Cafferty, 1986; Longenecker, Sims & Gioia, 1987; Zedeck & Cascio, 1982) that suggests that motivational issues such as the purpose for which appraisal ratings are to be used may affect the accuracy of performance ratings. Since performance ratings are used for several purposes (Cleveland et al, 1989; Locher & Teel, 1988), one objective of this dissertation is to investigate how the purpose of the appraisal affects leniency and accuracy of performance ratings as well as two important personnel decisions: recommending pay raises and recommending subordinates for

training. Previous studies investigating the effect of purpose have reached contradictory results. The contradictory findings reported may be due to the failure to consider the motivational effect of appraisal purpose and individual differences among raters. Rater's perception of consequences of ratings for the ratee as well as himself/herself is likely to elicit motivational concerns associated with appraisal purpose. Rater's perception of consequences of ratings and hence motivation to provide accurate ratings may be influenced by varying the amount of funds in the merit-raise budget. Therefore, besides purpose, this dissertation also focuses on the impact of merit-raise budget (a situational factor) on the accuracy of performance appraisal ratings and personnel decisions contingent upon those ratings. Additionally, individual differences in self-monitoring are expected to moderate the influence of 'purpose' and 'budget constraints' on the characteristics of performance ratings and subsequent personnel decisions. Thus, the dissertation will also examine the moderating effect of self-monitoring. The prototypic high self-monitor is one who, out of common concern for the situational and interpersonal appropriateness of his or her social behavior, is particularly sensitive to the expression and self-presentation of relevant others in social situations and uses these cues as guidelines for monitoring (that is, regulating and controlling) his or her own verbal and nonverbal self-presentation (Snyder, 1979, p.89). Low self-monitors on the other hand, lack either the ability or the motivation to do so.

This dissertation synthesizes theory and research relating to these three themes: purpose, budget constraints and rater self-

monitoring and will examine their influence on rating accuracy and important personnel decisions in a laboratory study. This endeavor has important theoretical and practical implications.

Implications for Theory

Focusing on the motivational issues confronting the rater has important theoretical implications. A number of researchers have suggested that purpose of appraisal should be considered as part of any model of the appraisal process (DeCotiis & Petit, 1978; Wherry, 1952). The major contribution of DeCotiis and Petit's (1978) model was the significance accorded to rater motivation. They proposed that rater motivation, rater ability and the availability of judgmental norms were the major determinants of rater accuracy. Even predominantly cognitive models of the appraisal process have emphasized the significance of contextual factors that address motivational issues confronting the rater such as purpose of appraisal. For example, Landy and Farr (1980), identified purpose of rating as a significant component in the rating process. Ilgen and Feldman (1983) expanded Feldman's (1981) model and emphasized the significance of organizational contextual variables to the rating process. Appraisal purpose is also an integral part of DeNisi, Cafferty and Meglino's (1984) cognitive model of the appraisal rating process. An important feature of DeNisi et al's (1984) model is the expanded role of purpose of appraisal. These researchers emphasized the cognitive as well as the motivational influence of appraisal purpose on characteristics of performance ratings. Another recent model of the rating process presented by Heneman, Moore and Wexley (1987) has further reiterated the significance of contextual factors such as time delay

between observation and rating of performance, the amount and method of observation of ratee behavior, and the purpose of appraisal for the rating process. Thus, models of the appraisal process with motivational as well as cognitive emphasis have all underscored the significance of appraisal purpose for the rating process. By addressing appraisal purpose, this dissertation will investigate the significance of a component critical to motivational as well as cognitive models of the performance appraisal process.

Previous studies that examined the influence of purpose on accuracy of performance ratings have obtained mixed results. While studies by Taylor and Wherry (1951), Bernardin, Orban and Carlyle (1981), Sharon and Bartlett (1969), Aleamoni and Hexner (1980) and Zedeck and Cascio (1982) have found significant effects; MaIntyre, Smith and Hassett (1984) found a very weak effect; yet others (Berkshire & Highland, 1953; Gmelch & Glasman, 1977; and Bernardin, Abbott, & Cooper, 1985) have failed to find a significant relationship between purpose and performance rating accuracy. Although, several factors such as sample characteristics, different appraisal purpose, strength of manipulations, research setting could have contributed to these mixed findings (to be discussed later), the extant literature has ignored two important considerations. These considerations and their theoretical implications are discussed below.

DeNisi et al (1984), while explicating their cognitive model of the appraisal process suggested that 'purpose of appraisal' has a motivational as well as a cognitive component. It appears that the motivational influence of purpose may be due to rater's perception of

the consequences of appraisal ratings for the ratee. This has two important theoretical implications. First, previous studies have treated purpose as if it were a purely cognitive variable, thus ignoring its motivational impact. The motivational impact of purpose on rating accuracy may be expected to operate through the rater's perception of the consequences of providing accurate ratings. If the consequences of providing accurate ratings are minimal, raters may not be as influenced by purpose as much as they would be if consequences were significant. Although related, purpose and consequences are distinct constructs and previous studies have failed to make this distinction. Therefore, it is likely that the impact of consequences may have served to confound the influence of purpose on accuracy of ratings leading to inconsistent results. Manipulating these two variables will permit a clearer explication of the purpose construct.

Second, the importance of individual differences for the rating process has been emphasized by several researchers (e.g. Feldman, 1981). The contradictory results reported in the literature may be due to the failure to distinguish raters who are low self-monitors from high self-monitors. Theoretical and empirical analyses of the self-monitoring construct suggest that this important (rater) disposition may be expected to moderate the relationship between appraisal purpose and rating accuracy. For instance, research on the self-monitoring construct suggests that only high self-monitors consider consequences of their actions whereas low self-monitors are not motivated to do so (Snyder, 1987). By considering the moderating influence of self-monitoring disposition among raters, this dissertation will endeavor to resolve the

conflicting findings reported in the literature between purpose and rating accuracy.

In summary, research on appraisal purpose has been too simplistic. It has focused exclusively on testing the association between appraisal purpose and characteristics of ratings without regard to potential moderators. By including merit-raise budget (a situational factor) to manipulate consequences of ratings, and rater self-monitoring (a person factor), this dissertation addresses the relationship from the much broader and richer interactional perspective.

Implications for Practice

Unlike most previous research on performance appraisals that has focused on rating format and the ability of the rater to provide accurate ratings, this dissertation focuses on motivational issues relevant to the appraisal process. This study is expected to yield several practical and useful suggestions for the practitioner. Specifically, this study is expected to produce three sets of results. Expected results and suggestions for the practitioners include the following.

First, it is expected that the rater's perception of purpose for which appraisal ratings will be used will influence the accuracy of ratings and subsequent personnel decisions. For instance, one may expect raters to inflate ratings when they perceive that the ratings they provide will be used for determining pay raises for their subordinates. On the other hand, supervisors are unlikely to be motivated to inflate or distort ratings when they perceive that the ratings will only be used for recommending subordinates for training. Since performance appraisals

are used for multiple purposes, one suggestion will be to clearly communicate the purpose(s) for which performance ratings will be used to the supervisors. Unless this information is clearly communicated, different raters may perceive different purposes and depending on their perception are likely to be motivated to provide inflated, distorted or accurate ratings. In complex organizations, it is easy to conceive of a scenario wherein one supervisor may believe that performance appraisal ratings are to be primarily used for distributing rewards, whereas another may believe that performance appraisal ratings are generally used for identifying training needs of subordinates. In the presence of such a scenario, one may expect the ratings provided by these supervisors to be inconsistent, even though, they may be appraising subordinates performing same jobs at the same performance levels. Such inconsistencies can be avoided by educating supervisors about the purpose(s) for which ratings will be used as well as how appraisal ratings are related to personnel decisions contingent upon those ratings.

Second, it is expected that the rater's perception of consequences of ratings will influence the accuracy of ratings and related personnel decisions. For instance, if the ratings and the outcomes contingent upon those ratings are of no real significance, raters may not be motivated to distort ratings. On the other hand, if consequences are perceived to be significant, raters may be motivated to inflate or otherwise distort ratings. In the appraisal context, the tendency to rate uncritically and leniently in order to avoid the ramifications of a deserved but harsh appraisal may be conceptualized as defensive behavior. While the

defensive behavior is manifested in the completion of the rating form, the source of the problem could be the anticipated encounters with the ratee - the object of the rating. Such dysfunctional influence of consequences of ratings can be minimized by enhancing rater's coping efficacy. Therefore, organizations should enhance rater's coping efficacy, in addition to increasing rater's ability to provide accurate ratings.

Finally, the influence of rater characteristics such as age, sex, education level, intelligence, etc., on performance ratings have been examined in the past. This dissertation focuses on an important disposition of raters, namely, self-monitoring. It is expected that ratings and related personnel decisions of high self-monitors will be more influenced by factors other than job performance such as purpose and anticipated consequences of ratings than those of low self-monitors. Training programs may be designed to help high self-monitors to overcome this tendency to consider information extraneous to job performance while appraising subordinates.

If the predicted results occur, practitioners may be able to improve the effectiveness of performance appraisal by communicating clearly the purpose(s) for which appraisal ratings will be used and educating raters about the hazards of considering information other than job performance. Failure to do so is likely to perpetuate inaccurate and distorted ratings leading to dissatisfaction with the appraisal process, feelings of inequity, and a sense of helplessness among employees, and may over time result in decreased organizational effectivness. Furthermore, in a litigious society, distorted ratings and personnel

decisions based on such inaccurate ratings may lead to charges of unfair discrimination.

Summary

This chapter briefly discussed various streams of research concerned with the accuracy of performance rating and presented the research problem. Dissertation objectives that address the accuracy of performance ratings were outlined followed by an analysis of expected theoretical and practical contributions of the dissertation. Performance appraisal literature and hypotheses relevant to this dissertation will be discussed in chapter II. Chapter III will outline the study methodology and an analysis of results of the study will be presented in chapter IV. Finally, chapter V will present a discussion of the results as well as conclusions drawn from that discussion.

Chapter Two

Literature Review

This chapter will begin with a brief analysis of performance appraisal research focused on the appraisal instrument and rater ability. These streams of research will be critically evaluated in terms of their potential to enhance accuracy of appraisal ratings.

This dissertation focuses primarily on motivational issues elicited by contextual factors. Therefore, research on contextual factors will be reviewed followed by a review and discussion of theory and research on appraisal purpose. As noted earlier, research on appraisal purpose has yielded inconsistent results such that the relationship between appraisal purpose and characteristics of performance ratings is not clear. This stream of research will be critiqued for failing to adequately operationalize appraisal purpose, explicitly manipulating perceived consequences of ratings, and consider individual differences (among raters) with potential to moderate the relationship between appraisal purpose and rating characteristics. Following a brief review of theory and research on self-monitoring, the rationale for expecting rater's self-monitoring disposition to moderate the relationship between appraisal purpose, consequences of ratings and rating characteristics will be provided. Finally, theory and research on appraisal purpose, budget constraints and self-monitoring will be synthesized to generate hypotheses.

Performance appraisal refers to the process by which an observer, typically a supervisor (or a peer), observes the behavior or the product of the behavior of his or her subordinate. The supervisor then has to make a judgment concerning the cause of the behavior. Since the purpose is to evaluate the subordinate, the supervisor typically attributes causality to the individual. This bias/error has been referred to as the fundamental attribution error (Jones & Davis, 1965). These judgments are then stored in memory (Cantor & Mischel, 1977, 1979) and at the time of the performance evaluation are recalled to evaluate/rate the subordinate on the appraisal instrument.

Performance appraisals are used for several purposes (Cleveland et al, 1989; Lawler, 1988; Locher & Teel, 1988) and the appraisal process has the potential to play an important role in enhancing organizational effectiveness (Latham et al, in press). Consequently, a legitimate concern of the practitioner as well as a recurring theme in the appraisal literature has been the accuracy of performance appraisal ratings.

The early appraisal research was based on the psychometric tradition. This research conceptualized accuracy as lack of errors such as leniency, stringency, central tendency and halo; and consequently, performance ratings free from these errors were regarded as accurate (see Feldman, 1981;1986). Recent theory (e.g. Funder, 1987) and research (e.g. Bernardin & Pence, 1980) has criticized this assumption and established the lack of relationship between errors and accuracy (see Bernardin & Pence, 1980). In the extant performance appraisal literature, two measures of accuracy, namely, differential/correlational

accuracy and distance accuracy, are used. Differential accuracy reflects the parallellism between subjects' and experts' ratings and several researchers (e.g. Bernardin & Cooke, 1992; Bernardin & Kane, in press; Borman, 1979) have argued differential accuracy to be the most important criterion for assessing accuracy of performance ratings. Distance accuracy, another frequently used measure of accuracy, is the average absolute value of the deviation of the obtained ratings from the true scores. Distance accuracy reflects the level difference between subjects' ratings and experts' ratings (see McIntyre, Smith & Hassett, 1984). These measures of accuracy will be used in this (dissertation) study and are fully operationalized in chapter three.

Most previous research on the accuracy of performance ratings has focused almost exclusively on the rating instrument and ability of the rater to generate accurate ratings. These streams of research will be critiqued in terms of their contribution towards enhancement of accuracy of performance ratings.

Appraisal Instrument

A large body of early research on performance appraisal has been concerned with the appraisal instrument. This body of research, in the classic psychometric tradition, focused on improving rating accuracy through improved design of the appraisal format. Appraisals are generally made on rating scales and various forms of rating scales have been investigated for their psychometric adequacy and the tendency to generate adequate evaluations (see Landy & Farr, 1980).

Over the last seventy years, several different rating scales have been recommended by the appraisal literature. These include Graphic

Rating Scales (Patterson, 1922), Behaviorally Anchored Rating Scales (Smith & Kendall, 1963), Mixed Standard Rating Scales (Blanz & Ghiselli, 1972), and Behavior Observation Scales (Latham & Wexley, 1977). Although these scales have progressively improved psychometric properties (see Landy & Farr, 1980), the same cannot be said about their ability to generate accurate ratings. This line of research seems to be based on the assumption that behaviorally-oriented scales are likely to yield more accurate ratings than those generated by trait type or numerically-oriented graphic rating scales. Research has established that such behaviorally oriented scales are also susceptible to biases. For example, Murphy and Constans (1987) found that the use of behavioral anchors in BARS may actually bias ratings. When BARS contained anchors that were actually observed, but not representative of overall performance, ratings were biased in the direction of the unrepresentative anchors.

BOS requires raters to observe and remember specific behaviors. However, people are simply incapable of such complex information processing (Miller, 1956; Simon, 1963). Much research on social cognition (Cantor & Mischel, 1977, 1979; Srull & Wyer, 1989; Winter & Uleman, 1984; Lord et al, 1982, 1984; Lord, 1985) has found that people generally categorize events into schemas based on prototype match. Thus, Murphy et al (1982) found that BOS measures general dimensions and not specific observable behaviors. Indeed, Nathan and Alexander (1985) found that both BARS and BOS yield ratings consistent with rater's cognitive schemas and not on the basis of observed behaviors.

There have been several extensive reviews of behaviorally oriented

scales and the conclusion seems to be that they are not much better than carefully constructed graphic scales or summated checklists for generating accurate ratings (Schwab, Henema, & DeCotiis, 1975; Bernardin et al, 1976; Bernardin, 1977; Landy & Farr, 1980, 1983). Thus this stream of research has not been very fruitful as variations in the instrument format do not appear to improve accuracy of ratings. Indeed, different formats of rating instruments account for only 4 to 8% of the variance in performance ratings (Landy & Farr, 1980; see also Smith, 1976). The most that can be concluded from this literature is that rigorous scale construction combined with behaviorally specific scale anchors and dimensional definitions reduces halo, leniency/stringency, and other biases. Format itself seems to have little effect on the accuracy of performance ratings.

Rater Ability

Two streams of research have been concerned with improving the rater's ability to generate accurate ratings. One stream of research has endeavored to enhance accuracy through 'rater training' to reduce appraisal errors, improve observation skills and use decision aids. A second stream, with a strong cognitive orientation, attempts to improve accuracy by studying judgmental processes and the accompanying distortions at various stages of information processing. A critical review of these two streams follows.

Early Approaches

Error Training

Many of the early attempts at rater training fall within the 'error training' model. With this approach, raters are given training on

common psychometric errors such as leniency, stringency, central tendency and halo, and then are admonished to avoid them.

An important assumption underlying this stream of research is that presence of errors indicate inaccuracy. A related assumption is that reducing these errors will improve accuracy. Theory and research has shown both these assumptions to be invalid (Funder, 1987; Murphy & Balzer, 1989). Consistently, although error training has been shown to be successful in reducing psychometric errors (Bernardin & Walter, 1977; Borman, 1975, 1979; Brown, 1968; Ivancevich, 1979; Latham, Wexley & Purcell, 1975) there is substantial evidence that reducing psychometric errors has little or no corrective effect on the accuracy of ratings (Bernardin & Pence, 1980; Borman, 1975, 1979; Murphy & Balzer, 1986; Nathan & Tippins, 1990; Smith, 1986; Smith, Hassett & McIntyre, 1982; Thornton & Zorich, 1980; Zedeck & Cascio, 1982). A recent meta-analytic study (Murphy & Balzer, 1989) designed to investigate the relationship between rating errors (halo and leniency) and rating accuracy concluded that the correlation between rating errors and accuracy is very near zero and, therefore, error measures are not good indicators of rating accuarcy, a conclusion consistent with the theoretical arguments of Funder (1987). Bernardin and Beatty (1984) have also argued that training programs that focus on minimizing rating errors simply exchange one response set for another without improving the accuracy of performance ratings. In fact, in one study, Bernardin and Pence (1980) found that rater error training led to a response set in raters that resulted in not only lower levels of leniency and halo error, but lower levels of accuracy as well. Another study conducted by Hedge & Kavanagh

(1988) reached the same conclusion. Thus, this line of research has not been successful in enhancing accuracy of appraisal ratings.

Observation Training

A variant of rater error training, particularly aimed at reducing halo is training to make raters better observers (Spool, 1978). Training to enhance observation skills is also based on the assumption that better observers are likely to yield more accurate ratings. This assumption has also been extended to rating formats. Thus, for instance, BOS require raters to observe and remember specific behaviors. People are incapable of such complex information processing (Miller, 1956; Simon, 1963). Much research on social information processing (e.g. Srull & Wyer, 1989) has found that people do not store specific events/behaviors as such in memory; instead, raters categorize information into preexisting schemas based on prototype match, thereby reducing the complexity of incoming information. During the rating process only the category labels are recalled and all traits and behaviors that comprise the category prototype are attributed to the ratee (see Cantor & Mischel, 1977, 1979; Winter & Uleman, 1984). Thus, Murphy et al (1982) found that BOS, which require raters to observe and remember specific behaviors, measures general impressions rather than specific behaviors. Additionally, Murphy and Balzer (1986) have documented that such general impressions might actually aid accuracy even though halo may be increased. These findings corroborate Cooper's (1981) concept of 'true halo' and his arguments for expecting a positive relationship between accuracy and halo.

Performance Diaries

Typically, performance is appraised once a year. Even if the raters observed all relevant ratee behavior, it is unreasonable to expect raters to remember examples of specific ratee behaviors. Therefore, research concerned with improving accuracy of performance ratings has also investigated the utility of decision aids such as behavioral diaries. Rater diaries have been advocated as a method of reducing the memory demands placed on raters (Bernardin & Walter, 1977; DeNisi & Williams, 1988). By documenting ratee performance, raters do not have to rely solely on their memory for appraising ratees. This approach assumes that rating accuracy will increase since memory loss, a key contributor to inaccuracy, will be less a factor. Several researchers (e.g. Balzer, 1986), on the other hand, have argued that initial impressions would have a biasing effect on the recording of incidents in behavioral diaries. In one study, Balzer (1986) found that raters were more likely to record information that was incongruent with initial impressions. Implications of this finding are that behavioral diaries, while designed to minimize bias in performance ratings, are themselves subject to cognitive distortion.

In summary, this stream of research focused on error training, training raters to be better observers, and the use of behavioral diaries has not been successful in enhancing accuracy of performance ratings.

Cognitive Processing Approach

Although Wherry (1952) presented the first formal cognitive model of performance appraisal (see also Wherry & Bartlett, 1982), the second

and more, cognitively-inclined stream of appraisal research is credited to Landy and Farr (1980). In their widely cited review of performance ratings, these researchers noted that cognitive processes of the rater are integral to the rating process. Following Landy and Farr, several theorists (e.g. Ilgen & Feldman, 1983, Cooper, 1981; DeNisi et al, 1984; DeNisi & Williams, 1988; Feldman, 1981, 1986) have presented complex cognitive models of the appraisal process. All models assume that the rating process is characterized by cognitive distortion affecting all stages of the information processing sequence.

Research on cognitive processing may be classified into that which is primarily concerned with information acquisition and that which emphasizes information processing (DeNisi & Williams, 1988). The former is more concerned with controlled processes and the latter with automatic processes (Feldman, 1981).

Studies on information acquisition have not only demonstrated that rater acquisition strategies exist (e.g. Balzer, 1986; Williams, DeNisi, Blencoe, & Cafferty, 1985; Zedeck & cascio, 1982), but also that these strategies can affect rating accuracy (e.g. Murphy, Garcia, Kerkar, Martin, & Balzer, 1982; see also DeNisi & Williams, 1988). Studies on memory issues such as encoding, storage, and retrieval have shown that raters possess categories of ratee effectiveness, and often rely upon them in the rating process (e.g. Borman, 1987; Murphy, Gannett, Herr, & Chen, 1986). Studies have also shown the critical role of memory processes in appraisal decisions (DeNisi & Williams, 1988: Williams, DeNisi, Meglino, & Cafferty, 1986), and have demonstrated how a variety of factors such as rater's expectations and job and ratee knowledge can

influence how performance information is processed (Hogan, 1987; Schmitt, Noe, & Gottschalk, 1986; Zedeck & Cascio, 1982).

Limitations of Cognitive Processing Approach

The ultimate challenge to cognitive approaches is to demonstrate the practical implications of this line of research. Although researchers have been able to demonstrate with some success that cognitive processes affect appraisal decisions, very few suggestions for the practitioner have been made and fewer yet applied.

A recurrent finding in cognitive processing research is that raters tend to rely on categories to encode (as well as store) information into pre-existing schemas, based on prototype match. Such categorization, while obscuring behavioral detail (codes) yields general impressions (impressionistic codes). During the rating process only the category labels or general impressions are recalled and then these are used to infer specific examples of behavior of the ratee. Such inference is subject to distortion as behaviors that comprise the category prototype are likely to be attributed to the ratee regardless of the actual behaviors of the ratee (behavioral codes).

Lord and Maher (1989) articulated the need to identify categories and prototypes of effective and ineffective performance held by raters. However, these categories and prototypes are likely to vary from individual to individual, department to department, organization to organization, and perhaps across cultures as well. DeNisi and Williams (1988) suggested the need to sensitize raters to the congruency (or incongruency) of their categories and prototypes (hence their general impressions or impressionistic codes) with the behavioral codes actually

exhibited by the ratees. To sensitize raters, incongruencies and inaccurate or inadequate prototypes need to be brought into rater's awareness and unlearned before accurate prototypes can be effectively learned (Lewin, 1938; see also Jawahar & Stone, 1992). These suggestions have seldom been applied in the appraisal literature.

To negate the influence of memory decay on rating accuracy, Bernardin and Walter (1977) suggested that raters use decision aids such as performance diaries. Furthermore, DeNisi and Williams (1988) argued that performance diaries may result in the use of behavioral codes, in addition to, or in place of impressionistic codes and therefore increase rating accuracy. However, Balzer (1986) has reported that the advantage of using behavioral diaries (increasing accessibility to a representative sample of behavioral codes) may be mitigated as raters tend to record only behavioral codes that are inconsistent with their initial impressions of the ratees. Such a biased sample of behavioral codes in combination with biases such as salience, availability and representativeness is likely to yield inaccurate or distorted ratings (see Kahneman & Tversky, 1979). Furthermore, a survey of actual raters by Williams (1987) has highlighted the impracticality of this suggestion to use behavioral diaries.

DeNisi and Williams (1988) have recommended imposing organization on performance data stored in memory (i.e., imposing appropriate schemata) to increase rating accuracy. Two types of organizations of performance data have been investigated. In the task-blocking approach, raters sequentially view performance of all ratees on one task at a time. Person-blocking involves sequentially viewing information about

one ratee at a time. Raters acquiring task-blocked information are more likely to organize information in memory by tasks. Conversely, those acquiring information blocked by persons may be expected to organize information in memory by person categories. Research investigating the relative efficacies of these two approaches has been inconclusive. For instance, Cafferty et al's (1986) study suggests task-blocking, whereas, studies by Williams et al (1986) and DeNisi et al (1987) suggest person-blocking as the best approach for generating accurate ratings. As DeNisi and Williams (1988) noted, perhaps there is no one method that is universally superior; moderating facors, such as appraisal purpose, or individual difference variables may exist. Clearly, more research is needed before sound recommendations can be made to the practitioner.

A final suggestion by cognitive theorists (DeNisi & Williams, 1988; Feldman, 1986; Pulakos, 1986) is to use rating scales as organizing devices. This approach attempts to provide raters with an organizing framework consistent with rating dimensions (Feldman, 1986) that are representative of behaviors defined through job analysis (DeNisi & Williams, 1988). Although DeNisi and Summers (1986) have provided some preliminary evidence, it is not clear why BARS, which inadvertantly achieves the same objective has been unable to generate more accurate ratings than other scales (see Landy & Farr, 1980).

Cognitive processing research has made rapid strides in the last decade, yet more systematic research is needed before practical suggestions to enhance rating accuracy in a significant manner can be made. Furthermore, suggestions based on cognitive processing research are likely to influence only ability of the rater rather than motivation

to provide accurate ratings. Performance ratings may be characterized more by a manager's or organization's need than by any attempt on the part of the rater to incorporate performance information accurately (Banks & Murphy, 1985; Lord & Maher, 1989). The appraisal process may be highly influenced by political variables (Longenecker, Gioia & Sims, 1987) with particular regard to future interaction and interdependence between the rater and ratee (Ilgen & Favero, 1985). These factors are likely to influence the types of ratings that are given, yet are not reflective of actual performance. Given these various organizational constraints, Banks and Murphy (1985) suggest that the rater's motivation to provide accurate ratings may lie outside the domain of cognitive processing research. Therefore, motivational issues confronting the rater are considered next.

Rater Motivation

As discussed in the previous section, research on the appraisal instrument and rater ability has had little success in enhancing accuracy of performance ratings. An assumption underlying these streams of research is that inaccuracy or errors in appraisal ratings are the result of lack of rater ability. Such an assumption ignores the political and motivational issues raters face in organizations (see Banks & Murphy, 1985; Fisher, 1989; Kane, 1980; Lord & Maher, 1989 for similar arguments). Even if raters are capable of rating accurately, there is no guarantee that they will choose to do so. In fact, surveys and field studies of actual raters suggest that rater behavior is highly calculative and motivated to accomplish desired goals, with little or no regard for rating accuracy (Bernardin & Villanova, 1986; Longenecker et

al, 1987; Longenecker & Gioia, 1988). For instance, Bernardin and Villanova (1986) found that superiors, adminstrators, and subordinates believed that ratings were often inflated to avoid confrontations with subordinates, to please certain employees, or because raters feel ill at ease in evaluating others. Longenecker, Sims and Gioia (1987) interviewed sixty executives and found that political considerations were nearly always involved in making performance appraisal ratings. Executives consciously used the appraisal process to attain desired ends such as to obtain a larger merit raise for a subordinate, encourage a subordinate with personal problems, teach a rebellious subordinate a lesson, jolt a subordinate into performing up to his or her potential, or the like. According to one interviewee, "accurately describing an employee's performance is really not as important as generating ratings that keep things cooking (p. 185)." Clearly then, rater motivation shaped by political considerations may be expected to significantly affect the accuracy of performance ratings.

This dissertation focuses on contextual factors with potential to influence accuracy of performance ratings. Apart from their influences on rater ability, contextual factors significantly affect rater motivation to provide accurate ratings. Such contextual factors include participation by ratee, opportunities to observe ratee behaviors, time delay between observation and performance rating, norms regarding proper ratings, consequences of providing accurate ratings for the rater as well as the ratee, and purpose(s) for which appraisal ratings will be used.

Although the beneficial results of involving raters and ratees in scale development and appraisal interview training programs have been noted (Dobbins, Cardy & Platz-Vieno, 1990; Giles & Mossholder, 1990; see also Landy, 1985), a recent review (Bretz et al, 1992) concluded that performance appraisal systems in U.S. organizations are designed primarily by personnel specialists with only limited input from managers (raters) who use the system and virtually no input from employees (ratees) affected by them. Additionally, Bretz et al (1992) noted that only few organizations provide rater training on an on-going basis. Ratees receive virtually no training in how to best use the process to receive feedback or improve performance.

The amount and method of observation has received some attention. Heneman and Wexley (1983) found that accuracy increased with the number of ratee behaviors observed by the rater. Favero and Ilgen (1983) reached the same conclusion. An early study by Maier and Thurber (1968) indicated that accuracy is greater when ratings are based upon a written or audio recording of ratee behavior rather than when the actual behavior is observed (cf. Heneman et al, 1987).

Several researchers (Heneman & Wexley, 1983; Nathan & Lord, 1983; Rush et al, 1981) have found that accuracy diminishes as a function of the time delay between the observation and rating of performance. The time delay influences memory decay, which introduces bias in the rating process. Specifically, memory decay affects the ability to recall job and ratee information and results in halo error and subsequently inaccurate ratings (Kozlowski & Kirsch, 1987). Rush et al (1981) found that the decline in accuracy with increasing time intervals was

independent of the memory capacity of the subject. Conversely, Smither and Reilly (1987) found that rater intelligence was significantly related to accuracy and concluded that rater intelligence, not rating delays, affected accuracy. Borman and Hallam (1991) also found that the past experience of the raters and their cognitive abilities influenced rating accuracy.

Social pressures and organizational norms may also affect the accuracy of performance ratings. For example, Quinn, Tabor, and Gordon (1968) provided evidence of the powerful influence of social pressures on discrimination in appraisals in a study of anti-Semitism. They found that social pressures to discriminate against Jewish employees led even those managers who were relatively egalitarian in their private views to discriminate against Jewish employees in evaluations of promotability. Perceived pressures from third parties also appeared to amplify the bias of those who were already anti-Semitic in their private beliefs. Similarly, Bowman, Worthy, and Greyson (1965) found that the reluctance of managers to promote women to supervisory roles was largely the result of anticipated resistance by co-workers. The influences of social pressures and norms are also reflected in the ratings received by most employees. A recent review (Bretz et al, 1992) indicated that it was common practice for 60 to 70% of an organization's workforce to be rated in the top two performance levels. Clearly, it is unlikely that all organizations have predominantly outstanding employees. Together these studies suggest that organizational norms that define proper/acceptable ratings significantly determine the accuracy of performance ratings.

Napier and Latham (1986) examined expectancies of raters from a social learning theory perspective. They identified outcome expectancies of raters in two disparate industries, newsprint and banking, using interviews and questionnaires. They found that managers perceived no consequences, positive or negative, of conducting performance appraisals, for themselves. This finding is not surprising as only 25% of the organizations even attempt to hold raters accountable for how they manage the appraisal process (Bretz et al, 1992). Conversely, Longenecker et al (1987) found that because of actual and perceived negative consequences of accurate appraisal, some managers knowingly make ratings that are inaccurate.

Although some contextual variables have been examined, very few studies have examined each of these variables. More research is needed before meaningful conclusions can be drawn. One contextual variable that has received the most attention is purpose of appraisal, the focus of this dissertation.

Purpose of Appraisal

Purpose of appraisal, a contextual factor with potential to influence motivation of the rater is the focus of this dissertation. Several models of performance appraisal have emphasized the role of rater motivation, in general, and purpose of appraisal, in particular, for generating accurate ratings (e.g. Ilgen & Feldman, 1983; Landy & Farr, 1980). For instance, in his theory of rating, Wherry (1952; see also Wherry & Bartlett, 1982) stated that accuracy of ratings can be improved when the rater's attention is focused on the rating task, when the rater is motivated to be objective, and when the behavior is readily

classified into specific categories, rather than overall schemata. A central feature of DeCotiis and Petit's (1978) model is the emphasis on rater motivation as a significant determinant of rating accuracy. Motivational issues confronting the rater, such as purpose of appraisal, have also been emphasized by predominantly cognitive models of the appraisal process. Purpose of appraisal was identified as a significant component of the rating process by Landy and Farr (1980). Ilgen and Feldman (1983), drawing on Feldman's (1981) earlier work, emphasized the significance of several contextual variables including appraisal purpose, norms regarding proper/acceptable rating, and opportunities to observe ratee behavior to the rating process. Appraisal purpose is an integral part of DeNisi et al's (1984) cognitive model of the appraisal process. Heneman et al (1987) have further underscored the significance of motivational factors, including appraisal purpose for the rating process. Thus, appraisal purpose is an integral component of cognitive (Ilgen & Feldman, 1983; DeNisi et al, 1984; Landy and Farr, 1980; Wherry, 1952) as well as motivational (DeCotiis & Petit, 1978) models of the appraisal process. These models suggest that purpose of appraisal has the potential to affect accuracy of performance ratings.

Since performance ratings are used for several purposes (Cleveland, Murphy & Williams, 1989; Locher & Teel, 1988), one objective of many empirical studies has been to investigate the influence of appraisal purpose on performance rating characteristics such as leniency and accuracy. Research on appraisal purpose based on cognitive models (e.g. DeNisi et al, 1984) has been limited. This stream of research addresses the effect of appraisal purpose on rating characteristics

indirectly by investigating the cognitive influence of appraisal purpose on information acquisition (e.g. Williams, DeNisi, Blencoe & Cafferty, 1985), memory processes (see DeNisi & Williams, 1988), and information processing and judgment (e.g. Zedeck & Cascio, 1982; see also DeNisi & Williams, 1988).

The focus of this dissertation though, as well as most of the research on appraisal purpose is on the motivational influence of appraisal purpose on rating characteristics. A substantial body of this research has focused on Wherry's (1952; see also DeCotiis & Petit, 1978; DeNisi et al, 1984) hypothesis that performance ratings provided for research and feedback purposes are likely to be more accurate than those provided for administrative purposes. While some studies have found support for this hypothesis, others have not. Understanding the reasons for these inconsistencies is of theoretical and practical importance as appraisal purpose has been referred to as a critical "boundary variable" with potential to limit the external validity of research findings (Bernardin & Kane, in press). However, it is not clear why studies investigating purpose effects have reached opposite conclusions. Furthermore, no attempt has been made to systematically identify and analyze the reasons for these inconsistencies so that firm conclusions may be drawn regarding the influence of appraisal purpose on rating characteristics. To advance this stream of research on purpose effects, this dissertation reviews the literature, and presents some plausible suggestions for resolving the inconsistencies. Specific hypotheses will be generated and tested in a laboratory study.

Appraisal Purpose Research

A number of studies have investigated the effect of intended use of the ratings on various psychometric properties of the ratings. Early studies investigating the effect of appraisal purpose focused on leniency of ratings and resistance of rating formats to those ratings. For example, in a military setting, Taylor and Wherry (1951) investigated the resistance of graphic rating scale and forced choice scales to leniency of ratings provided for either "research purpose only" or for making "administrative decisions," such as promotions or demotion of ratees. They reported that ratings provided for administrative purposes were more lenient than those provided for research purposes. In addition, the graphic rating scale was found to be more susceptible than the forced choice scale to this purpose effect. Sharon and Bartlett (1969) compared ratings generated by graphic rating scale and forced choice scale under four conditions: control - research purpose only, evaluation - ratings may be used for evaluation purposes by supervisor, identification - raters required to include their names so that they may be identified, justification - raters will have to explain his/her ratings to the ratee. They found that ratings provided by undergraduate students under evaluation and justification conditions were significantly more favorable than those in control and identification conditions but not significantly different from each other. The ratings in the control and identification conditions were also not significantly different from each other. Furthermore, no significant differences were reported between means or variances of forced-choice ratings. Sharon (1970) subsequently replicated these

results. Driscoll and Goodwin (1979) have shown that teacher ratings are more lenient when students are led to believe that ratings will be used to make administrative decisions.

In another study, Aleamoni and Hexner (1980) reported that ratings provided for "salary and promotion" purposes were significantly more favorable than those generated under standard teacher evaluation instructions. Results reported by this study are suspect for two reasons. First, standard teacher evaluation instructions could have had the effect of creating heteroscedasticity by allowing students (raters) themselves to conjure up various purposes for which teaching evaluations may be used. Second, the Illinois Course Evaluation Questionnaire used in this study has items that refer to the course as well as to the instructor. Since, only the overall ratings generated by the appraisal instrument were compared between the experimental and control groups it is not clear if the reported results are due to ratings on items that refer to the instructor or those that refer to the course. In a recent study, Bernardin and Orban (1990) examined the influence of three variables including appraisal purpose on leniency/severity of performance ratings. In this field study, thirty-two sergeants from two large municipal police departments evaluated sixty-five rookie patrol officers. As predicted, Bernardin and Orban found higher ratings when appraisals were used for personnel decisions than when they were used for feedback. This hypothesis received support with ratings from the graphic rating scale but not with the ratings from the mixed standard scale. However, in this study appraisal purpose was naturally confounded with department, as one department used performance ratings for

personnel decisions whereas the other used ratings for feedback purposes only. Although Bernardin and Orban argue that multiple sources of information revealed no indication that the naturally occuring confound was a threat to the internal validity of the study, it is impossible to rule out all rival explanations (Cook & Campbell, 1979) and therefore these results should be interpreted with caution.

Several studies have also found that appraisal purpose does not influence leniency/severity of ratings. For instance, Berkshire and Highland (1953), using a military sample, reported a lack of significant differences between ratings obtained for administrative purposes and those obtained for research purposes for both the graphic rating and the forced choice rating scales. Similarly, in a set of field studies, Hollander (1957, 1965) reported that the reliability and validity of peer-nomination scores of naval officers obtained under administrative conditions did not significantly differ from those obtained for research purposes. Borrensen (1967) found no differences among the student evaluations of teachers obtained under three different conditions. In another study, Centra (1976) also found no differences between teaching evaluations provided by students under administrative (tenure, salary, promotions) and feedback (improvement) conditions. Gmelch and Glasman (1977), using a within-subject design, compared ratings obtained for promotion/advancement and feedback/improvement purposes. In this study, students indicated whether they would rate their instructor differently if the evaluations were to be used for a different purpose, opposite from what original instructions had stated (either promotion/advancement or feedback/improvement). If the response was affirmative, students were

then asked if they would rate the instructor more favorably, somewhat more favorably, somewhat less favorably, or much less favorably. Only eleven percent of the students stated that they would rate the instructor either "somewhat less favorably" for purposes of instructor's improvement or somewhat more favorably" for purpose of instructor's promotion/advancement. The other eighty-nine percent stated that they would have rated the instructor same for both purposes. Therefore, Gmelch and Glasman reported that appraisal purpose did not significantly impact leniency/severity of ratings. Results reported by these authors should be interpreted cautiously for two reasons. First, ratings provided under original instructions, (that is, between-subject data) were not analyzed by the authors. Second, questioning whether one would rate differently if evaluations were to be used for a different purpose is likely to trigger social desirability concerns (such as honesty, consistency), rendering the second set of ratings suspect. In a well-controlled experiment, Meier and Feldhusen (1979) found no effect of purpose on the leniency of ratings. Murphy, Balzer, Kellam and Armstrong (1984) examined the influence of rating purpose on multivariate measures of accuracy in observing teacher behavior as well as measures of accuracy in evaluating teaching performance. Forty-five undergraduate students viewed a set of four videotaped lectures delivered by graduate students and were informed that their ratings would be used for either research purposes or to make decisions about the teachers they rated. Purpose of appraisal had no effect on characteristics of performance ratings.

Thus, while studies by Taylor and Wherry (1951), Sharon and

Bartlett (1969), Driscoll and Goodwin (1979), Aleamoni and Hexner (1980), Bernardin and Orban (1990) and others (Bernardin & Cooke, 1992; Kirkpatrick, Ewen, Barrett, & Katzell, 1968; Meyer, Kay & French, 1965; Zedeck & Cascio, 1982) have found a significant relationship between appraisal purpose and leniency/severity of ratings, several others (Berkshire & Highland, 1953; Bernardin, Abbott & Cooper, 1985; Borrensen, 1967; Centra, 1976; Gmelch & Glasman, 1977; Hollander, 1957, 1965; Meier & Feldhusen, 1979; McIntyre, Smith & Hassett, 1984) have reported the absence of such a relationship.

Apart from leniency/severity, studies have also investigated the influence of appraisal purpose on the accuracy of ratings. Using undergraduate students as raters, Zedeck and Cascio (1982) provided some evidence that perceived purpose affects several psychometric features of ratings such as leniency/severity, discriminability and accuracy. In this study, undergraduate students trained to avoid psychometric errors (e.g. leniency/severity and halo) rated thirty-three paragraphs describing performance of supermarket checkers for one of the following purpose: merit raise, development or retention. Subjects who rated hypothetical supermarket checkers for allocating merit increases provided less discrimination in their ratings (as indicated by the standard deviation of their ratings across ratees) than did subjects who provided ratings for retaining probationary employees. Additionally, the merit-purpose group differed from other rating-purpose groups in terms of their global differential accuracy (see Zedeck & Cascio, 1982). Therefore, Zedeck and Cascio concluded that ratings differ as a function of appraisal purpose with the difference being strongest between ratings

generated for merit increases and those generated for either development or retention purposes.

Interpretation of results reported by Zedeck and Cascio is difficult for the following reasons. First, interpretation of the differences in accuracy is difficult since they used a nontraditional accuracy measure: R values to represent rater accuracy. Second, it is unclear why the authors chose to use only rating discriminability as the criterion when a measure of differential accuracy was available. Finally, the study appears to have confounded scale anchors with the manipulation of purpose, making it difficult to assess the extent to which the results reported can be attributed to purpose independent of scale anchors (see Bernardin & Cooke, 1992). Two studies by Bernardin et al (1985) indicated that the significant differences found in Zedeck and Cascio study may be due to a purpose **X** scale anchor confound as Zedeck and Cascio had used a different rating scale in each of the three purpose conditions. When an independent scale was used for each of the three conditions, Bernardin et al (1985) found that there were no significant differences in accuracy among the three conditions. Recently, Bernardin and Cooke (1992) replicated the Zedeck and Cascio study, and the effect of purpose was also tested with identical anchors across rating purposes. Using the standard deviations of the subjects' responses as data points they found a significant effect for purpose as well as for the rating format. Contrary to the Zedeck and Cascio results, the greatest distinctions were not between merit and the other purposes; rather, they were between retention and the other purposes. When differential accuracy was used as the criterion, no significant

effect was obtained for appraisal purpose. Differential or correlation accuracy reflects the parallelism between subjects' and experts' ratings. Therefore, Bernardin and Cooke concluded that although appraisal purpose affects discriminability, it has no effect on rating accuracy.

In another study, McIntyre, Smith and Hassett (1984) used videotaped performances of male drama students acting as lecturers and compared ratings provided by undergraduate students for the following purposes: hiring, feedback and research. Though successful, the manipulation of appraisal purpose appears to be confounded with evaluation apprehension in the research condition (see McIntyre et al, 1984, p. 150). Nonetheless, they found subjects in the research condition to be more severe than subjects in the feedback and hiring conditions at a marginally significant level. These subjects were also more accurate at a marginally significant level with respect to correlation accuracy. With respect to the distance accuracy measure, no effect for perceived purpose was detected. Distance accuracy is the average absolute value of the deviation of the obtained ratings from the true scores. Experts' ratings are regarded as true scores (discussed in chapter 3). Since no analysis accounted for more than five percent of the total variance, McIntyre et al concluded that the effect of percieved purpose, if it does exist, may be weak.

Synthesis and Critique

In summary, research on appraisal purpose has generated contradictory results such that the relationship between appraisal purpose and leniency/severity as well as accuracy of ratings is not

clear. Is it necessary that we understand these inconsistencies? In other words, are these inconsistencies important. These inconsistencies are important for several reasons including the following. First, inconsistencies are always of theoretical interest to the researcher. Second, as a boundary variable, appraisal purpose has the potential to limit the external validity of performance appraisal research as performance ratings obtained for research purposes may be more or less accurate and/or lenient (severe) than those obtained for administrative purposes. Finally, since organizations use appraisal ratings for making several important administrative and personnel decisions, suggestions based on ratings obtained for research purposes are likely to be of little value to the practitioner. Therefore, these inconsistencies have theoretical importance as well as practical relevance and therefore need to be addressed. The importance of these inconsistencies and the implications they have elicits a second set of questions: why do these inconsistencies exist and how can they be resolved. This study is the first attempt to address these important questions.

First, and perhaps the easiest approach would be to analyze if the inconsistencies could be viewed as artifacts created by factors including differences in sample characteristics, appraisal purposes studied, manipulations, instruments used, and research setting. This, however, does not appear to be the case, as studies using undergraduate students as raters (e.g. Sharon & Bartlett, 1969; Zedeck & Cascio, 1982) and those using raters from organizations (e.g. Taylor & Wherry, 1951; Berkshire & Highland, 1953) have both reported contradictory findings regarding the effect of appraisal purpose on characteristics of ratings.

Similarly, studies investigating identical (e.g. Taylor & Wherry, 1951; Berkshire & Highland, 1953) as well as different (e.g. Zedeck & Cascio, 1982; McIntyre et al, 1984) purposes have both reached opposite conclusions. Studies with weak or confounded manipulations (e.g. Zedeck & Cascio, 1982; McIntyre et al, 1984) as well as those with appropriate and sound manipulations (e.g. Bernardin & Cooke, 1992; Bernardin et al, 1985) have both reported inconsistent results for appraisal purpose. Appraisal instruments have ranged from graphic rating scale (e.g. Taylor & Wherry, 1951), forced-choice scales (e.g. Berkshire & Highland, 1953), standard teaching evaluation forms (e.g. Aleamoni & Hexner, 1980), and scales developed using BARS methodology (e.g. Zedeck & Cascio, 1982). Appraisal formats do not seem to affect the conclusions drawn as studies using the same scale formats, namely, graphic rating scales and forced-choice rating scales (e.g. Taylor & Wherry, 1951; Berkshire & Highland, 1953) have reached opposite conclusions. Studies conducted in academic settings (e.g. Zedeck & cascio, 1982; McIntyre et al, 1984) as well as those conducted in military settings (e.g. Berkshire & Highland, 1953; Taylor & Wherry, 1951) have both reached contradictory conclusions regarding the effect of appraisal purpose on rating characteristics. Ironically, studies conducted by the same authors, Bernardin and his colleagues (Bernardin & Cooke, 1992; Bernardin & Orban, 1990; Bernardin et al, 1985) have yielded contradictory results regarding the effect of appraisal purpose. Although differences in measurement of rating characteristics and the constellation of factors considered above could have influenced results of particular studies, such a conclusion seems unlikely and perhaps a little premature.

A second reason for the contradictory findings reported in the literature may be the operationalization of the central variable: appraisal purpose. In explicating their model of the performance appraisal process, DeNisi et al (1984) emphasized the influence of both the cognitive and motivational components of appraisal purpose. Therefore, an adequate operationalization and manipulation of appraisal purpose would involve operationalizing and manipulating the cognitive as well as the motivational components of appraisal purpose. However, studies focused on the motivational influence of appraisal purpose have inadvertantly operationalized purpose as if it were a purely cognitive construct. For example, McIntyre et al's (1984) operationalization of "hiring-decision" purpose, shown below, illustrates the typical manner in which appraisal purpose has been operationalized in the literature.

"Hiring-decision instructions: Subjects were told that the psychology department was in the process of selecting graduate-student instructors for the upcoming semester. They were led to believe that the actors in the videotapes were real graduate students applying for the teaching positions and that the ratings would be used to make hiring selections" (McIntyre et al., 1984, p. 150).

A careful reading of these instructions indicate that only the cognitive component of appraisal purpose is manipulated as only information about the task (appraising for a particular purpose - hiring) is presented. No information about how appraising for this particular purpose, that is, hiring, will affect each of the actors portrayed as graduate students is presented. Apparently, the decision to hire or not hire will affect some candidates (actors) more than others.

Though the focus of this stream of research is the motivational effects of appraisal purpose, the above illustration clearly suggests that the affective component of appraisal purpose is neither manipulated nor considered. By failing to explicitly manipulate affective aspects of appraisal purpose, studies on purpose effects have inadequately operationalized the purpose construct. Additionally, since only information about the task is presented, this study, like all others on appraisal purpose, in essence, has attempted to impose a cognitively-oriented set on raters. Imposing a cognitively-oriented set, however, does not rule out affective experiences as affect and cognition are under the control of partially dependent systems that can potentially influence each other in a variety of ways (Zajonc, 1980). For instance, in his influential article, Zajonc (1980) argued that affect is always present as a companion to thought such that cognition (information processing) always elicits affect. Without exception, previous studies on the effects of appraisal purpose have failed to consider this possibility.

Consequently, when purpose is operationalized as a cognitive construct and affective influences are neither explicitly manipulated nor controlled, purpose is likely to elicit motivational concerns manifested in the form of rater's perception of consequences of appraisal ratings for the ratee as well as for the rater him/herself. Perceived consequences of ratings for the ratee may operate such that if consequences to the ratee are perceived to be minimal raters may not be motivated by purpose as much as they would be if consequences to the ratee are perceived to be significant. In many settings, for instance,

raters are likely to be strongly motivated to avoid the responsibility or the negative interpersonal consequences of giving low ratings, especially when decisions based upon those ratings are anticipated to be of significance to the ratee (for example, see Bernardin & Villanova, 1986). Thus, failure to control for or manipulate the effect due to perceived consequences could have had the effect of creating heteroscedasticty in the distribution of ratings, within each appraisal purpose, by allowing raters themselves to conjure up various levels of emotional imagery concerning the ratees. Therefore, it is conceivable that the impact of the consequences may have served to confound the influence of purpose on accuracy of ratings leading to contradictory results. This study avoids such confounding by explicitly manipulating both appraisal purpose and rater's perception of consequences of ratings for the ratee. The latter will be accomplished by manipulating a situational characteristic: merit-budget size - amount of funds available in the merit-budget for disbursement.

Budget Constraints

Several researchers (e.g. Ilgen & Favero, 1985) have pointed out that the anticipated influence of ratings on future interaction and interdependence between the rater and the ratee are likely to be of concern for the rater and may have the effect of biasing those ratings. Rater's perception of consequences of ratings for the ratee is likely to operate such that if consequences to the ratee are perceived to be minimal raters may not be motivated by appraisal purpose as much as they would be if consequences to the ratee are perceived to be significant. Indeed, employees including supervisors, administrators, and

subordinates have reported that raters often inflate ratings in order to avoid confrontations with subordinates, to please certain employees, or because they feel ill at ease in evaluating others (Bernardin & Villanova, 1986). Such evidence is supportive of the earlier argument that the motivational concerns associated with appraisal purpose are likely to be manifested in rater's perception of consequences of ratings for the ratee as well as himself/herself.

In this study, rater's perception of consequences of ratings for the ratee will be manipulated by manipulating the amount of funds in the pay-raise budget. Manipulating the amount of funds available for pay increases will lead raters to perceive different consequences depending on the particular appraisal purpose. For instance, the amount of funds available for providing pay increases is not likely to have any effect on performance ratings when those ratings are primarily used for identifying training needs of employees. On the other hand, when performance ratings are primarily used for recommending pay increases, raters are likely to perceive significant consequences of the ratings they provide as higher ratings are likely to result in larger pay increases for their subordinates than lower ratings.

Although, rater's perception of consequences of ratings are neither manipulated nor controlled, the extant literature on pay-for-performance focuses primarily on such an unidirectional influence between performance ratings and pay increases. This focus assumes an abundance of funds for pay allocation. Such an abundance of funds may facilitate purpose effects as supervisors may be motivated to obtain higher pay increases for their subordinates by providing lenient

ratings. In many settings, for instance, raters are likely to be strongly motivated to avoid the responsibility or the negative interpersonal consequences of giving low ratings, especially when decisions based upon those ratings are anticipated to be significant to the ratee (see Bernardin & Villanova, 1986).

During recent times, due to economic hardships, merit-raise budgets have shrunk. When very little or no funds are available for pay allocation purposes, the relationship between performance ratings and pay is weakened; and consequently, the ratings provided by raters are likely to have only a minimal impact on their subordinates. Under such circumstances raters may not be motivated to inflate ratings. Alternatively, it is quite possible that reduced budgets may force raters to be more stringent in their evaluations in order to avoid the embarassment of having to confront the "inequitable" as well as "dissonance-producing" scenarios of high performance-low/no merit raise. Indirect evidence for such a strategy may be found in the reaction of congress to the initial efforts by the Small Business Administration and NASA to use appraisal data as a basis for awarding bonus pay. Congress was displeased that over 50% of eligible employees were recommended for a bonus and responded by attaching a proviso to an appropriations bill providing that "no more that 25% of the number of Senior Executive Service positions, or positions under similar personnel system, in any agency may receive performance awards" (U.S. General Accounting Office, 1980, p. 2).

In summary, when performance ratings are exclusively used for identifying training needs, rater's perception of consequences of

ratings for the ratee are likely to be low, irrespective of the amounts of funds in the pay-raise budget. Consequently, no rating inflation may be observed. However, when performance ratings are exclusively used for pay raise purposes, raters are likely to perceive significant consequences of the ratings they provide for the ratees. For instance, when funds are abundant raters may be more motivated to inflate ratings and thereby secure larger pay increases for their subordinates than when funds are non existent. Alternatively, when no funds are available for providing pay increases, raters may be motivated to deflate ratings in order to avoid the embarassment of having to confront the "inequitable" and "dissonance-producing" scenario of high performance-no merit raise.

When appraisal purpose is operationalized as a cognitive construct and affective influences are neither manipulated nor controlled, purpose is likely to elicit motivational concerns manifested in the form rater's perception of consequences of ratings for the ratee. Failure of previous studies to control for or explicitly and uniformly manipulate the effect due to perceived consequences could have had the effect of creating heteroscedasticty by allowing raters themselves to conjure up consequences of different intensities. Therefore, it is conceivable that the impact of consequences may have served to confound the influence of purpose on accuracy of ratings leading to contradictory results. This study avoids such confounding by explicitly manipulating both appraisal purpose and rater's perception of consequences of ratings for the ratee.

A **third explanation** for the contradictory results reported in the literature for appraisal purpose may be the failure of previous studies

to identify individual differences with potential to moderate the influence of appraisal purpose. As mentioned earlier, the motivational influence of purpose may also be due to the rater's perception of consequences of providing accurate ratings for both the ratee and him/herself. Although, the anticipated influence of ratings on future interaction and interdependence between the rater and the ratee are likely to be of concern to the rater and may have the effect of biasing those ratings (see Ilgen & Favero, 1985), all raters may not have the ability as well as the motivation to provide ratings in anticipation of the expected effect of those ratings for themselves as well as for the ratees. Indeed, research on self-monitoring suggests that, in contrast to low self-monitors, high self-monitors are likely to take into account extraneous information such as appraisal purpose and consequences of ratings while appraising performance and making related personnel decisions.

Self-Monitoring

The construct of self-monitoring belongs to the family of self theories that emphasize the variability of the presented self. Its intellectual ancestry can be traced back to the "many social selves" of William James (1890), to the societal origins of self as set forth by the symbolic interactionists (e.g. Mead, 1934) and to the life-as-theater metaphor elaborated by Erving Goffman (1956). Present in all of these approaches is the notion that individuals actively strive to influence what others think of them by carefully orchestrating the impressions they convey. Goffman (1956) drawing on earlier work, suggested that we behave the way others expect us to, that we are alert

to subtle cues in our social environment, and that in general we engage in self-presentation.

A sociologist, Goffman ignored individual differences. Mark Snyder (1974) pointed out that there are striking individual differences in the extent to which individuals can and do monitor their self-presentation, expressive behavior, and nonverbal affective display. Research on the self-monitoring construct suggests that high self-monitors are adept at deciphering and interpreting cues in the social environment and using these cues as guidelines for monitoring (that is, regulating & controlling) their own verbal and non verbal self-presentations (Snyder, 1979, p. 89). High self-monitors tend to adopt what they see as a "pragmatic" interpersonal orientation, strategically creating social interaction patterns that promote situationally appropriate interaction outcomes. The high self-montoring social style is one that chronically strives to present the appropriate type of person called for in every situation. On the other hand, individuals low in self-monitoring lack either the ability or the motivation to regulate their expressive self-presentations. Their expressive behaviors, instead, functionally reflect their own enduring and momentary inner states, including their attitudes, traits, and feelings. Low self-monitors tend to adopt what they regard as a "principled" interpersonal orientation, which is reflected in the correspondence between their feelings and attitudes and their behavior. The low self-monitoring orientation is geared toward displaying a person's true dispositions and attitudes in every situation (Snyder, 1987).

Research with the self-monitoring scale has provided empirical

support for many hypotheses about the cognitive, behavioral, interpersonal consequences of self-monitoring and has been found to be predictive in several areas including the nature of friendships, romantic relationships, sexual involvements, advertising, psychopathology, and personnel selection (see Snyder & Gangestad, 1986; Snyder, 1987; Snyder, Berscheid & Matwychuk, 1988).

A key element of self-monitoring is sensitivity to situational cues. Several investigations have confirmed this "sensitivity" hypothesis. One study used excerpts from the television program "To tell The Truth". On this program, one of the three guest contestants is the "real Mr. X. However, all three claim to be Mr. X. Participants in this study watched each excerpt and then tried to identify the real Mr. X. High self-monitors were much more accurate than low self-monitors at spotting the truthful contestant and seeing through the deceptions of the other two (Geizer, Rarick & Soldow, 1977). High self-monitors are particularly sensitive to any information that might guide their expressive self-presentations. Indeed, when given the opportunity to do so, they consult information about the typical self-presentations of their peers more often and for longer periods of time than their low self-monitoring counterparts (Rhodewalt & Comer, 1981). So important is such social comparison information to high self-monitors that, at times, they may even go as far as to "purchase" at some cost to themselves, information that may help them choose appropriate self-presentations (Elliott, 1979). Additionally, high self-monitors are adept at intentionally controlling their nonverbal expressive behaviors. In one investigation, subjects read aloud an emotionally neutral paragraph

(e.g. "I am going out now. I won't be back all afternoon. If anyone calls, just tell them I'm not here") in ways that conveyed as accurately and naturally as possible each of seven different emotions, namely, happiness, sadness, anger, fear, surprise, disgust, or remorse. Naive judges after watching films or listening to tapes, indicated which emotion the person expressed. High self-monitors were much better able than low self-monitors to communicate accurately the intended emotion, in both the vocal and facial channels of expression (Snyder, 1974). They could, with little apparent difficulty, look and sound in quick succession happy and then sad, fearful and then angry, and so on through the list of emotions.

In social situations, high self-monitors invest considerable effort to "read" and understand others in search of information to aid them in choosing their own self-presentations. In one investigation, Berscheid, Graziano, Monson and Dermer (1976) gave subjects an opportunity to observe someone they expected to date. They found that men and women high in self-monitoring were more likely than those low in self-monitoring to notice and remember information about their prospective dates, make inferences about their personalities, and express liking for them. Thus, high self-monitors are motivated to use their impressions of others as cues to guide their own self-presentations in any ensuing social interaction. Another example of the differences in self-monitoring dispositions of high and low self-monitors is illustrated in a study conducted by Caldwell and O'Reilly (1982). In this study conducted in a corporate setting, Caldwell and O'Reilly found that success on a boundary spanning job to be a function

of self-monitoring. Since boundary spanning jobs require attention to cues in the environment, interpretation of these cues and appropriate responses, high self-monitoring field representatives performed more effectively and had longer tenure than low self-monitoring field representatives.

Taken together, these studies suggest that high self-monitors are capable of tailoring their behaviors to be congruent with the situational and appropriateness of those behaviors. The behaviors of low self-monitors, on the other hand, reflect their inner feelings, traits and attitudes without regard to the situational or interpersonal appropriateness of those behaviors. Additionally, high self-monitors appear to carry their concern with their own public appearances to a concern with the images conveyed by people with whom they may be associated. Consistently, high self-monitors place relatively greater emphasis on, and therefore pay considerable attention to external appearances when choosing whether or not to date someone. Similarly, low self-monitors appear to carry their concern with their own personal dispositions over to a concern with the suitability of the personal dispositions possessed by people they select for a relationship partner. Consistently, low self-monitors place relatively great emphasis on, and are therefore sensitive to, the internal qualities of their prospective dating partners. For instance, in a study by Snyder, Berscheid and Glick (1985) the high self-monitoring men paid more attention to physical appearances. They spent proportionately more time than low self-monitoring men inspecting the photographs of their potential partners. Low self-monitoring men devoted their attention to the psychological

characteristics of their potential partners. They spent proportionately more time than low self-monitoring men studying the personality sketches.

In a related experiment, other college-aged men chose between two prospective dating partners (Snyder & Simpson, 1984). One had a physically attractive exterior but, as revealed in the file, a rather moody, withdrawn, and self-centered personality. The other was much less attractive on the outside (and, in fact, was of below average physical attractiveness), but had, as revealed in the file, a highly desirable (sociable, outgoing & open) personality. Here, when forced to sacrifice one feature for another, 69 % of the high self-monitoring men chose the physically more attractive date even though she possessed a relatively undesirable personality. In contrast, 81% of the low self-monitoring men preferred the partner with the sterling inner qualities, even though this desirable personality was housed in an unattractive exterior. This research suggests that high and low self-monitors adopt sytematically different approaches to gathering, weighing and acting on information. Specifically, while high self-monitors are influenced by extraneous factors, low self-monitors are influenced by factors relevant to the task at hand. These factors may include cues, attributes of objects or persons.

This line of research has also been extended to the domain of personnel selection. Snyder and his colleagues (Snyder, Berscheid & Matwychuk, 1985, 1988) hypothesized that high and low self-monitors would adopt distinctly different strategies in personnel selction. In one study with college students, for example, Snyder, Berscheid, and

Matwychuk (1985) found that high self-monitors wanted to hire the attractive, well-dressed applicant who appeared to be concerned with her appearance for the position of sales clerk in a women's clothing store, even though she was rather unsociable and lacked organizational ability. Low self-monitors, on the other hand, preferred to hire the person who had the abilities but did not look the part. In another instance, low self-monitors gave a camp counselor's job to the applicant with a very gregarious and emphathetic personality, even though he looked and dressed more like a junior account executive than a camp counselor. Snyder, Berscheid and Matwychuk (1988) conducted two experiments to replicate these results. In both experiments, undergraduate students examined information about the physical appearance and personalities of two applicants for a specific job and then decided which applicant should receive a job offer. In experiment I, information about the applicant's physical attractiveness and job-appropriate dispositions were varied. In experiment II, job appropriateness of the applicant's physical appearance and their personalities were both varied. In each experiment, high self-monitoring individuals placed greater weight on extraneous information, namely, physical appearance than did low self-monitoring individuals. By contrast, low self-monitoring individuals placed greater weight on information about relevant personal dispositions than high self-monitoring individuals. This line of research suggests that high self-monitors are likely to be influenced by factors extraneous to job performance whereas low self-monitors base their decisions on job relevant information.

Once in a job setting, considerations similar to those involved in

the selection situation may be invoked when evaluations are made about who should be awarded pay raises, promotions, etc. When the evaluations are made by high self-monitoring individuals, matters of appearance, style, or other extraneous factors may come into play. Motivational tendencies of high self-monitors to regulate their expressive self-presentations stem from their concern with the situational and interpersonal appropriateness of his/her social behavior. Since high self-monitors are concerned about the situational and interpersonal consequences of their behaviors they may be expected to take into account extraneous information such as purpose of appraisal and actual or perceived consequences of ratings while appraising and providing performance ratings.

On the other hand, low self-monitoring individuals are likely to base their evaluations on the actual job performance of their subordinates. Furthermore, low self-monitors have high attitude-behavior consistency (Snyder, 1979); are motivated by their internal states and lack either the ability or the motivation to regulate their behavior in anticipation of the consequences of those behaviors. Therefore, they are not likely to consider extraneous information such as appraisal purpose or consequences of ratings when appraising employee performance. Thus, the effect of appraisal purpose and consequences of ratings on rating characteritics and related personnel decisions will be moderated by the self-monitoring disposition of raters.

Summary Conclusions and Hypotheses

The objective of this section is to summarize the literature reviewed thus far and present hypotheses specifying the relationship

between appraisal purpose and characteristics of performance ratings. In conclusion, research focused on improving performance rating accuracy has proceeded along three broad lines:

1. attempting to design the ideal instrument that would generate valid ratings

2. attempting to enhance ability of the rater to provide accurate ratings

3. examining the impact of motivational issues confronting the rater.

The first two streams of research were reviewed, critiqued and their inability to substantially influence rating accuracy noted. Next, research on appraisal purpose, a contextual variable, with potential to influence motivation of the rater to provide accurate ratings was reviewed. The contradictory findings on appraisal purpose were attributed to

1. operationalization of appraisal purpose as a purely cognitive variable, thus, disregarding its motivational influence

2. failure of previous studies to control for consequences of ratings for the ratee, thereby confounding consequences with appraisal purpose

3. failure of previous studies to consider individual differences (among raters) that predispose raters to be differentially motivated and consider extraneous information such as appraisal purpose and consequences of ratings while evaluating employee performance.

Next, theory and research on appraisal purpose, consequences of ratings, and self-monitoring were integrated. Briefly, Snyder's

(1974) self-monitoring theory and empirical evidence suggests that low self-monitors base their evaluations/judgments of others on attributes relevant to the task at hand (performance in the case of performance appraisal task) whereas high self-monitors are likely to place more emphasis on extraneous or task irrelvant attributes (see Snyder, 1987; Snyder et al, 1985, 1988). Therefore, while high self-monitors are likely to consider task irrelevant attributes such as appraisal purpose and consequences of ratings during performance appraisal, low self-monitors are not likely to do so. Accordingly, when performance ratings are used for determining pay increases, high self-monitors are likely to provide ratings in anticipation of the consequences those ratings may have for the ratees as well as themselves. Consequently, in contrast to low self-monitors, high self-monitors may be expected to inflate ratings if funds for allocating pay raises are abundant and deflate ratings when funds are non-existent.

When ratings are used exclusively for training purposes, perceived consequences of ratings are likely to be minimal, irrespective of the amount of funds available for allocating pay increases. In the absence of consequences for themselves as well as the ratees, high self-monitors will have no reason to inflate or otherwise distort ratings and hence will provide ratings similar to those provided by low self-monitors. These ideas are more formally expressed in the form of hypotheses.

Dependent Variables: Leniency, and accuracy

H1: Ratings provided for merit raise will be more lenient and less accurate than those provided for training purposes.

H2: Ratings provided in the 'plenty of funds' condition will be more lenient than those provided in the 'no funds' condition.

H3: Ratings provided by high self-monitors will be more lenient and less accurate than those provided by low self-monitors.

H4: Purpose, pay-raise budget, and self-monitoring will interact to affect leniency and accuracy. For low self-monitors no significant pay-raise budget X purpose interaction effects will be noted. Ratings provided by high self-monitors will be

a. most lenient in the merit raise-plenty of funds condition

b. most stringent in the merit raise-no funds condition

c. least accurate in the above two conditions.

Dependent Variable: recommendation for merit raise, and training.

H5: Merit raise recommendations will be more inflated than training recommendations.

H6: Merit raise recommendations in the plenty of funds condition will be stronger than those in the no funds condition.

H7: High self-monitors will strongly recommend merit raises than low self-monitors.

H8: Self-monitoring and pay-raise budget will interact to affect merit raise recommendations. High self-monitors will make stronger recommendations in the plenty of funds condition and weaker recommendations in the no funds condition than low self-monitors.

H9: Training recommendations in the plenty of funds condition will not be significantly different from those made in the no funds condition.

H10: Training recommendations made by high self-monitors will not be significantly different from those made by low self-monitors.

H11: Self-monitoring will not interact with pay raise budget to influence training recommendations.

Chapter Three

Methods

The purpose of this chapter is to outline the research study that examined the relationship among appraisal purpose, budget constraints, rater self-monitoring and characteristics of performance ratings. Following an outline of data collection procedures, the experimental design and experimental procedure will be presented. After a description of sample characteristics, operationalizations of the constructs will be discussed. Finally, statistical techniques for analyzing the data will be presented.

Data Collection Procedures

In the first part of the study, subjects completed the 'self-monitoring' scale. Administration of the self-monitoring scale and the actual experiment was separated in time (at least 3 weeks) to avoid cuing or carry over effects. The self-monitoring scale is enclosed as appendix A.

Experimental Design

The study consisted of two appraisal purposes: pay raise and training, and two levels merit-raise budget (to manipulate consequences of ratings): plenty of funds, no funds. Self-monitoring was used as a block (high and low self-monitors) and then treated as an independent variable. Therefore, this study is a 2 (pay raise, training) X 2 (plenty of funds, no funds) factorial, with self-monitoring as a blocking

factor. Manipulation checks were administered for appraisal purpose and consequences of ratings.

Experimental Procedure

In the first part of the study, subjects were asked to complete a self-monitoring scale and low and high self-monitors were identified. In the second part of the study, subjects were blocked on the self-monitoring variable and low and high self-monitors were randomly (and independently) assigned to each of the four conditions. Each subject received an information packet containing a letter, scenario, appraisal purpose (either merit-raise or training), availability of funds (either plenty of funds or no funds), and performance stimuli. The scenario contained 1. a brief description of a mail-order company specializing in a wide range of outdoor products, and 2. a job description of sales representatives. Subjects were provided with performance information of two subordinates, Pat and Chris. This information was presented in the form of critical incidents. For instance, one critical incident read "made a recommendation about adding Spencer fishing poles because of numerous customer suggestions" and another "lost temper when dealing with an upset customer." Twenty-five such incidents captured performance of each subordinate. Additionally, the order in which critical incidents capturing Pat's and Chris's performance were presented was counterbalanced within each cell. The role of the subject was detailed in the letter. Subjects were instructed to first familiarize themselves with the scenario, performance appraisal form, and critical incidents and 1. evaluate performance of Pat on a Behaviorally Anchored Rating Scale, designed for this study and 2. recommend either merit raises or

training for Pat on a scale similar to that used by Bernardin and Cooke (1992).

Following Bernardin and Cooke (1992) we used the same rating format for both merit raise and training purposes. They used the following scale points: 1 = strongly oppose personnel decision, 4 = neutral regarding personnel decision, and 7 = strongly support personnel decisions. It is quite possible that the words 'personnel decision' may lead subjects to infer that the personnel decision has already been made and that they were to either oppose or support this decision. To avoid such confusion we used the following anchors: 1 = strongly oppose, 3 = somewhat oppose, 5 = neutral, 7 = somewhat support, 10 = strongly support. This scale was prefaced by two statements. Depending on the appraisal purpose the wordings of these statements were slightly altered. The first read "please make a <u>training</u> (or merit raise) decision for Pat." The second read "decision is whether <u>to send subordinate for training</u> (or give a merit raise)."

To support the underlying theoretical rationale for the study we measured subjects' perceptions of consequences, the extent to which subjects' considered consequences while making decisions regarding merit raise or training, as well as subjects' willingness to assist Pat. Subjects' perceptions of consequences are expected to vary across treatment combinations. Additionally, examining subjects' responses to consequences in terms of decision accuracy (i.e. inflated/deflated decisions) as well as willingness to assist ratee(s) is likely to shed light on the motivational mechanisms underlying purpose effects. Lastly, manipulation checks for purpose, and availability of funds were

administered. Subjects were debriefed, thanked for their participation and dismissed.

Sample Characteristics

A pilot study was conducted to refine measures, experimental procedure and manipulation checks. Approval to conduct the study has been obtained from Oklahoma State University's Institutional Review Board for Human Subjects Research. Subjects were 320 undergraduate students enrolled in management classes at Oklahoma State University. There were 11 sophomores, 143 juniors, and 166 seniors. Of these 184 were male and 136 were female.

Operationalization of Constructs

The independent variables that must be operationalized to examine the relationship between appraisal purpose and characteristics of performance ratings include appraisal purpose, consequences of ratings, and self-monitoring. Dependent variables include measures of leniency and accuracy of ratings as well as decisions regarding merit raise and training. Operationalization of these constructs and measures will be discussed next.

Independent Variables

Appraisal Purpose

The appraisal purpose was stated while describing the task/role of the subject (rater). This information was attached to the cover letter contained in the information packet provided to each subject.

Merit raise - purpose instructions

Please note that this company uses performance appraisal ratings for merit-raise purposes only (i.e. pay increases). Therefore,

after rating Pat's performance, please make a decision regarding merit-increase for Pat.

Training - purpose instructions

Please note this company uses performance appraisal ratings for training purposes only. Training is provided to improve job knowledge, skills or abilities. The duration of training typically varies from 1 to 3 days. When the employee is attending a training program the company provides regular wages/salary and a temporary worker is assigned to replace the trainee during the employee's absence. After rating Pat's performance, please make a decision regarding training for Pat.

Budget Constraints

In this study rater's perception of consequences of ratings for the ratee were manipulated by manipulating the amount of funds available for allocating pay increases.

Plenty of Funds Condition:

Last year was a normal year for the company. This year the company made unusually large profits and consequently funds in the pay-raise budget have been tripled. So this year there will be plenty of funds/money for pay increases. Funds in the pay-increase budget are expected to remain at the current level for at least another 2 to 3 years.

No Funds Condition:

Last year was a normal year for the company. This year the company incurred heavy losses and consequently there are no funds in the pay-raise budget. So this year there will be no funds/money for

pay increases. Funds in the pay-raise budget are not likely to increase dramatically for at least another 2 to 3 years.

Self-Monitoring

Self-monitoring was measured using the self-monitoring scale constructed by Snyder and his colleagues (Gangestad & Snyder, 1985; Snyder & Gangestad, 1986). Snyder and his colleagues' new 18-item measure has an internal consistency of .70. This new measure is more factorially pure than the original measure (Snyder, 1974). The first unrotated factor emerging from a principal-axes factor analysis accounts for 62% of common variance (3 factors) compared to 51% accounted for by the first unrotated factor emerging from a factor analysis of the original 25-item measure. More importantly, total scale scores of the new 18-item measure are uncorrelated with an estimate of the second, relatively minor, source of variation, $r = .03$. By contrast, the total scale scores on the original 25-item measure are mildly correlated with an estimate of the second source of variation, $r = .15$. Further information on this new measure is provided in Gangestad and Snyder (1985). A complete discussion of the construct and construct validity of the scale is available in Snyder's (1987) book.

Dependent Variables

Rating Scales

Performance ratings were obtained on a BARS developed specifically for this study as well as on a one-item (global) rating scale. The job description of ratee(s) (service representative) was written to yield five performance dimensions: interpersonal and communication skills, dependability, quality, knowledge, and initiative. A BARS based on this

job description was constructed to evaluate performance along these dimensions. Additionally, to facilitate the ratees' task performance information presented in the form of behaviors and results were written to correspond to these five dimensions as well.

True Scores

Mean "expert" ratings were used as true score measures of service representatives' performance. Two management faculty members with a combined experience of over 30 years served as expert raters. Agreement between raters was high (inter-rater reliability, r=.86). Using expert ratings as true scores is widely accepted in the appraisal literature (see, for example, Borman, 1979; McIntyre et al, 1984; Murphy et al, 1982).

Accuracy Measures

Ratings provided by subjects were manipulated to yield measures of leniency/severity, differential accuracy and distance accuracy. Leniency/severity, and distance accuracy indices described in McIntyre et al (1984) were used in this study. Differential accuracy index (Borman, 1979; Cronbach, 1955) was also used.

Leniency/Severity

Leniency/severity is defined as a rater's tendency to assign ratings that are higher (leniency) or lower (severity) than those warranted by a ratee's performance. McIntyre et al (1984) used the following formula to operationalize leniency/severity

$$\text{Leniency} = \frac{\sum_{j=1}^{r}\left[\frac{\sum_{i=1}^{d}(T_{ij}-R_{ijk})}{d}\right]}{r}$$

where d is the number of items (in this case, 5)
r is the number of ratees (in this case, 1)
K is the subscript referring to the k rater
R - refers to the obtained rating
T - refers to the true score

Distance Accuracy

Distance accuracy is the average absolute value of the deviation of the obtained ratings from the true scores. It reflects the level difference between subjects ratings and expert ratings. It is computed as

$$\text{Distance accuracy} = \frac{\sum_{j=1}^{r}\left[\frac{\left(\sum_{i=1}^{d}|T_{ij} - R_{ijk}|\right)}{d}\right]}{r}$$

Differential Accuracy

The differential accuracy (DA) measure (Borman, 1979; Cronbach, 1955) provides accuracy scores for each rater on each job dimension. DA is usually determined by correlating each rater's responses with the true score for a dimension or construct. Hence, each rater receives a DA score for each dimension. The Fisher r to Z transformation is then applied to each DA correlation. DA reflects the parallelism between subjects' and experts' ratings. Bernardin and Cooke (1992) have argued that DA is the most important criterion for assessing the effectiveness of ratings (see also Bernardin & Kane, in press). These accuracy scores will then be used in analysis of variances to assess purpose and neediness effects on accuracy.

Merit-raise Decision

The merit-raise decision will be operationalized as follows

Decision is whether to give a merit raise.

Oppose means - merit raise should not be given
Support means - merit raise should be given

1	2	3	4	5	6	7	8	9	10
strongly oppose		somewhat oppose		neutral		somewhat support		strongly support	
(merit raise should not be given)								(merit raise should be given)	

Training Decision

The training decision will be operationalized as follows

Decision is whether to send subordinate for training.

Oppose means - training not required
Support means - training required.

1	2	3	4	5	6	7	8	9	10
strongly oppose		somewhat oppose		neutral		somewhat support		strongly support	
(training not required)							(training required)		

Manipulation Checks

Appraisal Purpose

Appraisal ratings can be used for several purposes. Most of these purposes can be placed in two broad categories involving either between comparisons or within comparisons. Purposes including merit-raises, promotions, and layoffs require the rater to compare between ratees whereas feedback, training and developmental purposes require raters to compare ratees with themselves, over time. In this study, the influence of two appraisal purposes, merit-raise and training are investigated.

These two purposes were chosen as most studies on purpose effects have investigated the influence of these two purposes. Additionally, while merit-raise purpose belongs to between-ratee class of purposes, training is representative of the within-ratee class of purposes. The following item was administered to verify the efficacy of the purpose manipulation.

For which of the following purposes does this company use performance appraisal ratings.

____ a. Merit-raise ___ c. Documentation

____ b. Training ___ d. Promotion

Budget Constraints

Rater's perception of consequences of ratings for the ratee were manipulated by varying the amount of funds available in the pay-raise budget such that in combination with appraisal purpose the consequences would be either low or high. Accordingly, the pay-raise budget contained either no funds or plenty of funds (manipulation discussed earlier). The following questions were administered to verify the efficacy of these manipulations.

The performance ratings that you just provided will significantly affect your subordinate Pat.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

The performance ratings you provided will not have any consequences for Pat.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

Please use the space below to write down the consequences of the ratings provided by you for Pat.

_____

_____

_____

Compared to last year, this year more funds are available for providing pay increases.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

This year due to lack of funds raises cannot be provided.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

Performance

The study was presented to the subjects as one designed to investigate the effectiveness of critical incident method of performance appraisal. Each subject was presented with information regarding the performance of two subordinates, Pat and Chris. This information was presented in the form of critical incidents. Twenty-five critical incidents captured the performance of each subordinate. Additionally, the oreder in which crtical incidents capturing Pat's and Chris's performance was counterbalanced within each condition. Subjects were instructed to rate the performance of Pat only. The following items were administered to verify the efficacy of the operationalization of performance.

Pat is a better performer than Chris

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

Chris is a better performer than Pat

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

Data Analysis Procedure

After ascertaining the efficacy of manipulation checks, the specific hypotheses were tested using traditionally accepted statistical data analysis approaches such as analyses of variance.

CHAPTER FOUR

RESEARCH RESULTS

INTRODUCTION

This chapter presents findings of the study. The presentation of results is organized by the following sections 1. manipulation checks 2. order effect 3. multivariate analysis of variance 4. analysis of variance 5. hypotheses testing. In the interest of brevity, throughout this chapter the names of variables/indices described in chapter three are abbreviated as follows.

ASSIST: The five-item scale measuring subjects' willingness to assist Pat by a. giving overtime b. arranging for a loan c. reducing workload d. assigning easy tasks e. providing coaching.

CONSEQ: The two-item scale measuring subjects' perception of consequences (of their performance ratings) for Pat.

CCONSEQ: The two-item scale measuring the extent to which subjects considered consequences of a. performance ratings while evaluating Pat's performance b. personnel decision while recommending the same for Pat.

PURPOSE: The manipulated independent variable 'purpose of appraisal.' Appraisal purpose (PURP) was either merit-raise (MR) or training (TRG).

FUNDS: The manipulated independent variable 'availability of funds/pay-raise budget.' Funds (FNDS) were either plentiful (PF) or not available (NF).

SM: The independent variable self-monitoring also served as a blocking variable. SM was either low (L) or high (H).

LENIENCY: Leniency of ratings provided by subjects.

DISTACCU: A measure of accuracy of subjects' ratings. For an elaborate discussion of distance accuracy, please see chapter three.

DA: A measure of accuracy of subjects' ratings. For an elaborate

discussion of differential accuracy, please see chapter three.

DECISION: Decision made by subjects for Pat. Subjects assigned to MR condition made a decision regarding pay-raise and those assigned to the TRG condition made a decision regarding training.

MANIPULATION CHECKS

In this study, two variables, PURPOSE and FUNDS were manipulated. In order to test the efficacy of manipulations, T tests were performed. A significant difference in subjects' perception of appraisal purpose based upon whether the subject was assigned to MR or TRG condition would yield evidence of a successful manipulation of purpose. As expected, subjects in the MR condition rated purpose significantly lower than subjects in the TRG condition (Satterthwaite $T = -49.0478$, $p < .001$, means = 1.019/1.994 (MR coded as 1/TRG coded as 2).

Similarly, a significant difference in subjects' perception of availability of funds based upon whether the subject was assigned to the PF or NF condition would yield evidence of a successful manipulation of funds. As expected, subjects in the PF condition rated availability of funds higher than subjects in the NF condition ($T = 32.5697$, $p < .0001$, means = 7.716/2.286 (PF/NF). Consequently, both manipulations were deemed successful.

Means and standard deviations of dependent variables LENIENCY, DISTACCU, DA, and DECISION are presented in table one.

TABLE I

MEANS AND STANDARD DEVIATIONS OF DEPENDENT VARIABLES

CELL #	CONDITION	LENIENCY	DISTACCU	DA	DECISION
1	L.MR.PF	-0.565 (.49)	0.755 (.42)	0.4087 (.58)	3.45 (1.97)
2	L.MR.NF	-0.565 (.558)	0.785 (.48)	0.13 (.67)	2.75 (1.63)
3	L.TRG.PF	-0.465 (.36)	0.715 (.31)	0.194 (.69)	8.38 (1.03)
4	L.TRG.NF	-0.41 (.432)	0.61 (.335)	0.29 (.58)	8.35 (.95)
5	H.MR.PF	-1.32 (.545)	1.39 (.47)	0.063 (.47)	5.85 (1.41)
6	H.MR.NF	-0.246 (.49)	0.67 (.39)	0.08 (.57)	2.18 (1.17)
7	H.TRG.PF	-0.52 (.61)	0.89 (.47)	-0.03 (.64)	8.3 (.97)
8	H.TRG.NF	-0.41 (.497)	0.66 (.38)	0.39 (.49)	8.08 (1.31)

Note: Numbers appearing in paranthesis represent standard deviations

ORDER EFFECT

Recall that subjects were provided with critical incidents capturing performances of two subordinates, Pat and Chris. All subjects rated Pat's performance only. However, the order in which critical incidents capturing Pat's and Chris's performance were presented was counterbalanced within each cell.

A significant difference on dependent variables based upon whether the subject was presented with Pat's performance first (coded 0) or

Chris's performance first (coded 1) would indicate order effects. No order effects were predicted. As expected, the means of each of the dependent variables were not significantly different when the order in which the performance information presented was varied (LENIENCY: T = 0.4577, p > .6475, means = -0.544/-0.574 (0/1); DISTACCU: T = -0.1428, p > .8866, means = 0.804/0.811 (0/1); DA: T = -0.6526, p > .5145, means = 0.171/0.216 (0/1); DECISION: T = 0.4446, p > .6569, means = 5.989/5.844 (0/1)).

MULTIVARIATE ANALYSIS OF VARIANCE

Before performing MANOVA, correlations among dependent variables were examined. Three correlations were significant. LENIENCY was negatively correlated with DISTACCU (r = -0.87745, p = .0001) and moderately correlated with DA (r = 0.11802, p = .0421). DISTACCU was negatively correlated with DA (r = -0.44331, p = .0001). Lower values indicate more leniency, and higher distance accuracy. A lower correlation indicates lower differential accuracy. Therefore, these correlations are logical and are in the expected direction.

The omnibus test, multivariate analysis of variance, examined the null hypothesis of no overall effect for the main effects of the three independent variables and the interactions among them on LENIENCY, DISTACCU, DA, and DECISION combined. The results of MANOVA are presented in table three. Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace and Roy's Greatest Root yielded same results. Therefore, only values of the Wilks' Lambda statistic are reported.

TABLE II

MULTIVARIATE ANALYSIS OF VARIANCE

VARIABLE	VALUE (WILKS'LAMBDA)	F VALUE	Pr > F
SM	0.92	6.154	.0001
PURP	0.229	239.708	.0001
SM*PURP	0.954	3.404	.0097
FNDS	0.81	16.77	.0001
SM*FNDS	0.863	11.365	.0001
PURP*FNDS	0.819	15.769	.0001
SM*PURP*FNDS	0.915	6.632	.0001

Note: Degrees of freedom for numerator and denominator were 4 and 286 respectively.

ANALYSIS OF VARIANCE

An analysis of variance was performed with each of the dependent variables. A summary of these results are provided in table three. The complete results of analysis of variance for the dependent variables LENIENCY, DISTACCU, DA, MR DECISION, and TRG DECISION are provided in tables four, five, six, seven and eight respectively.

TABLE III

SUMMARY OF ANALYSIS OF VARIANCE RESULTS

	LENIENCY	DISTACCU	DA	MR	TRG
SM	4.60 (.03)	16.23 (.0001)	3.29 (0.07)	14.47 (.0002)	1.06 (.304)
PURP	16.09 (.0001)	15.46 (.0001)	0.47 (.4915)	----	----
SM*PURP	3.11 (.07)	2.57 (.1096)	1.18 (.2781)	----	----
FNDS	29.22 (.0001)	31.07 (.0001)	0.82 (.3655)	76.69 (.0001)	0.54 (.462)
SM*FNDS	24.36 (.0001)	22.80 (.0001)	5.32 (.02)	34.73 (.0001)	0.34 (.557)
PURP*FNDS	15.81 (.0001)	3.43 (.06)	8.34 (.004)		
SM*PURP*FNDS	20.04 (.0001)	11.07 (.0001)	0.01 (.9100)		

Note: Numbers represent F values and probabilities (Pr > F)

TABLE IV

ANALYSIS OF VARIANCE WITH LENIENCY
AS DEPENDENT VARIABLE

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
MODEL	7	28.946	4.135	16.18	.0001
ERROR	311	79.507	0.256		
CORR. TOTAL	318	108.453			

R-SQUARE: 0.2669

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
SM	1	1.175*	1.175	4.60	.0328
		1.157**	1.157	4.52	.0342
		1.129***	1.129	4.42	.0363
PURP	1	4.113	4.113	16.09	.0001
		4.046	4.046	15.83	.0001
		4.003	4.003	15.66	.0892
SM*PURP	1	0.796	0.796	3.11	.0786
		0.756	0.756	2.96	.0866
		0.743	0.743	2.91	.0892
FNDS	1	7.469	7.469	29.22	.0001
		7.469	7.469	29.22	.0001
		7.587	7.587	29.68	.0001
SM*FNDS	1	6.227	6.227	24.36	.0001
		6.259	6.259	24.48	.0001
		6.295	6.295	24.62	.0001
PURP*FNDS	1	4.043	4.043	15.81	.0001
		4.043	4.043	15.81	.0001
		4.072	4.072	15.93	.0001
SM*PURP*FNDS	1	5.123	5.123	20.04	.0001
		5.123	5.123	20.04	.0001
		5.123	5.123	20.04	.0001

Note * refers to Type I sum of squares
** refers to Type II sum of squares
*** refers to Type III & Type IV sum of squares (in a balanced design types III & IV sum of squares will be same).

TABLE V

ANALYSIS OF VARIANCE WITH DISTACCU
AS DEPENDENT VARIABLE

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
MODEL	7	17.383	2.483	14.66	.0001
ERROR	311	52.673	0.169		
CORR. TOTAL	318	70.055			

R SQUARE: 0.248

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
SM	1	2.748*	2.748	16.23	.0001
		2.733**	2.733	16.14	.0001
		2.705***	2.705	15.97	.0001
PURP	1	2.619	2.619	15.46	.0001
		2.575	2.575	15.20	.0001
		2.559	2.559	15.11	.0001
SM*PURP	1	0.436	0.436	2.57	.1096
		0.415	0.415	2.45	.1184
		0.41	0.41	2.42	.1209
FNDS	1	5.262	5.262	31.07	.0001
		5.262	5.262	31.07	.0001
		5.322	5.322	31.42	.0001
SM*FNDS	1	3.862	3.862	22.80	.0001
		3.872	3.872	22.86	.0001
		3.889	3.889	22.96	.0001
PURP*FNDS	1	0.581	0.581	3.43	.0650
		0.581	0.581	3.43	.0650
		0.588	0.588	3.47	.0635
SM*PURP*FNDS	1	1.875	1.875	11.07	.0010
		1.875	1.875	11.07	.0010
		1.875	1.875	11.07	.0010

Note * refers to Type I sum of squares
** refers to Type II sum of squares
*** refers to Type III & Type IV sum of squares (in a balanced design types III & IV sum of squares will be same).

TABLE VI

ANALYSIS OF VARIANCE WITH DA
AS DEPENDENT VARIABLE

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
MODEL	7	6.753	0.965	2.78	.0083
ERROR	289	100.367	0.347		
CORR. TOTAL	296	107.119			

R-SQUARE: 0.063

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
SM	1	1.143*	1.143	3.29	.0707
		1.221**	1.221	3.51	.0618
		1.169***	1.169	3.37	.0675
PURP	1	0.165	0.165	0.47	.4915
		0.127	0.127	0.37	.5459
		0.151	0.151	0.43	.5102
SM*PURP	1	0.410	0.410	1.18	.2781
		0.364	0.364	1.05	.3068
		0.365	0.365	1.05	.3060
FNDS	1	0.284	0.284	0.82	.3665
		0.284	0.284	0.82	.3665
		0.295	0.295	0.85	.3577
SM*FNDS	1	1.849	1.849	5.32	.0217
		1.827	1.827	5.26	.0225
		1.827	1.827	5.26	.0225
PURP*FNDS	1	2.898	2.898	8.34	.0042
		2.898	2.892	8.34	.0042
		2.899	2.899	8.35	.0042
SM*PURP*FNDS	1	0.004	0.004	0.01	.9100
		0.004	0.004	0.01	.9100
		0.004	0.004	0.01	.9100

Note * refers to Type I sum of squares
** refers to Type II sum of squares
*** refers to Type III & Type IV sum of squares (in a balanced design types III & IV sum of squares will be same).

TABLE VII

ANALYSIS OF VARIANCE WITH MR-DECISION
AS DEPENDENT VARIABLE

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
MODEL	3	312.464	104.155	41.96	.0001
ERROR	155	384.719	2.482		
CORR. TOTAL	158	697.182			

R-SQUARE: 0.448

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
SM	1	35.909*	35.909	14.47	.0002
		34.869**	34.869	14.05	.0003
		34.172***	34.172	13.77	.0003
FNDS	1	190.347	190.347	76.69	.0001
		190.347	190.347	76.69	.0001
		191.975	191.975	77.35	.0001
SM*FNDS	1	86.208	86.208	34.73	.0001
		86.208	86.208	34.73	.0001
		86.208	86.208	34.73	.0001

Note * refers to Type I sum of squares
** refers to Type II sum of squares
*** refers to Type III & Type IV sum of squares (in a balanced design types III & IV sum of squares will be same).

TABLE VIII

ANALYSIS OF VARIANCE WITH TRG-DECISION
AS DEPENDENT VARIABLE

SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
MODEL	3	2.25	0.75	0.65	.5833
ERROR	156	179.65	1.152		
CORR. TOTAL	159	181.9			
R-SQUARE: 0.012					
SOURCE	DF	SUM OF SQ.	MEAN SQ.	F VALUE	Pr > F
SM	1	1.225*	1.225	1.06	.3040
		1.225**	1.225	1.06	.3040
		1.225***	1.225	1.06	.3040
FNDS	1	0.625	0.625	0.54	.4624
		0.625	0.625	0.54	.4624
		0.625	0.625	0.54	.4624
SM*FNDS	1	0.4	0.4	0.35	.5565
		0.4	0.4	0.35	.5565
		0.4	0.4	0.35	.5565

Note * refers to Type I sum of squares
** refers to Type II sum of squares
*** refers to Type III & Type IV sum of squares (in a balanced design types III & IV sum of squares will be same).

HYPOTHESIS TESTING

Hypotheses 1, 2, 3, and 4 were tested with dependent variables LENIENCY, DISTACCU, and DA. Hypotheses 5 was tested with the dependent variable DECISION. Hypotheses 6, 7 and 8 were tested with the dependent variable DECISION (recommendation for pay-raise). Hypotheses 9, 10, and 11 were tested with the dependent variable DECISION (recommendation for training).

H1: Hypothesis 1 predicted that ratings provided for MR purpose would be more lenient and less accurate than those provided for training purposes. Hypothesis 1 was supported for LENIENCY ($F = 16.09$, $p > .0001$, means -.67/-.45 (MR/TRG), effect size (d) = .55, $R = .27$) and DISTACCU ($F = 15.46$, $p > .0001$, means 0.9/0.72 (MR/TRG), $d = .44$, $R = .213$), but not for DA ($F = 0.47$, $p > .4915$, mean correlations .17/.23 (MR/TRG), effect size (q) = .045).

H2: Hypothesis 2 predicted that ratings provided in the PF comdition would be more lenient and less accurate than those provided in the NF condition. Hypothesis 2 was supported for LENIENCY ($F = 29.22$, $p > .0001$, means -.72/-.41 (PF/NF), $d = .61$, $R = .29$), DISTACCU ($F = 31.07$, $p > .0001$, means .94/.68 (PF/NF), $d = .49$, $R = .24$), but not for DA ($F = 0.82$, $p > .3655$, mean correlations .17/.223 (PF/NF), $q = .063$).

H3: Hypothesis 3 predicted that ratings provided by high self-monitors would be more lenient and less accurate than those provided by low self-monitors. Hypothesis 3 was supported for LENIENCY ($F = 4.60$, $p > .03$, means -.50/-.62 (Low self-monitor:LSM/high self-monitor:HSM), $d = .31$, $R = .1509$), DISTACCU ($F = 16.23$, $p > .0001$, means .72/.9 (LSM/HSM), $d = .45$, $R = .22$), and DA ($F = 3.29$, $p > .07$, mean correlations .254/.142 (LSM/HSM), $q = .125475$).

H4: Hypothesis 4 predicted that purpose, funds, and self-monitoring will interact to affect leniency and accuracy. For low self-monitors no significant funds*purpose interaction was expected. Ratings provided by

high self-monitors were expected to be a. most lenient in the MR-PF condition b. least lenient (i.e. stringent) in the MR-NF condition, and c. least accurate in the above two conditions. Figures 1, 2, and 3 depict the predicted three-way interaction with LENIENCY, DISTACCU, and DA respectively, as dependent variables.

The SM X PURP X FNDS interaction was significant at the multivariate level (Wilks' Lambda = 0.915, F = 6.632, p > .0001). At the univariate level the interaction was significant for dependent variables LENIENCY (F = 20.64, p > .0001) and DISTACCU (F = 11.07, p > .0001) but not for DA (F = 0.01, p > .9100). Follow-up tests were conducted to test the interaction captured by the hypothesis. Such tests were performed for dependent variables LENIENCY, DISTACCU, and DA. The LSMEANS procedure was used to make pre-planned comparisons between means.

LENIENCY:

.....

insert figure 1 about here

.....

As predicted, purpose and funds interacted to influence ratings of high self-monitors (d = .888445) but did not interact to influence ratings provided by low self-monitors (d = .0587651). As expected, purpose and funds did not interact to influence ratings provided by low self-monitors (cell 1 vs cell 4, p > .1714). Additionally, as expected ratings provided by high and low self-monitors did not differ when appraisal purpose was training (cell 3 vs cell 7, p > .6586; cell 4 vs cell 8, p > .9648). When appraisal purpose was merit-raise, ratings provided by high and low self-monitors differed as predicted. More

specifically, in the MR-PF condition, ratings provided by high self-monitors were significantly more lenient than those provided by low self-monitors (cell 1 vs cell 5, $p > .0001$). Furthermore, ratings provided by high self-monitors were most lenient in the MR-PF condition. In the MR-NF condition, ratings provided by high self-monitors were significantly less lenient than those provided by low self-monitors (cell 2 vs cell 6, $p > .0054$). Additionally, ratings provided by high self-monitors were significantly more stringent in the MR-NF condition than those provided in the MR-PF and TRG-PF conditions. Thus, hypothesis 4 was fully supported with respect to leniency.

ACCURACY: With respect to accuracy, hypothesis 4 predicted no differences in accuracy of ratings across treatment combinations for low self-monitors. For high self-monitors, it was predicted that ratings in the MR-PF and MR-NF conditions will be less accurate than those in TRG-PF and TRG-NF conditions.

DISTANCE ACCURACY:

.....

insert figure 2 about here

.....

Contrary to expectations, ratings provided by low self-monitors differed in accuracy across treatment combinations. Specifically, the difference between the farthest means (cell 2 & cell 4) was significant at $p > .0581$ with the LSMEANS procedure. As expected, ratings provided by low and high self-monitors in the TRG-NF condition did not differ in accuracy. However, contrary to expectations, high self-monitors provided less accurate ratings in the TRG-PF condition than low self-monitors

(LSMEANS cell 3 vs cell 7, $p > .0581$).

As expected, ratings provided by high self-monitors in the MR-PF condition were significantly less accurate than those provided by low self-monitors in the same condition (cell 1 vs cell 5, $p > .0001$). Furthermore, ratings provided by high self-monitors were least accurate in the MR-PF condition than in any other treatment combinations (cell 5 vs 6, 7, & 8, $p > .0001$). Ratings provided by high self-monitors in the MR-NF condition were also expected to be less accurate than those provided by low self-monitors. This, however was not the case (cell 2 vs cell 6, $p > .2023$). One interesting observation was that high self-monitors were less accurate when funds were plentiful, than when no funds were available suggesting that the funds variable may be exerting more influence on ratings than the purpose variable.

DIFFERENTIAL ACCURACY:

.....

insert figure 3 about here

.....

This measure of accuracy was obtained by correlating ratings provided by subjects and experts. The Fisher r to z transformation was used to transform these Pearson correlations to z correlations. When the z-test was used to test predicted differences between correlations, none of the correlations were significantly different. Thus, hypothesis 4 was fully supported for the dependent variable leniency, partially supported for distance accuracy, but not supported for differential accuracy.

H5: Hypothesis 5 predicted that merit raise recommendations would be more inflated than training recommendations. Unlike previous studies

that used standard deviation (of subjects' recommendations as data points) to test for purpose effects, we used standardized values of recommendations. A subject's recommendation for merit raise (or training) is actually a data point in the distribution of merit raise (or training) recommendations. Hence, recommendations for merit raise and training are data points in two different distributions. Since these two distributions (and hence the data points contained in them) are not directly comparable (see Zedeck & Cascio, 1982) we obtained standardized values of merit raise and training recommendations ($z = \text{sigma } (x - X)/s$). To test for purpose effects we compared the mean standardized estimate of merit raise ($z = .82$) with that of training ($z = .05$). Effect size computed as the difference between these two mean standardized values ($d = z\text{-}z = .77$) indicates that recommendations for merit raise and training are .77 standard deviations apart. Since a 'd' of size .8 is regarded as a large effect (Cohen, 1977), we conclude strong support for the hypothesized purpose effect.

The 2 X 2 analysis of variance with merit raise decision was significant ($F = 41.96$, $p > .0001$), whereas, that with training decision as the dependent variable was not ($F = 0.65$, $p > .5833$).

H6: Hypothesis 6 was fully supported. As expected, merit raise recommendations in the plenty of funds (PF) condition were significantly stronger than those in the no funds (NF) condition ($F = 76.69$, $p > .0001$, means 4.65/2.452 (PF/NF), $d = 1.397$, $R = .573$).

H7: Hypothesis 7 predicted that high self-monitors (HSM) would make stronger merit raise recommendations than low self-monitors (LSM). As predicted, high self-monitors did make stronger recommendations than low self-monitors ($F = 14.47$, $p > .0002$, means 4.0145/3.088 (HSM/LSM), $d = .589$, $R = .283$).

H8: Hypothesis 8 predicted that self-monitoring and funds would interact to influence merit raise recommendations. This hypothesis was supported ($F = 34.73$, $p > .0001$, $d = .94$, $R = .424$). This interaction is shown in figure 4.

.....

insert figure 4 about here

.....

The LSMEANS procedure was used to make pre-planned comparisons between means. As predicted, recommendations by high self-monitors were significantly stronger than those by low self-monitors in the PF condition (cell 1 vs cell 5, $p > .0001$, means 3.45/5.85 (LSM/HSM)) and weaker in the NF condition (cell 2 vs cell 6, $p > .0729$, means 2.725/2.179 (LSM/HSM)) than low self-monitors. High self-monitors made the strongest recommendations in PF condition than in NF condition (cell 5 vs cell 6, $p > .0001$, means 5.85/2.179 (PF/NF)). Availability of funds was expected to have no influence on recommendations of low self-monitors. Contrary to expectations, recommendations by low self-monitors were considerably stronger in PF condition than in NF condition (cell 1 vs cell 2, $p > .0167$, means 3.45/2.725 (PF/NF)).

H9: Hypothesis 9 predicted that training recommendations made in the PF condition would not be significantly different from those made in the NF condition. As expected, funds had no influence on training recommendations ($F = 0.54$, $p > .4624$, means 8.34/8.213 (PF/NF), $d = .1165$, $R = .058$).

H10: As predicted in hypothesis 10, self-monitoring had no effect on training recommendations ($F = 1.06$, $p > .3040$, means 8.36/8.19 (LSM/HSM), $d = .16$, $R = 081$).

H11: Hypothesis 11 predicted that self-monitoring would not interact with funds to influence training recommendations. As expected no interaction was noted ($F = 0.34$, $p > .5565$, $d = .096$, $R = .048$).

.....

insert figure 5 about here

.....

Please note that in hypotheses 9, 10, and 11 the research hypothesis was also the null hypothesis. The nonsignificant results suggest that the difference between means is negligible or trivial. In terms of effect size, .2 or less may be regarded as trivial (Cohen, 1977). Given an N of 40 subjects per cell, for an effect size of .2, power to correctly reject the null reduces to .22. Consequently, beta, the probability of incorrectly accepting the null increases to .78. Accepting the null hypothesis, that effect size is trivial, at beta=.78 is a risky endeavor. This situation can be avoided by setting beta at .2 and testing the hypothesis at power of .8. At this level, if the results

are nonsignificant, it would be proper to conclude that the population effect size is not more than .2, that is, it is negligible. This conclusion can be offered as significant at the specified beta level. This approach is functionally equivalent to affirming the null hypothesis with a controlled error rate beta (Cohen, 1977). However, since 310 subjects per cell would be required to test an effect size of .2 with beta at .2 (and power.8), this test could not be performed. Therefore, the nonsignificant results reported for hypotheses 9, 10 and 11 should be regarded as inconclusive. The results of the study are summarized in table nine.

TABLE IX

SUMMARY OF STUDY RESULTS

HYPOTHESES	DEPENDENT VARIABLES		
	LENIENCY	**DISTACCU**	**DA**
H1	**S** (d=.55) R=.265	**S** (d=.4352435) R=.2126	**NS** (q=.045)
H2	**S** (d=.611) R=.29	**S** (d=.4929456) R=.239	**NS** (q=.063)
H3	**S** (d=.305) R=.1509	**S** (d=.45) R=.218	**S** (q=.125)
H4 FNDS*PURP	**S**	**PS**	**NS**
LSM	(d=.059) R=.029	(d=.1629268) R=.08	(q=.1897)
HSM	(d=.89) R=.4059	(d=.56) R=.2684	(q=.1798)

	PAY-RAISE	**TRAINING**
H5: S (d=.77) R=.359	Z = .82	Z = .05
H6: S (d=1.397) R=.573		NA
H7: S (d=.589) R=.283		NA
H8: S (d=.936) R=.424		NA
H9:	NA	S (d=.1165) R=.058
H10:	NA	S (d=.1623) R=.081
H11:	NA	S (d=.0961094) R=.048

Note: S - supported, PS - partially supported, NS - not supported, NA - not applicable, d - effect size (with means), q - effect size (with correlations), Z - mean standardized values.

FIGURE 1

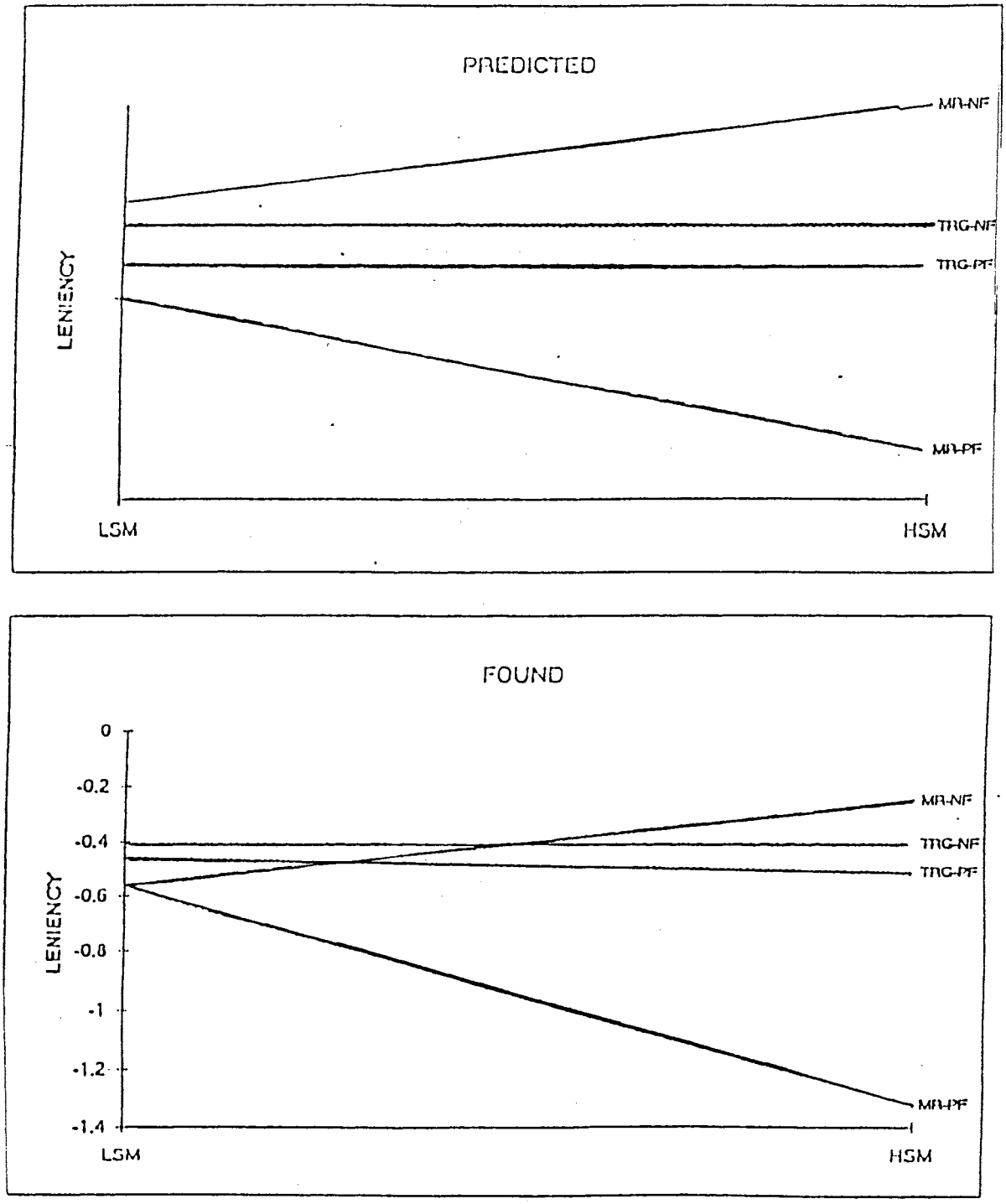

FIGURE 1. Three-way interaction of Self-Monitoring, Budget Constraints, and Appraisal Purpose with LENIENCY

FIGURE 2

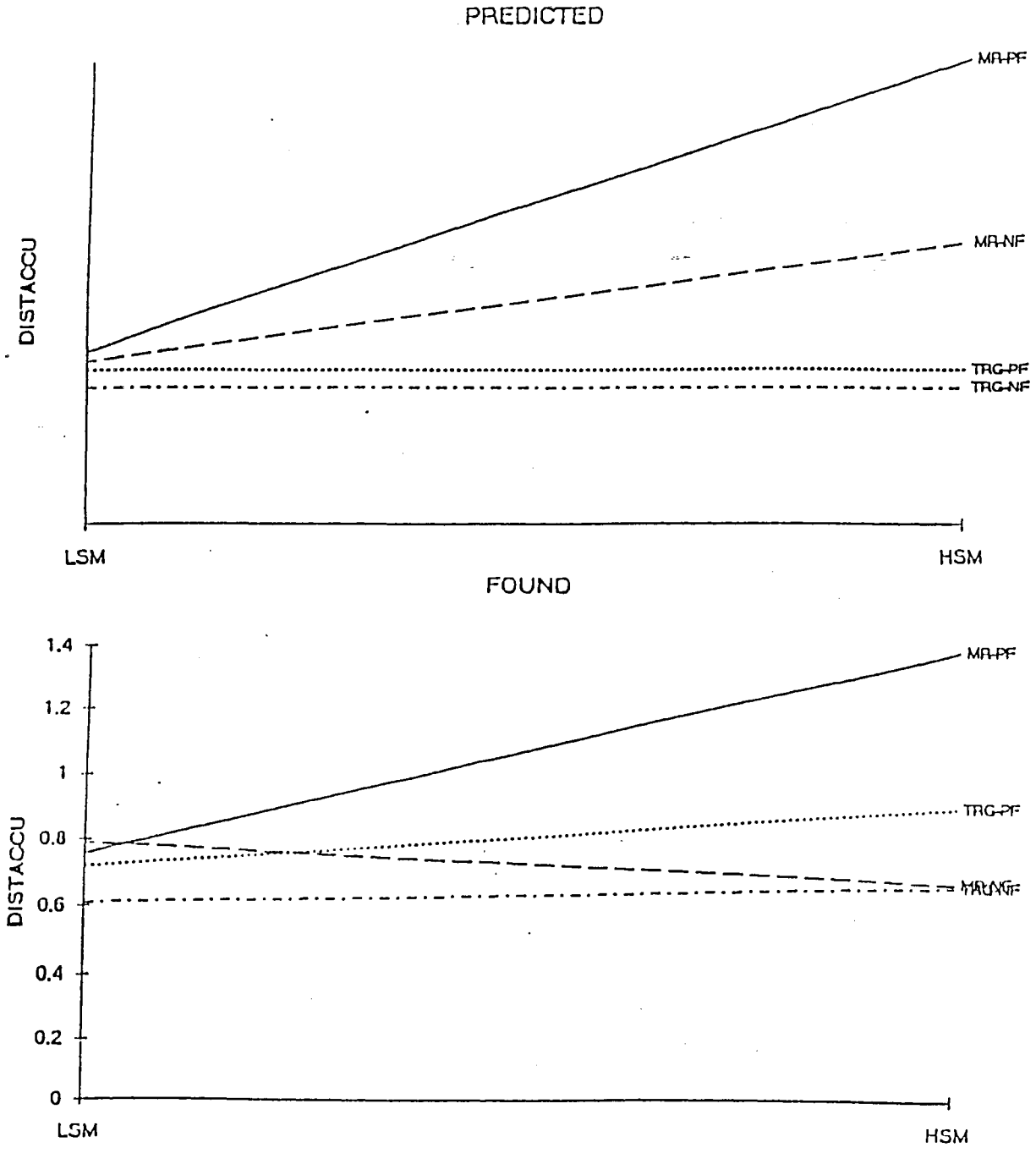

FIGURE 2. Three-way interaction of Self-Monitoring, Budget Constraints, and Appraisal Purpose with DISTACCU

FIGURE 3

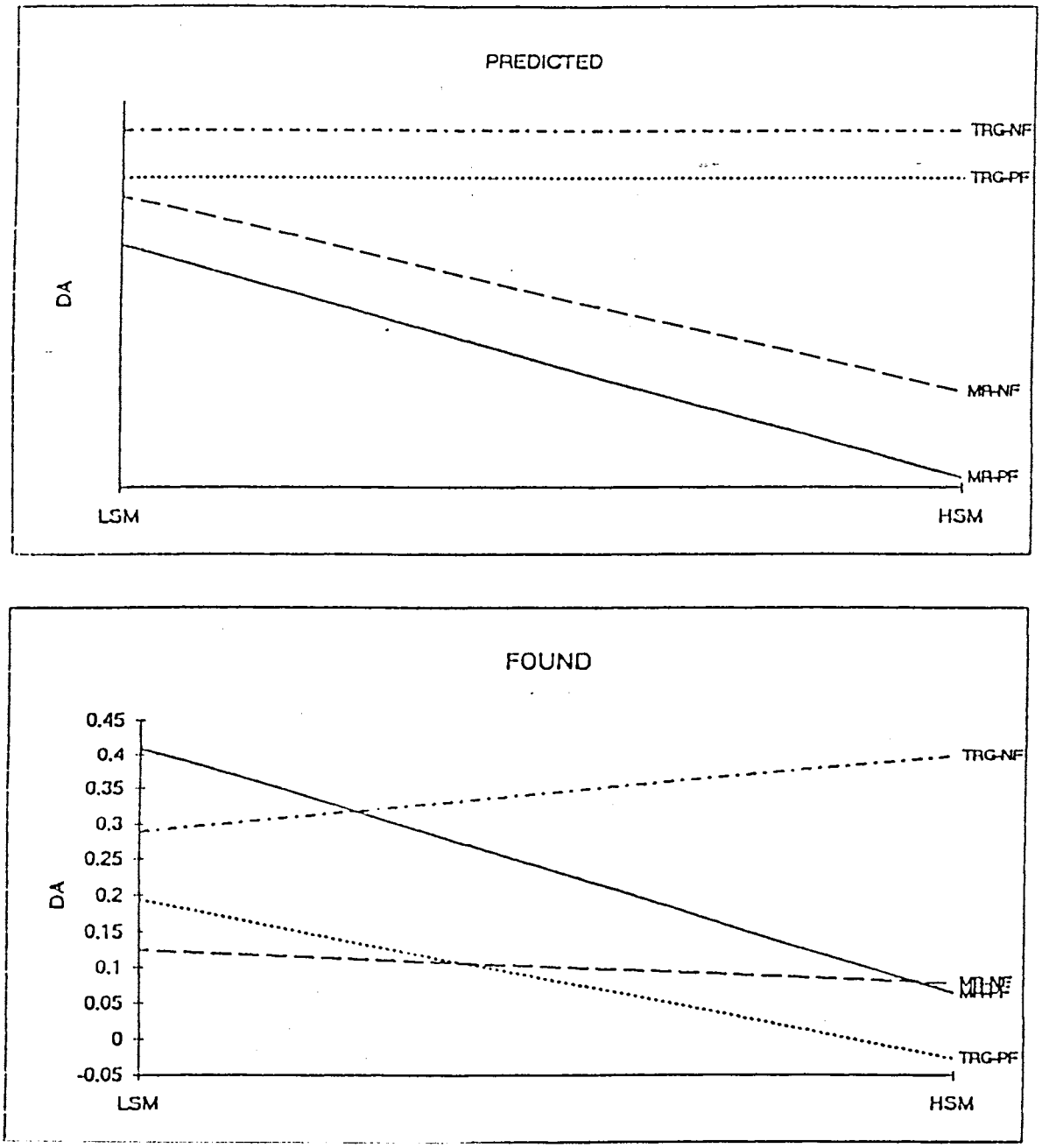

FIGURE 3. Three-way interaction of Self-Monitoring, Budget Constraints, and Appraisal Purpose with DA

FIGURE 4

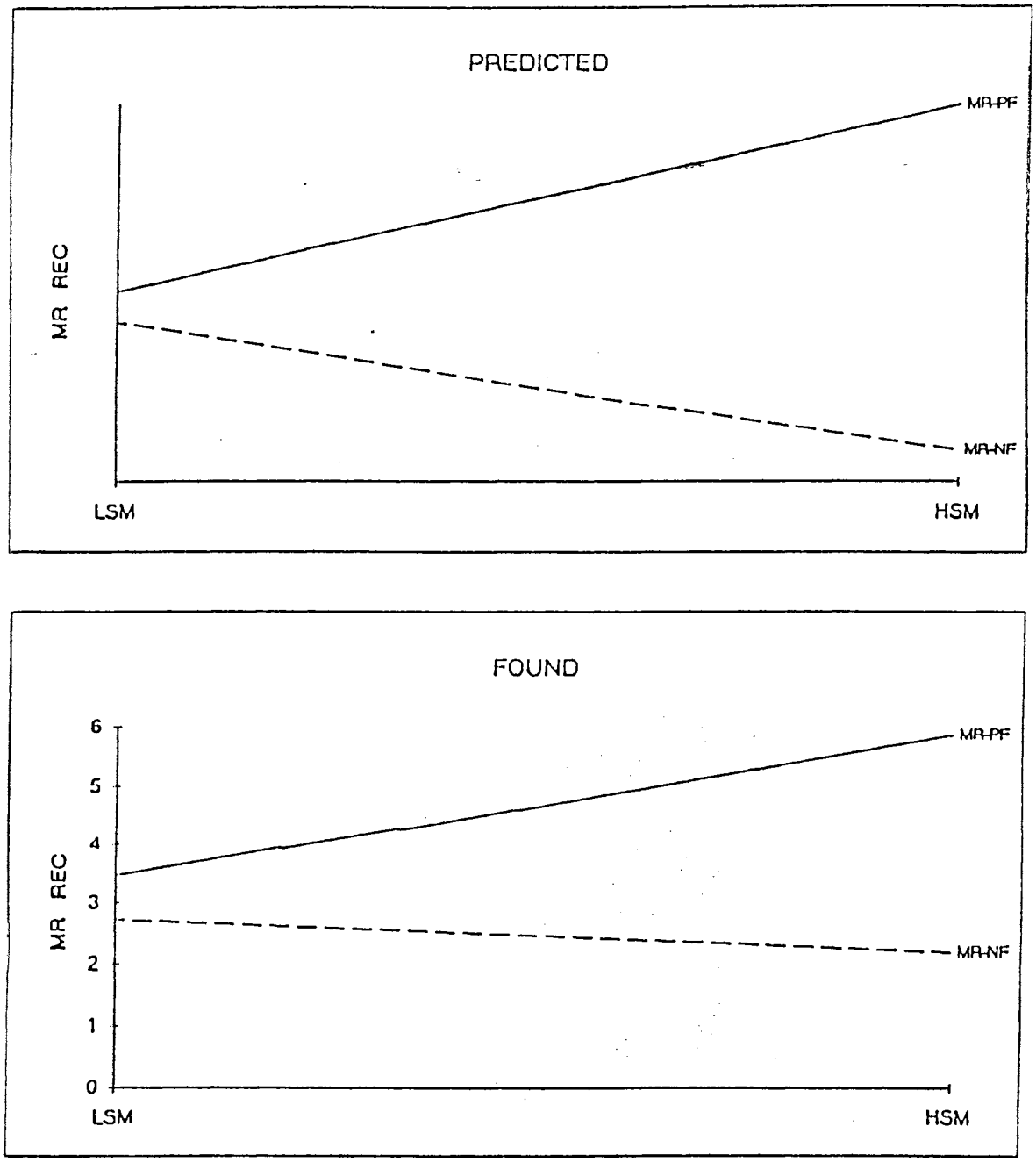

FIGURE 4. Two-way interaction of Self-Monitoring, and Budget Constraints with MR-Decision

FIGURE 5

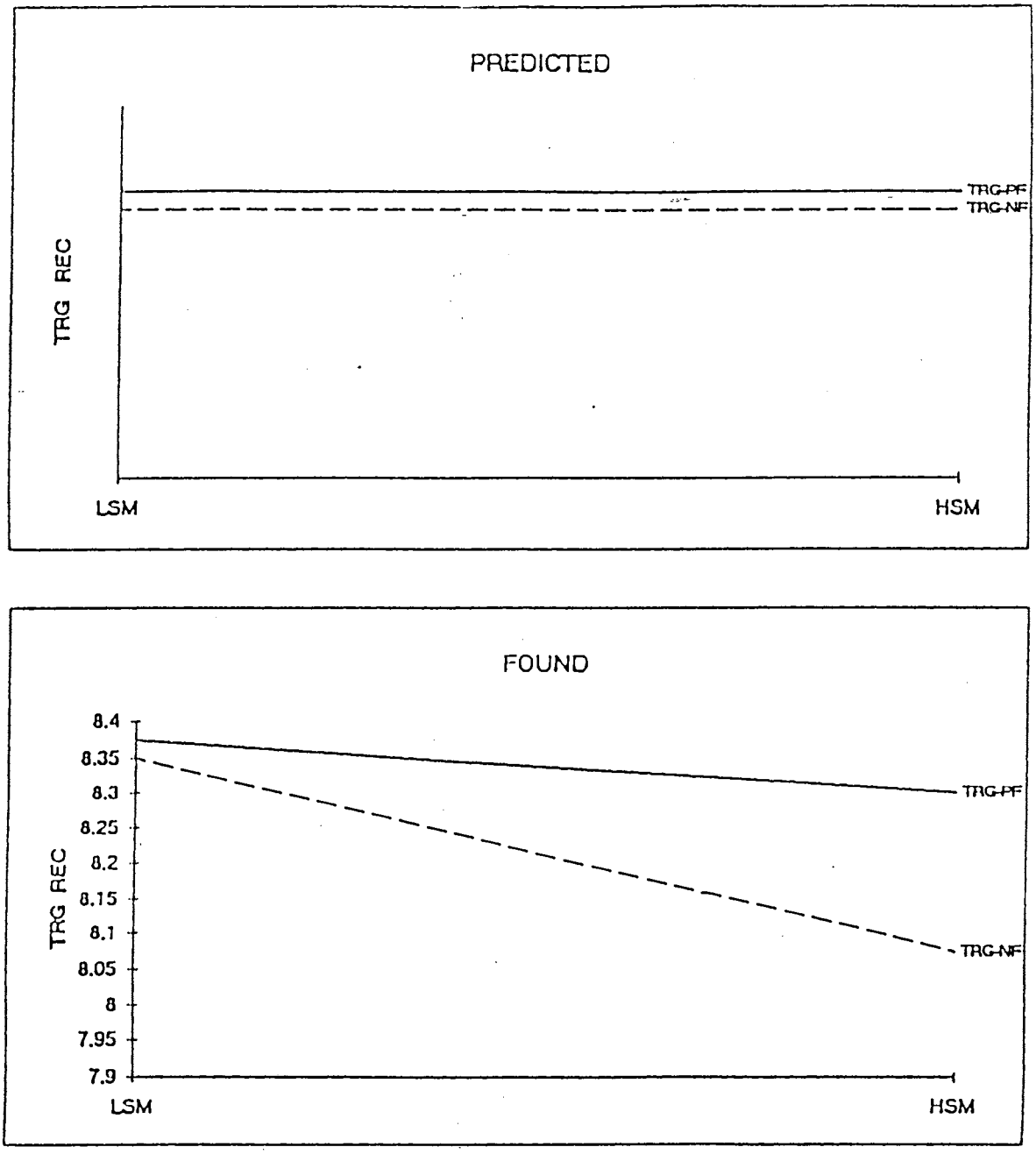

FIGURE 5. Two-way interaction of Self-Monitoring, and Budget Constraints with TRG-Decision

CHAPTER FIVE

SUMMARY AND CONCLUSIONS

This chapter will present a brief discussion of the study and results of the study. Next, implications for theory and future research will be presented followed by a discussion of limitations of the study. Finally, specific suggestions to enhance quality of performance appraisal ratings will be offered.

DISCUSSION

STUDY

It is important to reiterate the research problem and the major objectives of the study. Organizations use performance appraisal ratings for making several important administrative and personnel decisions. Since performance ratings are used for multiple purposes, many studies have investigated the effect of appraisal purpose on characteristics of ratings. These studies have reached contradictory results such that the relationship between appraisal purpose and leniency as well as accuracy of performance ratings is not clear. As a boundary variable, appraisal purpose has the potential to limit the external validity of research findings as performance ratings obtained for research purposes may be more or less accurate and/or lenient (severe) than those obtained for administrative purposes. Consequently, suggestions based on ratings obtained for research purposes may be of little value to practitioners who typically obtain ratings for making several important administrative

and personnel decisions. Because of the theoretical and practical significance of the research problem, this study was designed to address and attempt to understand the reasons for the contradictory results reported for appraisal purpose.

After analyzing the reasons for the inconsistent results, two plausible explanations were put forth. First, previous studies had expected purpose effects as different purposes may elicit different levels of motivation from raters to provide accurate (or inaccurate) ratings. Though focused on the motivational influenec of appraisal purpose, these studies had inadvertantly operationalized appraisal purpose as if it were a purely cognitive variable, thus ignoring it's motivational influence. The motivational influence of appraisal purpose on rating characteristics may be expected to operate through rater's perception of consequences of providing accurate ratings such that if consequences of providing accurate ratings are minimal raters may not be as influenced by purpose as much as they would if consequences were significant. Although related, purpose and consequences are distinct constructs and previous studies have failed to make this distinction. Consequently, these studies did not systematically vary consequences. Therefore, it is likely that consequences may have served to confound the influence of appraisal purpose on rating characteristics leading to inconsistent results. This study avoided such confounding by explicitly manipulating appraisal purpose and another situational variable, availability of funds (pay-raise budget), such that consequences would vary systematically across treatment combinations of appraisal purpose and availability of funds.

Second, previous studies had ignored individual differences among raters with potential to moderate the influence of appraisal purpose on rating characteristics. Research on self-monitoring suggests that high self-monitors, in comparison to low self-monitors, are more likely to be influenced by extraneous information such as purpose, funds and more importantly, consequences of ratings while evaluating performance. In this study, this non-manipulated independent variable, self-monitoring, was also used as a blocking variable. In summary then, this study examined the influence of appraisal purpose (MR, TRG), availability of funds (PF, NF), and self-monitoring (LSM, HSM) on rating characteristics leniency and accuracy. Several specific hypotheses were proposed and tested in a laboratory study.

RESULTS

The major impetus of this study is the inconsistency of prior research investigating the effects of appraisal purpose primarily on leniency/severity of ratings but also on accuracy. For example, while Bernardin and Cooke (1992) found a significant relationship between appraisal purpose and leniency, Bernardin, Abbott and Cooper (1985) failed to find such a relationship. Resolution of this inconsistency is important not only from a researcher's internal validity concerns but also for external validity concerns that are important for organizational application. I contend that many of the studies investigating Wherry's hypothesis of greater leniency of ratings for administrative purposes than when they are for development or research purposes have inadvertantly manipulated consequences at the same time as purpose. Though related, purpose and consequences of ratings are

distinct concepts.

This study avoided such confounding by systematically varying consequences across treatment combinations of purpose (merit raise versus training) and budget constraints (plenty of funds versus none). Indeed, subjects' perceptions of consequences in these conditions (MR-PF: 6.7875, MR-NF: 6.4376, TRG-PF: 5.875, and TRG-NF: 4.6125) provided further evidence that efforts to systematically vary consequences were successful. The means are in the predicted direction and five out of the seven possible T tests were significant. Additionally, it should be pointed out that these results are also consistent with the basic tenets of self-monitoring theory (Snyder, 1977, & 1987). As expected, low and high self-monitors did not differ in their perception of consequences of ratings for Pat (Satterthwaite $T = 0.9575$, $p > .3391$, means 6.031/5.825 (LSM/HSM)). Support for hypothesis 4 was obtained primarily because consequences varied across treatment combinations and high self-monitors not only considered consequences of their ratings and decisions for Pat more than low self-monitors [($T = -2.3440$, $p < .009$, means 2.9625/3.2281 (LSM/HSM); ($T = -1.8477$, $p < .032$, means 3.1125/3.375 (LSM/HSM))], but were also more willing to assist Pat than low self-monitors ($T = -1.9425$, $p < .02648$, means 2.836/2.934 (LSM/HSM)). These results not only provide additional support for the research findings reported in this study but also corroborate the underlying theoretical rationale for the study. We suggest that future research should carefully consider the effects of perceived consequences and the degree to which they may vary with other manipulations. Given the interdependent nature of the rater-ratee relationship, consequences of decisions are likely to be very high

and real in the field (see Bernardin & Villanova, 1986; Ilgen & Favero, 1985). Therefore, we suggest that the effect sizes reported in this study be regarded as underestimates of true effect sizes.

The support for hypothesis 1 for both leniency and distance accuracy by these data is consistent with that of many prior studies that found a strong effect of appraisal purpose on both leniency and at least one measure of rating accuracy. However, support for hypothesis 2, which predicted greater leniency when raise budgets were substantial than non-existant, has not been examined in prior studies. While support for this hypothesis is not surprising, the fact that effect sizes (see table nine) are slightly larger than those for purpose suggests that availability of funds may have a greater effect on ratings than purpose. Support for hypothesis 3, which predicted that high self-monitors will be more lenient and less accurate than low self-monitors corroborates theory as well as research on characteristics of high and low self-monitors.

The support for hypothesis 5 by these data is consistent with that of many prior studies that found a strong effect for appraisal purpose. This replication is valuable because unlike previous studies, in this study, consequences were systematically varied. This study also extends this stream of research by testing hypotheses investigating the influence of budget constraints and self-monitoring on merit raise and training decisions. The effects of these two variables on personnel decisions have not been examined in previous research.

As hypothesized, merit raise recommendations in the PF condition were significantly inflated than those in the NF condition. High self-

monitors made significantly stronger recommendations than low self-monitors. Thus, main effects for funds as well as self-monitoring were significant. The predicted two-way interaction between funds and self-monitoring was also supported as high self-monitors made stronger recommendations for merit raises when funds were substantial and weaker raises when funds were non-existent than low self-monitors. Although low self-monitors also strongly recommended raises when funds were substantial than when funds were non-existent, the effect size of the SM X FNDS interaction for high self-monitors (d = 2.84) was seven times larger than that for low self-monitors (d = .40). To further clarify this interaction, effect sizes were computed for budget constraints. The effect size of the SM X FNDS interaction for PF (d = 1.4) was only three and half times larger than that for NF (d = .385). Comparison of effect sizes suggests that raters' self-monitoring disposition overwhelmed the effect of budget constraints on merit raise recommendations.

As expected, neither funds nor self-monitoring had any effect on training recommendations. Additionally, as predicted, they did not interact to influence training recommendations.

IMPLICATIONS FOR THEORY AND RESEARCH

The extant literature on appraisal purpose has been too simplistic. It has focused exclusively on testing the association between appraisal purpose and characteristics of ratings without regard to potential moderators. By considering 'availability of funds,' a situational factor, and rater self-monitoring, this dissertation addressed the relationship from the much broader and richer interactional perspective.

Within the appraisal purpose literature, this study is the first attempt to examine the influence of rater personality and the rater X context interaction. This study is also the first attempt to address the contradictory results reported in the appraisal purpose literature. More importantly, in attempting to shift the literature on purpose effects which appears to have reached a "deadlock," this study study makes a significant contribution and paves the way for future research.

The significant results of the study encourage identification of other factors with potential to moderate the relationship between appraisal purpose and rating characteristics. Person factors with potential to moderate the relationship include trust in the appraisal process, empathy, self-consciousness, machiavellianism, coping-efficacy and supervisory dependence. Situational factors such as organizational climate, work group norms, organizational rewards and accountability as they relate to the appraisal process may also moderate the relationship between appraisal purpose and characteristics of performance ratings (see Jawahar, 1993 for a detailed discussion).

LIMITATIONS

College students served as subjects for this laboratory study. Additionally, the study employed 'paper-people' stimuli as opposed to live or videotaped stimuli. This section of the chapter addresses the potential of the setting, sample, and stimuli to limit validity of findings.

LABORATORY STUDY

This study was designed to demonstrate the effect of consequences of ratings and re-examine the relationship between appraisal purpose and

characteristics of performance ratings. Since this study was essentially concerned with theory-testing, it was important to control for threats to internal validity. Cook and Campbell (1979) and Calder, Phillips and Tybout (1981) have suggested that in research designed for theory-testing, concerns about addressing threats to internal validity should take precedence over generalizabilty of results. Consequently, this study was conducted in a laboratory setting.

A major criticism of laboratory research is it's purpoted lack of generalizability. Despite such criticisms, the findings of research in the laboratory do not appear to differ dramatically from the findings of field research. For instance, many of the more robust laboratory findings have been successfully replicated in the field (Locke, 1986). In perhaps the most comprehensive review, Bernardin and Villanova (1986) compared research in three areas of performance appraisal (rating formats, rater training, and rating purpose) and found little compelling evidence of difference in results (see also Dobbins, Cardy, & Truxillo, 1988).

The continuing resistance to laboratory research may be attributed to faulty conceptions of external validity. External validity - the potential for generalizability is not something that can be achieved in any one study, but is an empirical question, and can only be inferred from replication across populations, settings, variables and time (Campbell & Stanley, 1967; Mook, 1983).

To facilitate generalizability, it is important to enhance similarity between the laboratory environment and targeted populations and settings, at least with regard to factors with potential to

influence the psychological mechanisms underlying the phenomena under investigation. Such an approach creates the potential to transcend particularistic attributes of specific settings and enhances our understanding of the basic processes involved. Thus, ecological validity should be an important concern of laboratory researchers.

In most studies, for instance, subjects are only presented with information that is immediately relevant to the experimental task, while raters in real organizations are confronted with both relevant and irrelevant information. In studies on performance appraisal that essentially involve decision-making, such differences may potentially alter the decision making process. This viewpoint has been aptly pointed out by Murphy and his colleagues (Banks & Murphy, 1985; Murphy, Herr, Lockhart & Maguire, 1986) who have rightly observed that laboratory studies on appraisal may be unrealisitic insofar as subjects do not have to seperate the signal in the form of valid information about performance from the noise as do appraisers in field settings.

This study included both signal in the form of valid performance information and noise in the form of information extraneous to the appraisal task such as appraisal purpose and availability of funds. To make the study more realistic the noise was embedded within the signal. Thus, like real raters, in real organizations, subjects were placed in a situation which required them to sort through the material, pick out and integrate relevant information, form judgments about performance, transform these judgments into appraisal ratings, and make a related personnel decision.

Another important concern of studies conducted in the laboratory

is the lack of consequences of ratings for the rater and ratee (Ilgen & Favero, 1985). Although, no real consequences were made available, subjects' perception of consequences were varied across treatment combinations of purpose and funds. Even such perception of consequences (as opposed to real consequences) played a significant role in shaping the results of the study. Given this finding, laboratory research that incorporates real consequences for raters and ratees will be more informative. Thus, an attempt was made to enhance the ecological validity of this study.

SAMPLE

Another concern related to laboratory studies is the use of college students as subjects. However, it should be realized that phenomena observed in homogeneously defined groups of subjects - be they workers in the "real world" or college students in a laboratory - may offer equal, limited potential for generalizability. To the degree that any sample, whether it is composed of college students or organizational employees represent a homogeneous group that does not add extraneous variance to the behavior in question, its use may be considered a strength, and not a weakness (Berkowitz & Donnerstein, 1982).

This study addressed a judgment formation process (evaluating performance) well within the experience and capability of college students. Considering that this research focused on very general processess that one would not expect to vary radically with subject populations, college students do not appear entirely inappropriate. Several other researchers have also pointed out that there is usually little or no evidence to assume that such processes are dependent upon

the sample used (Locke, 1986). Dipboye and Flanagan (1979) and Greenberg (1987) each have contended that many of the processes of interest in organizational research yield more similar than dissimilar results between student and employee samples.

As Campbell (1986) noted, "perhaps college students really are people...why their disguise fools many observers into thinking otherwise is not clear" (p. 276). The typical laboratory subject, the college undergraduate, is not quite the barrier to generalizability as previously believed (see Dipboye, 1990).

PAPER-PEOPLE

In this study, performance information was presented in the form of discrete incidents. These incidents captured various levels of performance and are more realistic than vignettes that summarize performance of ratees. These designs labelled paper-people designs are regarded as inferior to those using live or videotaped stimuli by Murphy and his colleagues (Murphy et al, 1986). Murphy et al (1986) conducted a meta-analysis and reported larger effect sizes for paper-people than live or taped subjects. However, a more recent study that directly compared written vignettes and actual raters found no such differences (Dobbins, Cardy, & Truxillo, 1988). Currently, the status of this issue is not clear.

Subjects, in this study, perceived varying levels of consequences across treatment combinations and as expected high self-monitors provided evaluations in anticipation of those consequences more than low self-monitors. Inspite of the fact that no real consequences were made available to the subjects, anticipated consequences had the predicted

effect. Clearly, considering the interdependent nature of the rater-ratee relationship, consequences of ratings are likely to be very high (and real) in the field (see Ilgen & Favero, 1985). Indeed, both surveys and field studies have amply documented the political nature of performance appraisal (Bernardin & Villanova, 1986; Longnecker et al, 1987). Given this, effect sizes reported in this study may be regarded as an underestimate of what may be observed with real people and real consequences (as opposed to paper-people).

IMPLICATIONS FOR PRACTICE

Both researchers and practitioners are concerned with leniency and accuracy of appraisal ratings. Readers familiar with the history of appraisal in military or civil service domains are well aware of the pervasivess of leniency in ratings that often renders an entire appraisal system worthless. Hyde (1982) discussed the "vast quantity of inflated reports filled with superlatives" (p. 296). A recent report from the U.S. Merit Systems Protection Board reflects similar conclusions from a larger scale study of attitudes toward the merit system within the federal government (U.S. Merit Systems Protection Board, 1989). Leniency can cause major problems when personnel decisions (e.g. promotions, pay-raises) are based upon comparisons of each worker's performance to some established standard. The imposition of a forced distribution rating system by Congress highlights the pervasiveness and adverse consequences of lenient ratings.

Results of this study suggests that some raters, particularly high self-monitors, may tend to give lenient and inaccurate ratings as well as distort important personnel decisions when they perceive significant

consequences for their ratees. Furthermore, from a broader perspective, the significant role played by consequences suggests that performance ratings and personnel decisions of high self-monitors are likely to be influenced by consequences of those decisions (for themselves as well as the employee(s) involved), as opposed to valid information relevant to the decision at hand. Such biased decision making suggests that raters' motivation to provide accurate decisions may be more important than generally believed. Two approaches to overcome such biased decision making. First, I recommend training to enhance coping-efficacy in addition to that provided to improve rater's ability to make accurate decisions. Alternatively, the impact of consequences and hence motivation to bias decisions may be reduced by encouraging raters to assist ratees through organizationally legitimate activities such as coaching or providing learning opportunities to increase competence, or the like. In other words, practitioners dissuade raters from engaging in organizationally illegitimate behaviors such as distorting important personnel decisions by encouraging raters to help ratees through organizationally legitimate behaviors. Practitioners should also clearly communicate appraisal purpose to avoid unnecessary variance in evaluations provided by different raters who may have different perceptions of the intended use of performance ratings. Additionally, practitioners should educate raters about the legal and psychological consequences of considering nonperformance information while evaluating employee performance.

BIBLIOGRAPHY

Aleamoni, L.M., & Hexner, P.Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. Instructional Science, 9, 67 - 84.

Balzer, W.K. (1986). Biases in the recording of performance-related information : The effects of initial impression and centrality of the appraisal task. Organizational Behavior and Human Decision Processes, 37, 329 - 347.

Banks, C.G., & Murphy, K.R. (1985). Toward narrowing the research - practice gap in performance appraisal. Personnel Psychology, 38, 335 - 345.

Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep. American Psychologist, 37, 245 - 257.

Berkshire, J.R., & Highland, R.W. (1953). Forced - choice performance rating -A methodological study. Personnel Psychology, 6, 355 - 378.

Bernardin, H.J. (1977). Behavioral expectation scales versus summated scales : A fairer comparison. Journal of Applied Psychology, 62, 422 - 427.

Bernardin, H.J., Abbott, J., & Cooper, D. (1985). The effects of appraisal purpose and rater training on rating characteristics. Paper presented at the Annual Academy of Management Meetings, San Diego.

Bernardin, H.J., Alvares, K.M., & Cranny, C.J. (1976). A recomparison of behavioral expectation scales to summated scales. Journal of Applied Psychology, 61, 564 - 570.

Bernardin, H.J., & Beatty, R.W. (1984). Performance appraisal : Assessing human behavior at work. Boston : Kent.

Bernardin, H.J., & Cooke, D.K. (1992). Effects of appraisal purpose on discriminability and accuracy of ratings. Psychological Reports, 70, 1211 - 1215.

Bernardin, H.J., & Kane, J.S. (in press). Performance appraisal : A contingency approach to system development and evaluation .

Boston, MA : PWS - Kent.

Bernardin, J.H., & Orban, J. A. (1990). Leniency effect as a function of rating format, purpose for appraisal, and rater individual differences. Journal of Business and Psychology, 5, 2, 197-211.

Bernardin, H.J., Orban, J.A., & Carlyle, J.J. (1981). Performance rating as a function of trust in appraisal and rater individual differences. Proceedings of the 41st Annual Academy of Management Meetings.

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

Bernardin, H.J., & Villanova, P. (1986). Performance appraisal. In E.A. Locke (Ed.), Generalizing from laboratory to field settings (pp. 43 - 62). Lexington, MA : Lexington.

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

Berscheid, E., Graziano, W.G., Monson, T., & Dermer, M. (1976). Outcome dependency: Attention, attribution and attraction. Journal of Personality and Social Psychology, 34, 978 - 989.

Blanz, F., & Ghiselli, E.E. (1972). The mixed standard scale : A new rating system. Personnel Psychology, 25, 185 - 199.

Borman, W.C. (1987). Personal constructs, performance schemata, and 'folk theories' of subordinate effectiveness : Explorations in an army officer sample. Organizational Behavior and Human Decision Processes, 40, 307 - 322.

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

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

Borman, W.C., & Hallam, G.L. (1991). Observation accuracy for assessors of work-sample performance : Consistency across task and individual differences correlates. Journal of Applied Psychology, 76, 11 - 18.

Borrensen, H.A. (1967). The effects of instruction and item content on three types of ratings. Educational and Psychological Measurement, 27, 855 - 862.

Bowman, G.W., Worthy, N.B., & Greyson, S.A. (1965). Problems in review: Are women executives people? Harvard Business Review, 43 (4), 52 - 67.

Bretz, Jr., Milkovich, G.T., & Read, W. (1992). The current state of performance appraisal research and practice : Concerns, directions and implications. Journal of Management, 18, 2, 321 - 352.

Cafferty, T.P., DeNisi, A.S., & Williams, K.J. (1986). Search and retrieval patterns for performance information : Effects on evaluations of multiple targets. Journal of Personality and Social Psychology, 50, 676 - 683.

Calder, B.J., Phillips, L.W., & Tybout, A.M. (1981). Designing research for application. Journal of Consumer Research, 8, 197 - 207.

Caldwell, D.F., & O'Reilly, C.A. (1982). Boundary spanning and individual performance: The impact of self-monitoring. Journal of Applied Psychology, 67, 124 - 127.

Campbell, J.P. (1986). Labs, fields, and straw issues. In E.A. Locke (ed.), Generalizing from laboratory to field settings (pp. 269-297), Lexington, Mass.: D.C. Heath.

Campbell, D.T., & Stanley, J.C. (1967). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology, vol. 12. New York : Academic Press.

Cantor, N., & Mischel, W. (1977). Traits as prototypes : Effects on recognition memory. Journa of Personality and Social Psychology, 35, 38 - 48.

Centra, J.E. (1976). The influence of different directions on student ratings of instruction. Journal of Educational Measurement, 13, 277-282.

Cleveland, J.N., Murphy, K.R., & Williams, R.E. (1989). Multiple uses of performance appraisal : Prevalence and correltes. Journal of Applied Psychology, 74, 130 - 135.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation. Chicago: Rand- McNally.

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

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

DeCotiis, T., & 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., & 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., Gardner, T., & Cafferty, T.P. (1987). Diary keeping and the organization of information for performance appraisals. Proceedings of the Annual Academy of Management Meetings.

DeNisi, A.S., & Summers, T. (1986). Rating forms and the organization of information : A role for appraisal instruments. Paper presented at the Annual Academy of Management Meetings, Chicago.

DeNisi, A.S., & Williams, K.J. (1988). Cognitive approaches to performance appraisal. In K.M. Rowland and G.R. Ferris (Eds.), Research in personnel and human resource management (vol. 6, pp. 109 - 155). Greenwich CT : JAI Press.

Dipboye, R.L. (1990). Laboratory vs. field research in industrial and organizational psychology. International Review of Industrial and Organizational Psychology, 5, 1 - 34.

Dipboye, R.L. (1985). Some neglected variables in research on discrimination in appraisals. Academy of Management Review, 10, 116 - 127.

Dipboye, R.L., & Flanagan, M.F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than those in the laboratory ? American Psychologist, 34, 141 - 150.

Dobbins, G.H., Cardy, R.L., & Platz-Vieno, S.J. (1990). A contingency approach to appraisal satisfaction: An initial investigation of the joint effects of organizational variables and appraisal characteristics. Journal of Management, 16, 619 - 632.

Dobbins, G.H., Cardy, R.L., & Truxillo, D.M. (1988). The effects of purpose of appraisal and individual differences in stereotypes of women on sex differences in performance ratings: A laboratory and field study. Journal of Applied Psychology, 73, 551 - 558.

Driscoll, L.A., & Goodwin, W.L. (1979). The effects of varying information about the use and disposition of results on university students' evaluations of faculty and courses. American Educational Research Journal, 16, 25-37.

Elliott, G.C. (1979). Some effects of deception and level of self-monitoring on planning and reacting to a self-presentation.

Journal of Personality and Social Psychology, 37, 1282 - 1292.

Favero, J.L., & Ilgen, D.R. (1983). The effects of ratee characteristics on rater performance appraisal behavior(Tech. Rep. 83-S). East Lansing, MI: Michigan State University. Departments of Psychology and Management.

Feldman, J.M. (1986). Instrumentation and training for performance appraisal : A perceptual - cognitive view point. In K.M. Rowland and G.R. Ferris (Eds.), Research in personnel and human resource management (vol. 4, pp. 45-99). Greenwich, CT : JAI Press.

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

Fisher, C.D. (1989). Current and recurrent challenges in HRM. Journal of Management, 15, 2, 157 - 180.

Funder, B.C. (1987). Errors and mistakes : Evaluating the accuracy of social judgment. Psychological Bulletin, 101, 75 - 90.

Gangestad, S., & Snyder, M. (1985). To carve nature at its joints : On the existence of discrete classes in personality. Psychological Review, 92, 317 - 349.

Geizer, R.S., Rarick, D.L., & Soldow, G.F. (1977). Deception and judgment accuracy: A study in person perception. Personality and Social Psychology Bulletin, 3, 446 - 449.

Giles, W.F., & Mossholder, K.W. (1990). Employee reactions to contextual and session components of performance appraisal. Journal of Applied Psychology, 75, 371 - 377.

Gmelch, W.H., & Glasman, N.S. (1977). The effects of purpose on student evaluation of college instructors. Educational Research Quarterly, 2, 45 - 55.

Goffman, E. (1956). The presentation of self in every day life. Edinburgh, Scottland. University of Edinburgh Press..

Greenberg, J. (1987). The college sophomore as guinea pig: Setting the record straight. Academy of Management Review, 12, 157 - 159.

Hedge, J.W., & Kavanagh, M.J. (1988). Improving the accuracy of performance evaluations : Comparison of 3 methods of performance appraiser training. Journal of Applied Psychology, 73, 68 - 73.

Heneman, R.L., & 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.

Heneman, R.L., Wexley, K.N., & Moore, M.L. (1987). Performance-rating accuracy. A critical review. Journal of Business Research, 15, 431 - 448.

Hogan, E.A. (1987). Effects of prior expectations on performance ratings. A longitudinal study. Academy of Management Journal, 30, 354 - 368.

Hollander, E.P. (1965). Validity of peer nominations in predicting a distant performance criterion. Journal of Applied Psychology, 49, 434 - 438.

Hollander, E.P. (1957). The reliability of peer nominations under various conditions of administration Journal of Applied Psychology, 41, 85 - 90.

Ilgen, D.R., & Favero, J.L. (1985). Limits in generalization from psychological research to performance appraisal processes. Academy of Management Review, 10, 311 - 321.

Ilgen, D.R., & Feldman, J.M. (1983). Performance appraisal : A process focus. In L.L. Cummings and B.M. Staw (Eds.), Research in organizational behavior (vol. 5, pp. 141 - 197). San Francisco : JAI Press.

Ivancevich, J.M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. Journal of Applied Psychology, 64, 502 - 508.

James, W. (1890). Principles of psychology. New York: Holt, Rinehart & Winston.

Jawahar, I.M., & Stone, T.H. (1992). A model of the training process. Published in the proceedings at the Annual Conference of the Administrative Sciences Association of Canada, Quebec City, Canada.

Jones, E.E., & Davis, K.E. (1965). From acts to dispositions. In L. Berkowitz (Ed.), Advances in experimental social psychology (vol. 2). New York : Academic Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory : An analysis of decision under risk. Econometrika, 47, 263 - 291.

Kane, J. (1980). Alternative approcahes to the control of systematic error in performance appraisals. Paper presented at the 1st Annual Scientist-Practitioner Conference in Industrial/Organizational Psychology, Old Dominion University.

Kirkpatrick, J.J., Ewen, R.B., Barrett, R.S., & Katzell, R.A. (1968). Testing and fair employment. New York, New York University Press.

Kozlowski, S.W., & Kirsch, M.P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. Journal of Applied Psychology, 72, 252 - 261.

Landy, F.J., Farr, J.S. (1983). The measurement of work performance - methods, theory, and applications. Orlando, FL: Avcadmeic Press.

Landy, F.J., & Farr, J.S. (1980). Performance rating. Psychological Bulletin,87, 1, 72 - 107.

Latham, G.P., Irvine, D., Skarlicki, D., & Siegel, J.P. (in press). The increasing importance of performance appraisals to employee effectiveness in organizational settings in North America. In C.L. Cooper and I. Robertson (Eds.), International Review of Industrial and organizational Psychology.

Latham, G.P., & Wexley, K.N. (1977). Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 30, 255 - 268.

Lawler, T.G. (1988). The objectives of performance appraisal - or ' Where can we go from here ?' Nursing Management, 19, 82-88.

Lewin, K. (1938). The conceptual representation and the measurement of psychological forces. Durham, N.C : Duke University Press.

Locher, A.H., & Teel, K.S. (1988). Appraisal trends. Personnel Journal, September, 139 - 145.

Locke, E.A. (ed.) (1986). Generalizing from laboratory to field settings. Lexington, Mass.: D.C. Heath.

Longenecker, C.O., Sims, H.P., & Gioia, D.A. (1987). Behind the mask : The politics of employee appraisal. The Academy of Management Executive, 1, 183 - 193.

Longenecker, C.O., & Gioia, D.A. (1988). Neglected at the top - executives talk about executive appraisals. Sloan Management Review, 21 (Winter), 41 - 47.

Lord, R.G., & Maher, K.J. (1989). Cognitive processes in industrial and organizational psychology. In C.L. Cooper and I. Robertson (Eds.), International Review of Industrial and Organizational Psychology (pp. 49-91). New York : John Wiley and Sons.

Lord, R.G. (1985). Accuracy in behavioral measurement : An alternative definition based on rater's cognitive schema and signal detection theory. Journal of Applied Psychology, 70, 66 - 71.

Lord, R.G., Foti, R.J., & DeVader, C.L. (1984). A test of leadership categorization theory : Internal structure, information processing, and leadership perceptions. Organizational Behavior

and Human Performance, 34, 343 - 378.

Lord, R.G., Foti, R.J., & Phillips, J.S. (1982). A theory of leadership categorization. In J.G. Hunt, V. Sekaran, & C. Schriesheim (Eds.), Beyond Establishment Views (pp. 104 - 121). Carbondale : Southern Illinois University Press.

Maier, N.F., & Thurber, J.A. (1968). Accuracy of judgments of deception when an interview is watched, heard, and read. Personnel Psychology, 21, 23 - 30.

McIntyre, R.M., Smith, D.E., & Hassett, C.E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 149 - 156.

Mead, G.H. (1934). Mind, self, and society. Chicago. University of Chicago Press.

Meier, R.A., Feldhusen, J.F. (1979). Another look at Dr. Fox: effect of stated purpose of evaluation, lecturer expressiveness, and density of lecture content on student ratings. Journal of Educational Psychology, 71, 339 - 345.

Meyer, H.H., Kay, E., & French, J. (1965). Split roles in performance appraisal. Harvard Business Review, 43, 123 - 129.

Miller, G.A. (1956). The magical number seven, plus or minus two : Some limits on our capacity for processing information. Psychological Review, 63, 81 - 97.

Mohrman, A.M., & Lawler, E.E. (1983). Motivation and performance appraisal behavior. In F.J. Landy, S. Zedeck., & J. Cleveland (Eds.), Performance measurement and theory (pp. 173 - 189). Erlbaum, Hillsdale, N.J.

Mook, D.G. (1983). In defense of external invalidity. American Psychologist, 38, 379 - 387.

Murphy, K.R., & Balzer, W.K. (1989). Systematic distortions in memory-based behavior ratings and performance evaluations : Consequences for rating accuracy. Journal of Applied Psychology, 71, 39 - 44.

Murphy, K.R., & Balzer, W.K. (1986). Systematic distortions in memory-based ratings: Consequences of rating accuracy. Journal of Applied Psychology, 71, 39-44.

Murphy, K.R., Balzer, W.K., Kellam, K.L., & Armstrong, J.G. (1984). Effects of the purpose of rating on accuracy in observing teacher behavior and evaluating teaching performance. Journal of Educational Psychology, 76, 45 - 54.

Murphy, K.R., & Constans, J.I. (1987). Behavioral anchors as a source of

bias in rating. Journal of Applied Psychology, 72, 573 - 577.

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

Murphy, K.R., Gannett, B.A., Herr, B.M., & Chen, J.A. (1986). Effects of subsequent performance on evaluations of previous performance. Journal of Applied Psychology, 71, 427-431.

Murphy, K.R., Herr, B.M., Lockhart, M.C., & Maguire, E. (1986). Evaluating the performance of paper people. Journal of Applied Psychology, 71, 654 - 661.

Murphy, K.R., Martin, C., & Garcia, M. (1982). Do behavior observation scales measure observation? Journal of Applied Psychology, 67, 562 - 567.

Napier, N.K., & Latham, G.P. (1986). Outcome expectancies of people who conduct performance appraisals. Personnel Psychology, 39, 827 - 837.

Nathan, B.R., & Alexander, R.A. (1985). The role of inferential accuracy in performance rating. Academy of Management Review, 10, 109-115.

Nathan, B.R., & Lord, R.G. (1983). Cognitive categorization and dimensional schemata : A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68, 102 - 114.

Nathan, B.R., & Tippins, N. (1990). The consequences of halo "error" in performance ratings : A field study of the moderating effect of halo on test validation results. JOurnal of Applied Psychology, 75, 290 - 296.

Paterson, D.G. (1922). The Scott company graphic rating scale. Journal of Personnel Research, 1, 361 - 376.

Pulakos, E.D. (1986). The development of training programs to increase accuracy with different rating tasks. Organizational Behavior and Human Decision Processes, 38, 76 - 91.

Quinn, R.P., Tabor, J.M., Gordon, L.K. (1968). The decision to discriminate: A study of executive selection. Ann Arbor, MI: Institute of Survey Research.

Rhodewalt, F., & Comer, R. (1981). The role of self-attribution differences in the utilization of social comparison information. Journal of Research in Personality, 15, 210 - 220.

Rush, M.C., Phillips, J.S., & Lord, R.G. (1981). Effects of temporal delay in rating of leader behavior descriptions: A laboratory

investigation. Journal of Applied Psychology,66, 442 - 450.

Schmitt, N., Noe, R.A., & Gottschalk, R. (1986). Using the lens model to magnify rater's consistency, matching, and shared bias. Academy of Management Journal, 29, 130 - 139.

Schwab, D.P., Heneman, H.G., III., & DeCotiis, T (1975). Behaviorally anchored rating scales : A review of the literature. Personnel Psychology, 28, 549 - 562.

Sharon, A. (1970). Eliminating bias from student ratings of college instructors. Journal of Applied Psychology, 54, 278-281.

Sharon, A.T., & Bartlett, C.J. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 22, 251 - 263.

Smith, D.E. (1986). Training programs for performance appraisal. A review. Academy of Management Review, 11, 22 - 40.

Smith, P.C. (1976). Behaviors, results, and organizational effectiveness : The problem of criteria. In M.D. Dunnette (Ed.), Handbook if industrial and organizational psychology. Chicago : Rand McNally.

Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations : An approach to the construction of the unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149 - 155.

Smith, D.E., Hassett, C.E., & McIntyre, R.M. (1982). Using student ratings for administrative decisions : Are ratings contaminated by perceived uses of the information. Paper presented at the 23rd Annual Meeting of the Western Academy of Management. Colarado Springs, Co.

Smither, J.W., & Reilly, R.R. (1987). True intercorrelation among job components, time delay in rating, and rater intelligence as determinants of accuracy in performance ratings. Organizational Behavior and Human Decision Processes, 40, 369 - 391.

Snyder, M. (1987). Public appearnaces private realities : The psychology of self-monitoring. New York : W.H. Freeman & Co.

Snyder, M. (1979). Self-monitoring process. In L. Berkowitz (Ed.), Advances in experimental social psychology(vol. 12.). New York : Academic Press.

Snyder, M. (1974). Self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30, 526 - 537.

Snyder, M., Berscheid, E., & Glick, P. (1985). Focusing on the exterior and the interior: Two investigations of the initiation of personal relationships. Journal of Personality and Social Psychology, 48,

1427-1439.

Snyder, M., Berscheid, E., & Matwychuck, A. (1988). Orientations toward personnel selection: Differential reliance on appearance and personality. Journal of Personality and Social Psychology, 54, 972-979.

Snyder, M., Berscheid, E., & Matwychuk, A. (1985). Unpublished paper., University of Minnesota.

Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring : Matters of assessment, matters of validity. Journal of Personality and Social Psychology, 51, 1, 125 - 139.

Snyder, M., & Simpson, J.A. (1984). Self-monitoring and dating relationships. Journal of Personality and Social Psychology, 47, 1281 - 1291.

Spool, M.D. (1978). Training programs for observers of behaviors : A review. Personnel Psychology, 31, 853 - 888.

Srull, T.K., & Wyer, R.S. (1989). Person memory and judgment. Psychological Review, 96, 1, 58 - 83.

Srull, T.K., & Wyer, R.S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgment. Journal of Personality and Social Psychology, 38, 841 - 856.

Srull, T.K., & Wyer, R.S. (1979). The role of category accessibility in the interpretation of information about persons : Some determinants and implications. Journal of Personality and Social Psychology, 37, 1660 - 1672.

Taylor, E.K., & Wherry, R.J. (1951). A study of leniency in two rating systems. Personnel Psychology, 4, 39 - 47.

Thornton, G.C., III, & Zorich, S. 91980). training to improve observer accuracy. Journal of Applied Psychology, 65, 351 - 354.

U.S. General Accounting Office, First Look at Senior Executive Service Performance Awards (Washington, D.C.: FPCD-80-74, August 15, 1980, p. 2).

Wherry, R.J. (1952). The control of bias in ratings : A theory of rating. Columbus : The Ohio State Research Foundation.

Wherry, R.J. & Bartlett, C.J. (1982). The control of bias in ratings : A theory of ratings. Personnel Psychology, 35, 521 - 552.

Williams, K.J., DeNisi, A.S., Blencoe, A.G., & Cafferty, T.P. (1985). The role of appraisal purpose: Effects of purpose on information

acquistion and utilization. Organizational Behavior and Human Decision processes, 35, 314 - 339.

Williams, K.J., DeNisi, A.S., Meglino, B.M., & Cafferty, T.P. (1986). Initial decisions and subsequent performance ratings. Journal of Applied Psychology, 71, 189 - 195.

Winter, L., & Uleman, J.S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. Journal of Personality and Social Psychology, 47, 237 - 252.

Zajonc, R.B. (1980). Feeling and thinking: preferences need no inferences. American Psychologist, 35, 2, 151-175.

Zedeck, S., & 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.

APPENDIXES

APPENDIX A

SELF-MONITORING SCALE

Dear Participant,

The following questionnaire is administered for research purposes only. Your responses will be guarded with strict confidentiality. Please answer the questions as **accurately** as you can. Your cooperation is essential for the success of this research project and will be gratefully acknowledged.

Sincerely,

Jim Jawahar.

Name (please print) _____

Social security # _____ Classification _____

Major area _____ Gender _____

Please answer the following items as accurately as possible by placing a T - for True, or an F - for False.

____ 1. I find it hard to imitate the behavior of other people

____ 2. At parties and social gatherings, I do not attempt to do or say things that others will like

____ 3. I can only argue for ideas which I already believe

____ 4. I can make impromptu speeches even on topics about which I have almost no information

____ 5. I guess I put on a show to impress or entertain others

____ 6. I would probably make a good actor

____ 7. In a group of people I am rarely the center of attention

____ 8. In different situations and with different people, I often act like very different persons

____ 9. I am not particularly good at making other people like me

___ 10. I'm not always the person I appear to be

___ 11. I would not change my opinions (or the way I do things) in order to please someone or win their favor

___ 12. I have considered being an entertainer

___ 13. I have never been good at games like charades or improvisational acting

___ 14. I have trouble changing my behavior to suit different people and different situations

___ 15. At a party I let others keep the jokes and stories going

___ 16. I feel a bit awkward in public and do not show up quite as well as I should

___ 17. I can look anyone in the eye and tell a lie with a straight face (if for a right end)

___ 18. I may deceive people by being friendly when I really dislike them

APPENDIX B

EXPERIMENTAL MATERIAL

Dear Participant

One of the most important jobs of management is evaluating the performance of employees. This study tests the effectiveness of a new method of performance evaluation. In this study, your role will be that of a supervisor who evaluates the performance of subordinates.

The new method of performance appraisal is called the journal entry method. This approach requires managers to keep a log on each of their employees. These logs describe behaviors the managers have noted about each employee over the course of 6 months. At appraisal time, managers review the logs and make the necessary personnel decisions. The purpose of this study is to determine the usefulness of the journal entry method for evaluating employee performance and making related personnel decisions.

Information about your role as a supervisor is presented on the next page. The performance information of your two subordinates, Pat and Chris follows the background scenario. Please read the scenario and performance information carefully. After you read and understand the material please evaluate Pat's performance. After rating Pat on the enclosed performance appraisal form please proceed to the final questionnaire. This should take about 20 minutes to complete.

Thank you for your time and attention. Your participation is essential to the success of this project and is greatly appreciated.

Sincerely,

Jim Jawahar

SCENARIO BACKGROUND

You work for a mail-order company that carries a wide range of outdoor products, everything from camping to sports equipment. Catalog sales are conducted both over the telephone and through the mail. Most of the contact with the customers is over the phone but there are also some walk-in contacts. A representative handles, on average, 20-30 calls per day. Customer questions usually pertain to product characteristics, warranties, delivery times, etc. The representatives are responsible for dealing with any problems or complaints a customer is having with the merchandise they ordered. A log of each call is kept by the representatives detailing the nature of the call, the caller's name (if available), as well as the information provided to the caller. The representatives must summarize their call logs and give these summaries to you each month.

If representatives are unable to promptly answer a customer's questions they are instructed to call the customer back with the proper information. Politeness, friendliness, and accuracy are stressed in all representative interactions with customers. Representatives are never to respond to the rudeness of a customer with anything but tact and a calm response. If a representative is unable to solve a customer's problem or is unsure of how to solve the problem then he or she is to transfer the call to you or consult with you before returning the customer's call. Representatives are to brief you about the nature of a call before it is transferred to you. Sometimes it may be necessary for the representatives to call product manufacturers for information needed to answer a caller's question.

The representatives are encouraged to make suggestions which will improve customer service and satisfaction. Also, representatives may be sometimes asked to do special projects. The representatives work from 8 am to 5 pm, Monday through Friday. Overtime, such as working evenings or weekends, is sometimes available.

YOUR TASK

The journal entry method of performance appraisal requires supervisors to keep a journal/log of each employee's behavior. Pat's and Chris's logs are provided on the next two pages. These logs contain a random sample of work events recorded over a 6 month period. These work events show their typical job performance.

You are asked to assume the role of a supervisor. Although you have two subordinates, Pat and Chris, only Pat needs to be evaluated now. When evaluating Pat's performance refer to the logs as often as you like; they are there to aid your decisions. However, before rating Pat please compare the performance information (log) of Pat with that of Chris.

Please note that this company uses performance appraisal ratings for merit-raise purpose only (i.e. pay increases). Therefore, after rating Pat's performance, please make a decision regarding merit-increase for Pat.

YOUR TASK

The journal entry method of performance appraisal requires supervisors to keep a journal/log of each employee's behavior. Pat's and Chris's logs are provided on the next two pages. These logs contain a random sample of work events recorded over a 6 month period. These work events show their typical job performance.

You are asked to assume the role of a supervisor. Although you have two subordinates, Pat and Chris, only Pat needs to be evaluated now. When evaluating Pat's performance refer to the logs as often as you like; they are there to aid your decisions. However, before rating Pat please compare the performance information (log) of Pat with that of Chris.

Please note that this company uses performance appraisal ratings for training purposes only. Training is provided to improve job knowledge, skills or abilities. The duration of training typically varies from 1 to 3 days. When the employee is attending the training program the company provides regular wages/salary and a temporary worker is assigned to replace the trainee during the employee's absence. After rating Pat's performance, please make a decision regarding training for Pat.

Chris's Log

* Turned in call log summary report on time.

* Told me that people do not appear to be very interested in our new line of White River hiking boots

* Noticed that somebody had been placing bogus orders in the Lawrence, Kansas area

* Was late for work this morning

* Turned in call log summary report on time

* Referred an irate customer to me

* Contacted UPS about the many reports of late shipments in the Dekalb, Illinois area

* Corrected mistakes on a customer's bill

* Reported that some of the orders made on the 24th had been sent out in duplicate

* Turned in call log summary report on time

* Could not keep an irate customer from cancelling his order

* Completed the assigned project on time

* Was unable to track down a missing order that had been sent to the wrong Sally Jones in Tulsa

* Turned in call log summary report late

* Was confused about how to figure out shipping costs as they are described in the latest catalog

* Lost temper when dealing with an upset customer

* Told me we were running low on office supplies

* Turned in call log summary report on time

* Contacted the wrong manufacturer. Could not figure out the manufacturer to be contacted

* Made a recommendation about adding Spencer fishing poles because of numerous customer suggestions

* Was unable to track down a customer's late order

* Misinterpreted a customer's requests

* Was late for work this morning

* Turned in call log summary report on time

* Called Brunswick for some warranty information

Pat's Log

* Noted that a lot of incorrect orders were originating from the third shift at the Grand Island warehouse

* Turned in call log summary report late

* Called in sick

* Made several mistakes on a customer's bill

* Turned in call log summary report late

* Lost temper with a customer who was upset about a late order

* Reported problems with jumbled customer orders issued by the computer

* Called in sick

* Turned in call log summary report on time

* Called Garcia about some product information

* Forgot to bring to my attention a moderately serious problem

* Failed to notice that somebody has been placing phony orders in the Des Moines, Iowa area

* Turned in call log summary on time

* Reported that customers were having problems getting Dobson to honor warranty repairs on their tobogans

* Called in sick

* Was noticed contacting Garcia when the subordinate should have been contacting Sewrite

* Turned in call log summary report late

* Noticed how messy the employee's work area was

* Told me that people seemed to be very interested in our new line of Shakespeare trolling motors

* Referred an irate customer to me

* Turned in call log summary report on time

* Referred a customer to me who demanded to speak to me. The customer was not happy with what the subordinate had told her about the lantern she had ordered

* Made a recommendation about discontinuing Prankston boat covers due to numerous quality problems

* Failed to keep an irate customer from cancelling her order

Last year was a normal year for the company. This year the company made unusually large profits and consequently funds in the pay-raise budget have been tripled. So this year there will be plenty of funds/money for pay increases. Funds in the pay-raise budget are expected to remain at the current level for at least another 2 to 3 years.

PERFORMANCE APPRAISAL FORM

Please use the following performance appraisal form to rate the performance of Pat. Performance is measured on 5 dimensions. The dimensions are interpersonal and communication skills, dependability, quality, knowledge and initiative. Please circle only one item per dimension.

Interpersonal and Communication Skills: This dimension assesses representative's interpersonal and communication skills.

Could be expected to respond to customers in a polite and friendly manner 6

Could be expected to effectively handle upset customers 5

Could be expected to enquire and understand the needs of customers and provide them with the relevant information..... 4

Could be expected to misunderstand customer's requests and provide him or her with irrelevant information..... 3

Could be expected to argue with customers in order to convince them about superiority of our products..... 2

Could be expected to lose temper when dealing with upset customers..... 1

Dependability: This dimension assesses dependability of representatives.

Could be expected to be at work, submit log summary and complete projects on time..... 4

Could be expected to turn in log summary report late..... 3

Could be expected to come in late for work 2

Could be expected to call in sick frequently..... 1

Last year was a normal year for the company. This year the company incurred heavy losses and consequently there are no funds in the pay-raise budget. So this year there will be no funds/money for pay increases. Funds in the pay-raise budget are not expected to increase dramatically for at least another 2 to 3 years.

PERFORMANCE APPRAISAL FORM

Please use the following performance appraisal form to rate the performance of Pat. Performance is measured on 5 dimensions. The dimensions are interpersonal and communication skills, dependability, quality, knowledge and initiative. Please circle only one item per dimension.

Interpersonal and Communication Skills: This dimension assesses representative's interpersonal and communication skills.

Could be expected to respond to customers in a polite and friendly manner 6

Could be expected to effectively handle upset customers 5

Could be expected to enquire and understand the needs of customers and provide them with the relevant information..... 4

Could be expected to misunderstand customer's requests and provide him or her with irrelevant information..... 3

Could be expected to argue with customers in order to convince them about superiority of our products..... 2

Could be expected to lose temper when dealing with upset customers..... 1

Dependability: This dimension assesses dependability of representatives.

Could be expected to be at work, submit log summary and complete projects on time..... 4

Could be expected to turn in log summary report late..... 3

Could be expected to come in late for work 2

Could be expected to call in sick frequently..... 1

Quality: This dimension assesses representative's quality of work, neatness and ability to enhance customer satisfaction.

Could be expected to detect and correct mistakes..... 4

Could be expected to be unorganized and maintain an unclean work area..... 3

Could be expected to fail to detect phony orders..... 2

Could be expected to make billing mistakes..... 1

Knowledge : This dimension assesses representative's awareness of procedures, products, etc.

Could be expected to be fully aware of products; manufacturers to be contacted and their delivery schedules..... 3

Could be expected to frequently ask representatives about product attributes, delivery schedules and particular manufacturers to be contacted..... 2

Could be expected to contact the wrong person/manufacturer for product information..... 1

Initiative : This dimension assesses representative's propensity to take initiative, assume responsibility and pay attention to detail.

Could be expected to anticipate problems and make suggestions to avoid them..... 4

Could be expected to report problems and make appropriate recommendations..... 3

Could be expected to report problems 2

Could be expected to fail to identify problems..... 1

Overall Evaluation

Pat's performance (circle one number only) is

1	2	3	4	5	6	7
Very Low			Medium			Very High

Please answer the following questions:

1. Please make a merit-raise decision for **Pat**.

Decision is whether to give a merit raise
Oppose means - merit raise should not be given
Support means - merit raise should be given

1	2	3	4	5	6	7	8	9	10
strongly oppose		somewhat oppose		neutral		somewhat support		strongly support	
(merit raise should not be given)						(merit raise should be given)			

2. On the following scale, please indicate if you would be willing to assist **Pat** by

1. very unlikely 2. unlikely 3. maybe 4. likely 5. very likely

a. giving overtime	1	2	3	4	5
b. arranging for a loan.....	1	2	3	4	5
c. reducing workload	1	2	3	4	5
d. assigning easy tasks	1	2	3	4	5
e. providing coaching.....	1	2	3	4	5

3. Your name (please print) _____

4. Social security # _____

5. Have you ever appraised/evaluated another employee?

0 (No) 1 (Yes). If Yes, how many times ? _____

6. To what degree do you feel that the performance appraisal form allowed you to accurately evaluate performance ?

1	2	3	4	5	6	7	8	9	10
very low degree				average				very high degree	

Please answer the following questions:

1. Please make a training decision for **Pat**.

Decision is whether to send subordinate for training
<u>Oppose</u> means - training not required
<u>Support</u> means - training required

1	2	3	4	5	6	7	8	9	10
strongly oppose		somewhat oppose		neutral		somewhat support		strongly support	
(training not required)								(training required)	

2. On the following scale, please indicate if you would be willing to assist Pat by

1. very unlikely 2. unlikely 3. maybe 4. likely 5. very likely

a. giving overtime	1	2	3	4	5
b. arranging for a loan	1	2	3	4	5
c. reducing workload	1	2	3	4	5
d. assigning easy tasks	1	2	3	4	5
e. providing coaching	1	2	3	4	5

3. Your name (**please print**) _____

4. Social security # _____

5. Have you ever appraised/evaluated another employee?

0 (No) 1 (Yes). If Yes, how many times ? _____

6. To what degree do you feel that the performance appraisal form allowed you to accurately evaluate performance ?

1	2	3	4	5	6	7	8	9	10
very low degree				average				very high degree	

7. What is your overall impression of the critical incident method for evaluating employee performance ?

_____

_____

8. The performance ratings that you just provided **will** significantly affect your subordinate, **Pat**.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

9. The performance ratings you provided **will not have** any consequences for **Pat**.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

10. Please use the space below to write down the consequences of your performance ratings for **Pat**.

_____

_____

_____

_____

11. To what extent did you consider consequences of performance ratings while evaluating **Pat's** performance:

1	2	3	4	5
Did not consider it at all		Somewhat considered it		Considered it a great deal

12. While recommending pay-raise, to what extent did you consider how your pay-raise decision would affect **Pat**:

1	2	3	4	5
Did not consider it at all		Somewhat considered it		Considered it a great deal

13. For which of the following purposes does this company use performance appraisal ratings:

_____ a. Merit-raise

_____ b. Training

_____ c. Documentation

_____ d. Promotion

14. Compared to last year, this year more funds are available for providing pay raises

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

15. This year due to lack of funds raises cannot be provided

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

16. Pat is a better performer than Chris.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

17. Chris is a better performer than Pat.

1	2	3	4	5	6	7	8	9
strongly disagree		disagree		neither disagree nor agree		agree		strongly agree

18. How would you rate your knowledge of appraisal methods ?

1	2	3	4	5
None	Limited	Average	Good	Excellent

19. What do you think is the real purpose of the study ?

_____

_____

Thank you very much for you time and attention. Your assistance is deeply appreciated.

VITA 2

JAWAHAR I. MOHAMMED

Candidate for the Degree of

Doctor of Philosophy

Thesis: PURPOSE OF APPRAISAL REVISITED: AN EXAMINATION OF THE RELATIONSHIP BETWEEN PURPOSE AND CHARACTERISTICS OF PERFORMANCE RATINGS

Major Field: Business Administration

Biographical:

Personal Data: Born in Madras, India, November 05, 1964, the son of Mr. and Mrs. Ibrahim.

Education: Graduated from Kesari Higher Secondary School, Madras, India, in 1982; received Bachelor of Science Degree in Physics from University of Madras, India, in 1985; received Master of Science Degree in Physics from University of Madras, India, in 1987; received Honors Post Graduate Diploma in Personnel Management and Industrial Relations from Madras School of Social Work, India, in 1988; received Master of Arts Degree in Industrial and Organizational Psychology from University of Tulsa, Oklahoma, in 1990; completed requirements for the Doctor of Philosophy Degree at Oklahoma State University in May, 1994.

Professional Experience: Graduate Teaching Associate, Department of Management, Oklahoma State University, August, 1989, to May, 1994. Personnel Officer, Greaves Chitram Limited, Madras, India, May, 1987, to December 1987. Consultant, January, 1986, to April, 1987.