

EFFECTS OF CONTEXT ON THE LENIENCY, ACCURACY, AND UTILITY
OF SELF-APPRAISALS OF PERFORMANCE:
SOCIAL COMPARISON INFORMATION AND PURPOSE OF APPRAISAL

by

Steven Craig Morgan

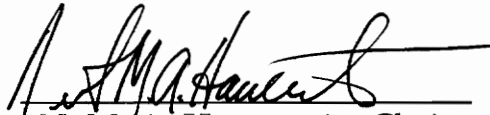
Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

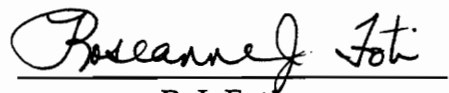
in

Psychology

APPROVED:


N. M. A. Hauenstein, Chair


S. B. Gustafson


R. J. Foti

October 17, 1994

Blacksburg, Virginia

c.2

LD
5655
V855
1994
M673
c.2

EFFECTS OF CONTEXT ON THE LENIENCY, ACCURACY, AND UTILITY
OF SELF-APPRAISALS OF PERFORMANCE:
SOCIAL COMPARISON INFORMATION AND PURPOSE OF APPRAISAL

by

Steven Craig Morgan

Committee Chairperson: Neil M. A. Hauenstein

Psychology

(ABSTRACT)

The current study examines the effects of social comparison information and purpose of appraisal on the leniency and accuracy of self-appraisals of performance. Previous research has shown that providing comparison information decreases the leniency and increases the accuracy of self-appraisals. While these effects of comparison information quantity are documented, no research to date has examined the effects of comparison information quality on self-appraisal. It was hypothesized that self-ratings would be less lenient and more accurate when comparison information was presented via a written medium as opposed to an observational medium. While the results clearly support the role of comparison information quantity, the role of quality was generally not supported. Moreover, there was a discrepancy between the free recall measure and self-ratings with regard to the role of quality of comparison information. Past research indicates that self-appraisals conducted for reward purposes are more lenient and less accurate than those conducted for feedback purposes. The current results provide further support for this trend. Possible explanations for the current results, implications for past research, and suggestions for future research are discussed.

ACKNOWLEDGMENTS

My sincerest gratitude is extended to Dr. Neil Hauenstein. Neil's experience and advice as well as his confidence in me have greatly facilitated my progress towards becoming a professional psychologist. Neil went beyond the obligatory roles of advisor and committee chair and became my mentor and friend. I will always be grateful for this.

I would also like to thank Dr. Roseanne J. Foti and Dr. Sigrid B. Gustafson for serving on my committee and providing valuable suggestions regarding this research as well as ideas for future endeavors.

Thanks also go to my friends. Wayne, Wanda, Paul, Paige, John, Judy, and Thomas helped provide much needed distractions from school.

I would also like to give thanks to my two families, the Ahrens' and the Morgans, for all their support over the years. Special thanks also go to my parents, Bill and Eileen Morgan, whose unwavering faith in me is responsible for my being at this point in my life today.

Last, but certainly not least, I thank Andrea Morgan, my wife, best friend, and biggest supporter. Words can't express my feelings for Andrea and what she means to me. She has and always will be there when I need her most. It is to Andrea that I dedicate this work.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS.....	iii
<u>chapter</u>	<u>page</u>
INTRODUCTION.....	1
Overview of Study.....	10
SELF-APPRAISAL: A LITERATURE REVIEW.....	12
Leniency.....	13
Accuracy.....	16
Utility.....	19
Quality.....	21
Social Comparison Information.....	21
Performance Appraisal.....	23
Purpose Of Appraisal.....	28
Hypotheses.....	32
METHOD.....	34
Subjects.....	34
Design.....	34
Experimental Task.....	34
Procedure.....	34
Independent Variables.....	35
Social Comparison Information.....	35
Purpose of Appraisal.....	37
Dependent Measures.....	38
Self-Ratings.....	38
Objective Performance Measures.....	38
Leniency.....	39
Accuracy.....	39
Free Recall of Social Comparison Information.....	40
Utilization of Social Comparison Information.....	41
Manipulation Check.....	41

<u>chapter</u>	<u>page</u>
RESULTS.....	42
Manipulation Checks.....	42
Leniency.....	44
Accuracy.....	51
Utilization of Social Comparison Information.....	53
DISCUSSION.....	56
REFERENCES.....	68
APPENDICES.....	76
A. PROOFREADING TEST SAMPLE.....	77
B. PERFORMANCE APPRAISAL FORMS.....	79
C. PILOT STUDY DATA.....	83
D. NORMATIVE DATA SHEETS.....	85
E. PROTOCOL: CONTROL GROUP.....	88
F. PROTOCOL: BEHAVIORAL OBSERVATION INFERENCE GROUP.....	90
G. PROTOCOL: STRUCTURED INFORMATION INFERENCE GROUP.....	92
H. CONSENT FORM.....	95
I. CONFIRMATION OF DEBRIEFING.....	97
J. OBJECTIVE SCORE DERIVATION.....	99
K. SUMMARY TABLES FOR MANIPULATION CHECKS.	101
L. SUMMARY TABLES FOR DEPENDENT VARIABLES...	104
VITA.....	126

INTRODUCTION

Performance appraisals may be conducted by using supervisor, peer, or self-ratings. Although each has its advantages and disadvantages (Landy & Farr, 1980), it has traditionally been assumed that supervisors provide the best source of ratings. However, many researchers (e.g., Cardy & Dobbins, 1994) believe that self-appraisals will become more prevalent in the future as organizations move towards the implementation of work teams (Sundstrom, Demeuse, & Futrell, 1990) and programs such as Quality of Work Life and Quality Control Circles that advocate worker autonomy and self-development. Moreover, other researchers (e.g., Lin, Fahr, Dobbins, Doyle, & Spaulding, 1992) have proposed that self-appraisal is often the only method of evaluation for jobs that are complex and/or require highly specialized knowledge, skills, and abilities. Unfortunately, as opposed to other rating sources, little research has been conducted that attempts to improve the quality of self-ratings. The present study will examine contextual factors that may affect the quality of self-ratings.

Self-Appraisal

Traditionally, self-appraisals have been criticized as inaccurate measures of job performance. Historically, the criticism surrounding the use of self-appraisals in formal appraisal contexts seems to arise from perceptions that: (1) self-appraisals are subject to self-enhancement biases such as leniency; and, (2) most people are not able to evaluate themselves objectively enough to provide an accurate self-appraisal (Boyle & Klimoski, 1993).

Leniency errors reflect an inappropriate judgmental standard on the part of the rater, whereby assigned ratings are unduly lenient or positive

(Bernardin & Beatty, 1984). Saal and Knight (1988) suggest that a rater may exaggerate the quality of performance because of (1) a desire to be liked; (2) a fear that other raters will exaggerate ratings, therefore giving them an unfair advantage when promotions or pay raises become available; and (3) abnormally low standards of excellence. Moreover, evidence suggests that self-ratings are more lenient than either supervisor or peer ratings (Klimoski & London, 1974).

Leniency has serious implications for the performance appraisal process. Because lenient ratings are, by definition, higher than they should be, any generated performance feedback based on those ratings will be inaccurate (Saal & Knight, 1988). Leniency also presents a problem to the extent that a restriction of range restricts the potential relationship between ratings and related criteria (Hoffman, Nathan, & Holden, 1991; Holzbach, 1978; Thornton, 1980; Shrauger & Osberg, 1981). Given the inherent measurement problems associated with self-ratings, it is not surprising that researchers are reluctant to encourage their use. Indeed, Campbell and Lee (1988) conclude that the traditional self-appraisal has little utility as an evaluation tool.

Although the self-appraisal, with its assumed inherent biases and psychometric limitations is perceived as having little utility, it is inappropriate to generalize this perception to all self-appraisals. A self-appraisal that minimizes measurement problems and biases has potential utility for an organization. Perhaps one of the greatest assets of self-appraisals to an organization is the increased likelihood of agreement among workers regarding organizational fairness and justice (Bies & Shapiro, 1988; Folger &

Greenberg, 1985; Greenberg, 1987; Sheppard, 1985). Self-appraisals also have the potential to facilitate employee development within an organization (Campbell & Lee, 1988); increase ratee involvement in the appraisal process (Bernardin & Beatty, 1984); improve the acceptance of decisions based on appraisal information (Farh, Werbel, & Bedeian, 1988; Folger & Greenberg, 1985; Greenberg & Folger, 1983; Lind, Kurtz, Musante, Walker, & Thibault, 1980); increase communication between the supervisor and the worker (Cardy & Dobbins, 1994); and, provide unique and valuable information to the appraisal process (Lin, Fahr, Dobbins, Doyle, & Spaulding, 1992).

The present study will examine contextual factors that may influence the quality of self-ratings during the self-appraisal process. To better understand how the self-appraisal process may be improved, this study will examine the effects of the availability of social comparison information and purpose of appraisal on the leniency and accuracy of self-appraisals.

Social Comparison Information

The presence or absence of social comparison information is one contextual factor that may influence the psychometric properties of self-ratings. The lack of accuracy and inherent bias (leniency) associated with the self-appraisal are assumed to be the fault of the self-rater rather than a flaw in the rating procedure (Somers & Birnbaum, 1991). However, the outcome of the rating process is dependent on the complex information-processing capabilities of the rater (Cairns & Green, 1979). As such, a performance appraisal requires the rater to process information via placing "raw" performance data on some type of anchored (Tversky & Kahneman, 1974) or standardized scale. Following the GIGO model (garbage in = garbage out), if

the information available to the rater is limited or of poor quality, the potential accuracy of the ratings will also be limited. Therefore, accuracy may be increased (and leniency decreased) by the provision of information that helps calibrate perceptions of performance. Heneman (1986) suggests that presenting information in a relative fashion may be helpful to the rater when evaluating performance. The importance of relative information in the evaluation process is evident in social comparison theory.

Festinger's (1954) social comparison theory proposes that individuals are motivated to form an accurate self-evaluation of their abilities. Festinger delineates two standards against which abilities may be evaluated: physical reality and social reality. Individuals first desire to conduct a self-evaluation based on a physical standard (e.g., number of widgets produced in one hour). However, when such standards are not available, individuals evaluate their ability via comparison to relevant others' abilities (i.e., social reality). Finally, when neither absolute nor relative information is available, evaluations of abilities will be unstable. Jellison and Arkin (1977) believe that individuals are motivated to seek social comparison information of others whom they will be competing or cooperating with in order to gain rewards and avoid punishments. Over the years, a large body of research has accumulated which supports these basic tenets of social comparison theory (see Suls & Wills, 1991, for a review).

In the same vein, White's (1959) theory of competence motivation proposes that individuals are motivated to master their environment. According to White (1959), seeking out performance feedback (e.g., social comparison information) facilitates the attainment of competence as well as

adaptation to one's environment. In short, an individual is motivated to accurately process information about the environment in order to gain competence.

There is a large body of evidence in the social psychology literature which suggests that women are superior to men in accurately perceiving and processing non-verbal communication (e.g., Hall, 1984; Mayo & Henley, 1981). Hall (1984) conducted a series of meta-analyses in which large gender differences in non-verbal behavior (e.g., decoding skill, amount of gaze, personal space) emerged. For example, one meta-analysis of 64 studies revealed that women, on average, were superior to men in decoding non-verbal cues. Another meta-analysis of 30 studies revealed that women, on average, engage in more eye contact during social interactions than do men. In short, these findings suggest that women may be more sensitive than men to non-verbal cues (e.g., social comparison information). Therefore, the present study, as a precautionary measure, will randomly assign each gender group to each experimental condition.

In organizational settings, absolute or objective standards against which performance can be compared are extremely rare (Farh & Dobbins, 1989). Furthermore, even if absolute standards are available, performance evaluations are meaningful only when given in terms of relative standing with regard to an organizational reference group (Mabe & West, 1982).

Mabe and West (1982) suggest that the measure of performance is more relative than absolute. They further suggest that a relative performance evaluation better approximates the "true" criterion measure of ability. In turn, a performance appraisal worded in relative terms (i.e., average, below

average, better than, etc.) should have a higher correlation with "true" performance because of its probability of eliciting a comparative evaluation process. A meta-analysis by Mabe and West (1982) on 55 studies was able to support these hypotheses by showing a higher correlation between performance measures and self-appraisals when social comparison terminology was employed.

The aforementioned studies, coupled with social comparison theory, indicate that quantity of social comparison information plays an integral part in both increasing the accuracy of and decreasing the leniency of self-appraisals. However, little research in the performance appraisal domain has been conducted regarding the effects of quality of social comparison information on self-report accuracy and leniency. In other words, do qualitatively different forms of social comparison information have different effects on the quality of self-appraisals?

The purpose of the present study is to examine this dimension through the analysis of two qualitatively different forms of social comparison information: behavioral observation inference and structured information inference. The behavioral observation inference condition entails subjects directly observing the performance of comparison others. This condition is labeled "behavioral observation inference" because it requires the self-rater to make judgments based on observation in order to infer performance attributes to both the salient comparison others and the general population of all subjects in the study. In contrast, the structured information inference condition is achieved through the presentation of an extensive statistical summary sheet which delineates the performances of comparison others.

This condition is labeled "structured information inference" because the self-rater is presented with structured data regarding the entire population of subjects in the study from which to make inferences. Thus, the self-rater is not required to observe others' performances in order to generate and infer performance standards for the population. Finally, a no-information condition serves as a control.

It is possible that in organizational settings where both types of information are readily available (i.e., the self-rater is in a position to observe co-workers in addition to having access to written normative data) and relevant (e.g., job dimensions can be meaningfully quantified on normative data sheets) that information presented in written, structured form will be the more powerful. Two reasons why this may be the case are: (1) differential motivation in attending to the two types of information; and, (2) differential information-processing demands.

In terms of motivation to attend to comparative information, research by Rakestraw and Weiss (1981) on the interaction of social influence and task experience suggests that individuals may be less motivated to attend to comparative information presented via an observational medium. Rakestraw and Weiss (1981) conducted a study in which they demonstrated that observing the performance of another worker (in their study, watching a video of a student working on the same task) can serve as a standard by which individuals evaluate their own performance. Yet, they also demonstrated that this observational effect was significantly moderated by task experience. Subjects who were more familiar with the task (i.e., performed it previously) were less influenced by the social comparison model (i.e., the student in the

film). Rakestraw and Weiss (1981) suggested that this occurred because individuals generally turn to models under conditions of uncertainty, when personal experience fails to provide behavior guidelines. Consequently, when an individual has task-experience, the usefulness of a comparison model decreases.

Still, Rakestraw and Weiss (1981) note that subjects in all conditions, regardless of task experience and achievement, used the model as a standard of self-appraisal. Thus, while the basic tenets of social comparison theory are supported (e.g., self-evaluation based on social standards), the power of comparative information may be limited by the manner in which it is presented. For task-experienced individuals, comparative information obtained via observation may not be as helpful for calibrating perceptions of performance as other types (e.g., written summaries) of comparative information (assuming that other types are available and relevant in a given organizational setting).

In terms of information-processing demands, the issue becomes one of inferences. When an individual uses comparative information obtained by observation, he or she must first infer the performance distribution of the entire population (i.e., all co-workers) based on the observation of a sample of co-workers. Next, the individual must infer his or her performance on the inferred performance distribution of the entire population. In contrast, comparative information obtained by normative data sheets requires only one inference: that of inferring where to place one's self in the performance distribution of the population. From this perspective, an individual could be equally motivated to attend to both types of information. However, the extra

inference required by information obtained through observation, as well as its more ambiguous nature, may render it the less powerful of the two.

Purpose of Appraisal

Another contextual factor that may affect the quality of self-ratings is purpose of appraisal. Purpose of appraisal has received much attention in appraisal literature. Murphy and Cleveland (1991) suggest that it is important to understand the role of purpose of appraisal in the appraisal process because: (1) performance appraisals are used for many different purposes within an organization; (2) purpose of appraisal influences performance ratings and the appraisal process as a whole; and (3) the purpose of appraisal facilitates the interaction between context of rating and rating ability of the rater (Cleveland, Murphy, & Williams, 1989).

The Landy and Farr (1980) process model of performance appraisal delineates purpose of appraisal as a significant factor in determining both the psychometric properties of ratings (e.g., leniency, accuracy) and the utility of the appraisal process as a whole. Indeed, Landy and Farr (1980) stressed the importance of purpose of appraisal by calling for more research including experimental manipulation of the purpose variable. For the most part, purpose of appraisal has been categorized as either administrative (e.g., salary decisions, promotion, layoffs, etc.) or research (e.g., performance feedback, identification of training needs, etc.). Both field research (e.g., Bernardin & Orban, 1990; Bernardin, Orban, & Carlyle, 1981; Sharon & Bartlett, 1969) and laboratory research (e.g., Farh & Werbel, 1985; McIntyre, Smith, & Hassett, 1984; Zedeck & Cascio, 1982) have found that ratings are more lenient when used for administrative purposes as opposed to research purposes. One

possibility for these results, as proposed by Mohrman and Lawler (1983), is that leniency is more likely to occur when ratings are linked to positive or negative outcomes for the rater (e.g., administrative decisions). The present study examines this contextual factor through the manipulation of two purposes of appraisal: reward and performance feedback.

Overview of the Study

To examine the effects of contextual factors on self-appraisal quality, the following factors are manipulated: availability of social comparison information and purpose of appraisal. Both factors are manipulated to test predictions about their effects on both the leniency and accuracy of self-appraisals.

The first factor, availability of social comparison information, is manipulated both quantitatively and qualitatively. Quantitatively, availability of social comparison information is manipulated by allowing the experimental groups (i.e., behavioral observation inference, structured information inference) access to comparative information while the control group is not. Past research suggests that subjects granted access to social comparison information will be less lenient and more accurate (e.g., Farh & Dobbins, 1989; Mabe & West, 1982).

Qualitatively, availability of social comparison information is manipulated by presenting the information via two distinct mediums: observational (i.e., videos) and written (i.e., normative data sheets). Research on task experience and social influence (e.g., Rakestraw & Weiss, 1981) indicates that, due to subjects' task experience, the observational medium may be less powerful. Furthermore, the additional inference required in the

utilization of the observational comparative information, as well as its more ambiguous nature, suggests that it may not be as powerful as the structured, written information. Therefore, self-appraisals are predicted to be less lenient and more accurate when the available social comparison information is presented in the structured, written form.

Of further importance to the present study are the thought processes involved in the utilization of comparative information. To date, no research in the performance appraisal domain has assessed how self-raters use the available comparison information to calculate self-ratings. One purpose of the present study is to examine these thought processes via a free-recall measure.

The second factor, purpose of appraisal, is manipulated by delineating to the subjects one of two purposes at the time of self-appraisal: performance feedback or reward. Past research suggests that ratings will be more lenient when ratings can be linked to a positive outcome for the rater (Farh & Werbel, 1986; Mohrman & Lawler, 1983). Hence, because ratings in the reward condition can be linked to a positive outcome for the self-rater (i.e., extra-credit point), ratings made for this purpose are predicted to be more lenient and less accurate than ratings made for the purpose of performance feedback.

SELF-APPRAISAL: A LITERATURE REVIEW

Performance appraisals may be conducted by using supervisor, peer, or self-ratings. Although each has its advantages and disadvantages (Landy & Farr, 1980), it has traditionally been assumed that supervisors provide the best source of ratings. However, many researchers (e.g., Cardy & Dobbins, 1994) believe that self-appraisals will become more prevalent in the future as organizations move towards the implementation of work teams (Sundstrom et al., 1990) and high levels of worker participation. Other researchers (e.g., Lin et al., 1992) have proposed that self-appraisal is often the only method of evaluation for jobs that are complex and/or require highly specialized knowledge, skills, and abilities. Furthermore, recent reviews by Muchinsky (1986) and Schmitt and Noe (1986) suggest that a properly conducted self-appraisal may be a better evaluation tool than other conventional personnel selection methods. These reviews compared the various selection methods using such criteria as fairness, validity, and cost. Self-appraisals were found to have high fairness, moderate validity, and be relatively inexpensive.

Despite these potential applications of self-appraisals, researchers and practitioners remain skeptical about the quality of self-appraisals in formal appraisal contexts (Boyle & Klimoski, 1993). Historically, this skepticism seems to arise from the beliefs that (1) self-appraisals are subject to self-enhancement biases such as leniency; and, (2) most people are not able to evaluate themselves objectively enough to provide an accurate self-appraisal (DeNisi & Shaw, 1977; Farh, Werbel, & Bedeian, 1988). Each of these beliefs will be addressed below.

Leniency

Leniency is commonly cited as the primary shortcoming of self-appraisals (e.g., Landy & Farr; Meyer, 1980; Thornton, 1980). Leniency errors reflect an inappropriate judgmental standard on the part of the rater, whereby assigned ratings are unduly lenient or positive (Bernardin & Beatty, 1984). Murphy and Cleveland (1991) define leniency errors as distributional errors thought to be present when the distribution of ratings is different than the (assumed) distribution of job performance. More specifically, if the mean rating given by a certain rater is substantially higher than the scale midpoint, the rater is thought to be overly lenient.

Leniency has serious implications for the performance appraisal process. Because lenient ratings are, by definition, higher than they should be, any generated performance feedback based on those ratings will be inaccurate (Saal & Knight, 1988). Leniency also presents a problem to the extent that a restriction of range restricts the potential relationship between ratings and related criteria (Hoffman, Nathan, & Holden, 1991; Holzbach, 1978; Shrauger & Osberg, 1981; Thornton, 1980).

The literature on leniency errors offers many possible explanations for the existence of leniency in self-appraisals. Traditionally, the existence of leniency has been assumed to be the result of a deliberate distortion of the rater's actual impression (i.e., private appraisal) (Bernardin & Beatty, 1984). As such, leniency is a function of motivational pressures weighing on the rater as he or she fills out the performance appraisal form (Hauenstein, 1992). For example, Saal and Knight (1988) suggest that a rater may be motivated to exaggerate the quality of performance at the time of formal appraisal because

of: (1) a desire to be liked; or, (2) a fear that other raters will exaggerate ratings, therefore giving them an unfair advantage when promotions or pay raises become available.

Hauenstein (1992) proposes the encoding bias model as an alternative explanation for the existence of leniency. As discussed by Hauenstein (1992), social information-processing is complex and susceptible to many distortions and biases. Therefore, it is improbable that the traditional motivational perspective, which assumes distortion occurs only at the time of formal appraisal, can offer a comprehensive explanation for the existence of leniency. The motivational perspective assumes that contextual factors don't affect information-processing but do motivate raters to inflate their ratings at the time of appraisal. In contrast, Hauenstein's (1992) encoding bias model proposes that contextual factors affect both the selective attention and encoding phases of the judgment process. As such, leniency is not solely a product of motivational biases occurring at the time of appraisal.

From this encoding bias perspective, it is plausible that the self-rater may be accurately reporting his or her private self-appraisal; however, the private self-appraisal may be inaccurate due to encoding biases. For example, if the self-rater knows that, somewhere down the road, self-appraisals of performance will be used to calculate raises, he or she may selectively attend to positive performance incidents. Hence, by the time there is a formal self-appraisal, the self-rater may have developed an inflated private appraisal that he or she genuinely believes to be accurate.

Empirical evidence suggests that self-ratings are more lenient than either supervisor or peer ratings (Holzbach, 1978; Klimoski & London, 1974).

Leniency was operationally defined by Holzbach (1978) as present when supervisor, peer, and self-ratings are significantly different. Holzbach found that, for all items on a performance questionnaire, self-ratings had higher mean ratings than either superior or peer ratings. However, opposite results have also been reported. Heneman (1974) compared self- and supervisor ratings on various performance dimensions and found that self-appraisals possessed less leniency, halo error, and restriction of range than did supervisor ratings.

Harris and Schaubroeck (1988) recently conducted a quantitative review of empirical research on agreement among self-, peer, and supervisor ratings. The meta-analysis, which was conducted on 36 pairs of correlations, found that self-ratings were approximately one-half standard deviation higher than supervisor ratings. Moreover, another meta-analysis, conducted by Mabe and West (1982), found that 15 of 21 studies reported data which indicated that people overestimate their abilities. The meta-analysis by Mabe and West also found a low correlation, on average, between self-ratings and other rating sources.

However, the large body of research comparing self- and supervisor ratings, offers a wide range of findings. For example, Klimoski and London (1974) reported an average correlation between self- and supervisor ratings of .05, while Williams and Seiler (1973) found an average correlation of .60. The Harris and Schaubroeck (1988) meta-analysis found an average correlation of .35 after correcting for measurement error and range restriction. As such, no firm conclusions can be drawn regarding the extent of agreement between self- and supervisor ratings (Harris & Schaubroeck, 1988). Indeed, Landy and

Farr (1980) have suggested that the limited agreement between self ratings and other rating sources is not due to anything beyond that to be expected from utilizing multiple raters.

In general, the available research on leniency tends to suggest that when individuals are asked to appraise their abilities or performance, they tend to furnish inflated evaluations. However, research on leniency also suggests that contextual factors may influence the extent to which leniency exists. Recent research by Farh and Dobbins (1989) showed that leniency decreased as comparative information was made available to help calibrate perceptions of performance. The present study will take this line of research a step further by examining the effects of *type* of comparison information on self-appraisal leniency. In other words, is one type of comparative information more powerful in terms of reducing leniency? Additionally, are the thought processes involved in self-appraisal differentially affected by comparative information type? Another contextual factor, purpose of appraisal, has also been shown to influence leniency. Research by Farh and Werbel (1986) demonstrated that self-appraisals conducted for administrative purposes were found to be more lenient than those conducted for research purposes. The present research will examine the effects of both of these contextual factors on the leniency of self-appraisals.

Accuracy

Self-appraisals are often assumed to be inaccurate because of the belief that they are inherently lenient. However, mounting evidence in the empirical literature presents a more favorable picture regarding the capacity for accurate self-appraisals. In a recent review of the literature on self-

appraisal, Shrauger and Osberg (1981) compared self-appraisals to other methods commonly used in evaluation (e.g., test scores, grades, external rater evaluations, past performance, peer ratings) in the prediction of academic performance, outcomes of psychotherapeutic interventions, vocational choice, and job performance. Their results indicated that self-appraisals are at least as valid, if not more, as the other assessment methods with which they had been compared. Further evidence suggests that self-appraisals converge with academic/mental ability (e.g., Maki & Swett, 1987), academic performance (e.g., LaVoie & Hodapp, 1987), personality traits (e.g., Hurley, 1988), and job performance (e.g., Harris & Schaubroeck, 1988). As such, the literature indicates that individuals do possess the capacity to appraise themselves accurately.

Nevertheless, there is a group of studies which has examined the validity of self-appraisals and found them to be inaccurate (e.g., Mabe & West, 1982; DeNisi & Shaw, 1977). For example, Mabe and West (1982) conducted a meta-analysis of 55 studies in which self-appraisals of ability were compared with performance measures. Self-appraisals of ability were found to be only moderately correlated with the measures of performance ($M=.29$, $SD=.25$). In another study investigating the validity of self-appraisals, DeNisi and Shaw (1977) had subjects complete a questionnaire asking them to rate themselves on 10 ability areas. Subjects were then given a battery of tests designed to test these same areas. The researchers found that the correlations between self-appraised and tested abilities were too small (although most were significant) to have any practical significance.

However, these results must be interpreted with caution. As emphasized by Cardy & Dobbins (1994), these studies, like most accuracy research on self-appraisals, focused on self-appraisal of ability as opposed to self-appraisal of performance. As such, self-appraisals of ability were correlated with objective tests of ability. In short, self-appraisals were used to predict ability scores. Cardy & Dobbins (1994) suggest that limited research has examined the accuracy of self-appraisals of performance in a postdictive design, because of the difficulty of obtaining "true" measures of job performance. The present research addresses this problem by generating objective scores of performance which are used in the analysis of self-appraisal accuracy.

Research on self-appraisals provides evidence that individuals possess the capacity for accurate self-appraisal. Nevertheless, self-appraisals are often found to be inaccurate. As such, the interesting question becomes one of how to improve upon self-appraisal quality. One approach when addressing this question is to assume that individuals are capable of accurate self-appraisal but may be motivated by contextual factors to do otherwise (Boyle & Klimoski, 1993). Indeed, Murphy and Cleveland (1991) propose that, "... it is more reasonable to assume that raters are indeed capable of accurately evaluating performance, and that deficiencies in ratings are more likely to be a result of the rater's willingness to provide accurate ratings than of his or her capacity to rate accurately." (p. 191).

Another approach when addressing this question is to assume that individuals are reporting what they truly believe to be accurate self-appraisals.

However, as discussed previously, contextual factors may be facilitating encoding biases which lead to inaccurate private self-appraisals.

The present study examines the effects of contextual factors on the quality of self-appraisals. More specifically, the effects of availability of social comparison information and the purpose of appraisal on self-appraisal leniency and accuracy are examined.

Utility

The available literature indicates that self-appraisals are potentially valuable sources of information for performance appraisal purposes (Farh et al., 1988). Indeed, researchers have delineated many potential advantages that may result from incorporating self-appraisals into the traditional performance appraisal process (e.g., Cardy & Dobbins, 1994; Farh et al., 1988; Latham & Wexley, 1981; Maroney & Buckeley, 1992).

Perhaps the greatest advantage of utilizing self-appraisals is the increased likelihood of agreement among workers regarding organizational fairness and justice (Bies & Shapiro, 1988; Folger & Greenberg, 1985; Greenberg, 1987; Sheppard, 1985). Dipboye and Pontbriand (1981) propound that the long-term effectiveness of a performance appraisal system is as subject to employee opinions as it is to the reliability and validity of the performance appraisal measures. They further posit that positive perceptions toward the process are dependent on participation in the process. Research by Bies and Shapiro (1988) on perceptions of procedural fairness also indicates that when ratees are allowed a voice in the appraisal process, ratings of procedural fairness are significantly higher than when ratees are not afforded the opportunity for input. More importantly, research suggests that the use of

self-appraisals increases the acceptance of decisions based on appraisal information (Farh, Werbel, & Bedeian, 1988; Folger & Greenberg, 1985; Greenberg & Folger, 1983; Lind, Kurtz, Musante, Walker, & Thibault, 1980), even when decisions are unfavorable to the employee (Dipboye & Pontbriand, 1981).

The use of self-appraisals is also advantageous in that it increases communication between supervisors and subordinates regarding job content, performance criteria, and mutual expectations. As Cardy & Dobbins (1994) suggest, formalizing the self-appraisal process facilitates the identification of major discrepancies between self- and supervisor ratings. Once identified, these discrepancies can be addressed, and attention can be focused on altering the perceptions of either the self-appraiser or supervisor. In turn, this process should lead to greater convergence in ratings and greater acceptance of feedback.

Another advantage of using self-appraisals is that they provide unique and valuable information for the appraisal process. Self-appraisers are often in the privileged position of being able to observe their own performance and job knowledge first-hand. Indeed, self-appraisers are frequently more familiar with their own performance than are supervisors; and, therefore, are in a position to provide accurate evaluations (Klimoski & London, 1974). Furthermore, as was previously mentioned, other researchers (e.g., Lin et al., 1992) have proposed that self-appraisal is often the only method of evaluation for jobs that are complex and require highly specialized knowledge, skills, and abilities. Self-appraisals are also valuable in situations where employees work in isolation or without supervision.

In sum, it is evident that a self-appraisal that minimizes measurement problems and biases has potential utility for an organization. This utility stems from three sources: (1) increasing the favorability of perceptions of the appraisal process as a whole; (2) increasing communication between the superior and subordinate; and, (3) the provision of valuable performance information for the appraisal process.

Quality

As discussed earlier, there are many potential advantages of incorporating self-appraisals into the traditional appraisal process. Hence, many researchers have acknowledged that self-appraisals are a valuable resource for an organization. This acknowledgment has facilitated a trend towards research that focuses on the factors which affect the quality of self-appraisals and on the boundary conditions that restrict their utility (Cardy & Dobbins, 1994). The current research focuses on the effects of social comparison information and purpose of appraisal on self-appraisal quality. Past research on these two contextual factors is presented below.

Social Comparison Information. The availability of social comparison information is one contextual factor that may affect the quality of self-appraisals. Much of the research on social comparison information is based on Festinger's (1954) seminal work *A Theory of Social Comparison Processes*. Festinger's (1954) social comparison theory proposes that individuals are motivated to form an accurate self-evaluation of their abilities. Festinger delineates two standards against which abilities may be evaluated: physical reality and social reality. Individuals first desire to conduct a self-evaluation based on a physical standard (e.g., number of widgets produced in one hour).

However, when such standards are not available, individuals evaluate their ability via comparison to relevant others' abilities (i.e., social reality). Finally, when neither absolute nor relative information is available, evaluations of abilities will be unstable. Over the years, a large body of research has accumulated which supports these basic tenets of social comparison theory (see Suls & Wills, 1991, for a review).

In the same vein, White's (1959) theory of competence motivation proposes that individuals are motivated to master their environment. According to White (1959), seeking out performance feedback (e.g., social comparison information) facilitates the attainment of competence as well as adaptation to one's environment. In short, an individual is motivated to accurately process information about the environment in order to gain competence.

There is a large body of evidence in the social psychology literature which suggests that women are superior to men in accurately perceiving and processing non-verbal communication (e.g., Hall, 1984; Mayo & Henley, 1981). Hall (1984) conducted a series of meta-analyses in which large gender differences in non-verbal behavior (e.g., decoding skill, amount of gaze, personal space) emerged. For example, one meta-analysis of 64 studies revealed that women, on average, were superior to men in decoding non-verbal cues. Another meta-analysis of 30 studies revealed that women, on average, engage in more eye contact during social interactions than do men. In short, these findings suggest that women may be more sensitive than men to non-verbal cues (e.g., social comparison information). Therefore, the

present study, as a precautionary measure, will randomly assign each gender group to each experimental condition.

Performance Appraisal. In the domain of performance appraisal, absolute or objective standards against which performance can be compared are extremely rare (Farh & Dobbins, 1989). Thus, it is necessary for workers to compare themselves to other workers in order to determine their ability levels. Indeed, Mumford (1983) believes that, in organizational settings, fellow workers are logical social comparison targets since they are likely to be working on similar tasks, under comparable conditions. Moreover, Jellison and Arkin (1977) conducted a literature review of social comparison research and concluded that individuals seek out information concerning the abilities of others with whom they will be competing or cooperating with (e.g., co-workers) in order to gain rewards and avoid punishments.

One implication of the social comparison viewpoint is that a poor quality self-appraisal may occur because the self-rater does not have adequate social comparison information. Hence, as suggested by Festinger (1954), when social standards are not available, self-evaluations will be unstable. In a traditional performance appraisal situation, the supervisor has access to performance information about the entire work group (e.g., how many widgets each worker produces in an hour, how many days each worker was late to work, etc.) from which to calibrate perceptions of performance. Workers, on the other hand, rarely have the benefit of such extensive information. As such, one possibility for the traditional findings that self-appraisals are more lenient and less accurate than supervisor ratings is the

disparity of the availability of social comparison information between the two rating sources.

Another implication of the social comparison viewpoint is that performance appraisal involves more of a relative comparison than an absolute one. As mentioned earlier, absolute standards against which performance can be appraised are very rare. Furthermore, Mabe and West (1982) propound that even when absolute standards are available, they are meaningful only when presented in terms of one's relative standing in a work group. For example, knowing that you produce 100 widgets a day provides little information about your level of performance unless you know that most people only produce 75 widgets a day. Hence, Mabe and West (1982) propose that the measure of performance is more relative than absolute. They further submit that using social comparison terminology (e.g., average, below average, better than, etc.) "exerts considerable influence on the validity of self-evaluations" (p. 294). Therefore, a performance appraisal worded in relative terms should have a higher correlation with objective performance because of its probability of eliciting a comparative evaluation process. A meta-analysis by Mabe and West (1982) on 55 studies was able to support these propositions by showing a higher correlation between performance measures and self-appraisals when social comparison terminology was employed.

A recent laboratory study conducted by Farh and Dobbins (1989) examined the effects of social comparison information on self-appraisal. In this study, subjects served as editors for a fictitious publication. Subjects were randomly assigned to either the social comparison or control condition. The social comparison manipulation occurred after the subjects had conducted

editorial tasks (e.g., proofreading) for 30 minutes. After this work period, subjects in the social comparison condition were given the opportunity to evaluate the work of their co-workers. Subjects were instructed to pay close attention to certain performance dimensions (e.g., quantity, quality). In contrast, subjects in the control condition were not able to evaluate their co-workers work. Furthermore, the laboratory was set up so that subjects from both conditions were unable to observe their co-workers' performance. Subjects were then asked to complete a self-appraisal of their performance. Analyses indicated that self-appraisals were more accurate (i.e., more highly correlated with objective performance measures) when subjects were given access to the social comparison information.

The aforementioned studies, coupled with social comparison theory, indicate that quantity of social comparison information plays an integral part in both increasing the accuracy of and decreasing the leniency of self-appraisals. For example, Farh and Dobbins (1989) demonstrated that presenting self-raters with comparative information in written form increases self-appraisal accuracy. However, little research in the performance appraisal domain has been conducted regarding the effects of quality of social comparison information on self-report accuracy and leniency. In other words, do qualitatively different forms of social comparison information (e.g., visual vs. written) have different effects on the quality of self-appraisals? The purpose of the present study is to examine this dimension through the analysis of two qualitatively different forms of social comparison information: behavioral observation information and structured information.

The behavioral observation inference condition entails subjects directly observing the performance of comparison others. This condition is labeled "behavioral observation inference" because it requires the self-rater to make judgments based on observation in order to infer performance attributes to both the salient comparison others and the general population of all subjects in the study. In contrast, the structured information inference condition is achieved through the presentation of an extensive statistical summary sheet which delineates the performances of comparison others. This condition is labeled "structured information inference" because the self-rater is presented with structured data regarding the entire population of subjects in the study from which to make inferences. Thus, the self-rater is not required to observe others' performances in order to generate and infer performance standards for the population. Finally, a no-information condition serves as a control.

It is possible that in organizational settings where both types of information are readily available (i.e., the self-rater is in a position to observe co-workers in addition to having access to written normative data) and relevant (e.g., job dimensions can be meaningfully quantified on normative data sheets) that information presented in written, structured form will be more powerful. Two reasons why this may be the case are: (1) differential motivation in attending to the two types of information; and, (2) differential information-processing demands.

In terms of motivation to attend to comparative information, research by Rakestraw and Weiss (1981) on the interaction of social influence and task experience suggests that individuals may be less motivated to attend to comparative information presented via an observational medium.

Rakestraw and Weiss (1981) conducted a study in which they demonstrated that observing the performances of other workers can serve as a standard by which individuals evaluate their own performance. Yet, they also demonstrated that this observational effect was significantly moderated by task experience. In their study, all subjects were given written instructions for a card sorting task, with approximately half the subjects given the opportunity to work on the task. All subjects then viewed a "training film" of either a high- or low-performing student (model). All subjects then worked on the task, and their performance, goals, and satisfaction were assessed. Results indicated that subjects who were more familiar with the task (i.e., performed it before watching the film) were less influenced by the social comparison model. Rakestraw and Weiss (1981) suggested that this occurred because individuals generally turn to models under conditions of uncertainty, when personal experience fails to provide behavior guidelines. Consequently, when an individual has task-experience, the usefulness of a comparison model decreases.

This finding is relevant to the current study in that subjects observe the comparison others after performing the proofreading task. As such, subjects are task-experienced at the time of observation and, as suggested by Rakestraw and Weiss (1981), may be less influenced by the comparative information. Nevertheless, it is important to present the comparative information in this sequence (i.e., after task performance) because this is most likely how it would occur in an organizational setting. After all, what good would it do to have a worker conduct a self-appraisal of performance before performing the task?

Still, Rakestraw and Weiss (1981) note that subjects in all conditions, regardless of task experience and achievement, used the model as a standard of self-appraisal. Thus, while the basic tenets of social comparison theory are supported (e.g., self-evaluation based on social standards), the power of comparative information may be limited by the manner in which it is presented. For task-experienced individuals, comparative information obtained via observation may not be as helpful for calibrating perceptions of performance as other types (e.g., written summaries) of comparative information (assuming that other types are available and relevant in a given organizational setting).

In terms of information-processing demands, the issue becomes one of inferences. When an individual uses comparative information obtained by observation, he or she must first infer the performance distribution of the entire population (i.e., all co-workers) based on the observation of a sample of co-workers. Next, the individual must infer his or her performance on the inferred performance distribution of the entire population. In contrast, comparative information obtained by normative data sheets requires only one inference: that of inferring where to place one's self in the performance distribution of the population. From this perspective, an individual could be equally motivated to attend to both types of information. However, the extra inference required by information obtained through observation, as well as its more ambiguous nature, may render it the less powerful of the two.

Purpose of Appraisal. Another contextual factor that may affect the quality of self-ratings is purpose of appraisal. Purpose of appraisal has received much attention in the performance appraisal literature. Murphy

and Cleveland (1991) suggest that it is important to understand the role of purpose of appraisal in the appraisal process because: (1) performance appraisals are used for many different purposes within an organization; (2) purpose of appraisal influences performance ratings and the appraisal process as a whole; and (3) the purpose of appraisal facilitates the interaction between context of rating and rating ability of the rater (Cleveland, Murphy, & Williams, 1989).

The Landy and Farr (1980) process model of performance appraisal delineates purpose of appraisal as a significant factor in determining both the psychometric properties of ratings (e.g., leniency, accuracy) and the utility of the appraisal process as a whole. Indeed, Landy and Farr (1980) stressed the importance of purpose of appraisal by calling for more research including experimental manipulation of the purpose variable. For the most part, purpose of appraisal has been categorized as either administrative (e.g., salary decisions, promotion, layoffs, etc.) or research (e.g., performance feedback, identification of training needs, etc.).

Farh and Werbel (1986) conducted a study which investigated the effects of purpose of appraisal (grading or research) on self-appraisal leniency. Undergraduate students enrolled in an organizational behavior course served as subjects in this study. Students' self-appraisals of their class participation were collected during the final exam via a short questionnaire. Purpose of appraisal was manipulated by the instructional sets on the questionnaires. The instructional set in the grading purpose informed students that borderline grades would be adjusted upward if class participation was high. The instructional set in the research purpose informed students that a survey

about class participation was being conducted for research purposes and that the survey results would not affect grades. Subsequent analyses showed that purpose of appraisal significantly affected self-appraisal leniency (i.e., overestimation of participation). Subjects in the grading condition had more lenient self-appraisals than those in the research condition.

Bernardin and Orban (1990) recently conducted one of the few studies investigating purpose of appraisal in a real organizational setting. The effects of purpose on appraisal leniency were studied in two law enforcement organizations. Ratings in one organization were to be used as a basis for promotional decisions, whereas ratings in the other organization were to be used only for feedback on past performance. With the exception of purpose, there were no potentially confounding differences between the organizations. Bernardin and Orban (1990) found, consistent with their hypothesis, that purpose of appraisal accounted for a significant amount of rating variance, with ratings in the "promotion" organization being systematically higher than those in the "feedback" organization.

For the most part, both field research (e.g., Bernardin & Orban, 1990; Bernardin, Orban, & Carlyle, 1981; Sharon & Bartlett, 1969) and laboratory research (e.g., Farh & Werbel, 1985; McIntyre, Smith, & Hassett, 1984; Zedeck & Cascio, 1982) have found that ratings are more lenient when used for administrative purposes as opposed to research purposes. One possibility for these results, as proposed by Mohrman and Lawler (1983), is that leniency is more likely to occur when ratings are linked to positive or negative outcomes for the rater (e.g., administrative decisions).

However, it must be noted that opposite results have also been found. Dobbins, Farh, and Werbel (1993) recently conducted a study which examined the effects of purpose on inflation of grade point averages (GPAs). Subjects reported their GPAs on both a research survey and an application form at the college placement service. Contrary to previous findings, inflation was found to be more prevalent on the research survey than on the application form. Dobbins, et al. (1993) suggest that this may have occurred because, while students had more to gain by inflating their GPAs on the application form, they probably had a higher expectation of validation on the application form. Thus, various contextual factors (e.g., expectation of validation) may decrease leniency in self-appraisals even when they are used for administrative purposes.

The present study examines this contextual factor through the manipulation of two purposes of appraisal: reward and performance feedback. Subjects in the reward condition are told that the top three performers will receive an additional extra credit point towards their psychology grade (i.e., positive outcome). Subjects in the performance feedback condition are told that a research team will be analyzing their self-appraisals in order to generate feedback regarding their perceptual performance. Therefore, although this purpose offers no "reward" or positive outcome, it is still self-relevant to the subjects.

This purpose of this study is to assess the effects of social comparison information and purpose of appraisal on the leniency and accuracy of self-appraisals of performance.

Hypotheses

Leniency Hypothesis 1: Leniency will decrease as information that helps calibrate perceptions of performance becomes available. Therefore, leniency in the no-information condition will be greater than leniency in the conditions where social comparison information is available. Moreover, because information in the structured information inference condition is more helpful for calibrating perceptions of performance than information in the behavioral observation inference condition, leniency in the behavioral observation inference condition, which will be greater than leniency in the structured information inference condition.

Leniency Hypothesis 2: Leniency is more likely to occur when ratings can be linked to a positive outcome for the rater (Fahr & Werbel, 1985; Mohrman & Lawler, 1983). Therefore, self-ratings in the reward condition will be more lenient than self-ratings in the performance feedback condition.

Leniency Hypothesis 3: Because self-ratings of quantifiable performance dimensions are easier to validate than self-ratings of ambiguous performance dimensions, they are less likely to be lenient (Farh & Werbel, 1989). As such, there will be a dimension by purpose of appraisal interaction such that there will be a smaller purpose of appraisal effect on the quantifiable dimensions (i.e., Quantity and Quality) than on the ambiguous dimensions (i.e., Effort and Conscientiousness).

Accuracy Hypothesis 1: Because leniency puts a ceiling on the accuracy of ratings (Saal & Knight, 1988), ratings in the structured information inference condition will be more accurate than ratings in the behavioral

observation inference condition, which will be more accurate than ratings in the no-information condition.

Accuracy Hypothesis 2: Because leniency puts a ceiling on the accuracy of ratings (Saal & Knight, 1988), self-ratings in the performance feedback condition will be more accurate than self-ratings in the reward condition.

Utilization of Social Comparison Information Hypothesis 1: The utilization of social comparison information elicits a relative or comparative evaluation (Festinger, 1954). Therefore, in a free recall measure assessing the self-appraisal process, subjects in the structured information inference condition will record more comparative statements than subjects in the behavioral observation inference condition, who will record more comparative statements than subjects in the no-information condition.

Utilization of Social Comparison Information Hypothesis 2: As social comparison information becomes more helpful, raters will report that they utilized it more. Therefore, in response to the self-appraisal item "Evaluation of other students' proofreading abilities", ratings will be higher in the structured information inference condition than in the behavioral observation inference condition, which will be higher than ratings in the no-information condition.

Utilization of Social Comparison Information Hypothesis 3: Because utilization of social comparison information may prohibit the attainment of a positive outcome (i.e., receiving extra-credit for being one of the top three performers), ratings for the self-appraisal item "Evaluation of other students' proofreading abilities" will be higher in the performance feedback condition than ratings in the reward condition.

METHOD

Subjects

Ninety male and ninety female introductory psychology students from Virginia Polytechnic Institute and State University served as subjects for the current investigation.

Design.

The study utilized a 2 (sex) X 2 (purpose of appraisal) X 3 (social comparison information) factorial design (15 subjects per cell).

Experimental Task

Subjects were given a proofreading test which consisted of passages taken from a psychology textbook. The test contained 520 total lines (20 pages) and 100 possible mistakes (5 mistakes per page). The text of the test was enlarged to insure easy readability (see Appendix A). A pilot study was run in which 57 subjects performed the identical proofreading task. The data from this pilot study (see Appendix C) confirmed that subjects were not able to complete the proofreading test during the 15 minute work period (the maximum number of pages proofread was 16).

Procedure

Subjects reported to the study on an individual basis. Subjects were told that the purpose of the study was to research perceptual performance and that a proofreading test was needed to collect the necessary data. After signing a consent form (see Appendix H), the proofreading test was explained to the subjects. The instructor then gave subjects the proofreading test. After 15 minutes, the instructor asked the subjects to stop working, and collected the test. Subjects were then told that there would be a five minute rest period

and that they would be evaluating their performance at the end of this period. At the end of the rest period, subjects were given the performance appraisal sheets, instructed on how to fill them out, and briefed on what the purpose of self-appraisal was. Subjects were then given 15 minutes to evaluate their performance. Finally, subjects were debriefed concerning the true purpose of the study (see Appendix I).

Independent Variables

Social Comparison Information.

Structured Information Inference Group. Subjects in this group were asked to examine two normative data sheets (see Appendix D) during the five minute rest period. The normative data sheets contained descriptive statistics on the objective measures of Quantity and Quality of the 57 subjects in the pilot study who had performed the identical task. These graphs provided three types of information regarding the Quantity and Quality of other students' performances: (1) the frequency of students performing at various performance levels; (2) the percentile rank of various performance levels; and, (3) the average performance score. The normative data sheets were explained to the subjects (see Appendix G for a full protocol) who were told that the normative data was based on other students participating in the study.

Behavioral Observation Inference Group. Subjects in this group were asked to watch videotapes of other participants in the study during the five minute rest period (see Appendix F for a full protocol). Subjects were told that the students in the videos had been proofreading for 10 minutes and that they were to watch them complete the task for the next five minutes. The

subjects were presented with the three five-minute videotapes simultaneously. One video monitor was positioned directly in front of subjects, with the other two monitors positioned on either side of it. The videotapes were of confederates giving representative performances based on data from the pilot study (see Appendix C). Each video represented a different performance level: poor, average, or excellent.

The poor performance video represented performance two standard deviations below the mean on both Quantity (5.5 pages completed) and Quality (an average of .5 necessary corrections made per page) dimensions. Since the video was of the final five minutes of performance, subjects observed the confederate proofreading from three-quarters of the way down the fourth page (i.e., 3.75) to half way down the sixth page (i.e., 5.5) while detecting an average of .5 mistakes per page. To provide a representative sample of Quality, proofreading errors (i.e., circling correctly spelled words or not circling misspelled words) were equally dispersed within the text proofread.

The average performance video represented performance at the mean on both Quantity (9.5 pages completed) and Quality (an average of 2.5 necessary corrections made per page) dimensions. Subjects observed the confederate proofreading from half way down the seventh page (i.e., 6.5) to half way down the tenth page (i.e., 9.5) while detecting an average of 2.5 mistakes per page.

The excellent performance video represented performance two standard deviations above the mean on both Quantity (13.5 pages completed) and Quality (an average of 4.5 necessary corrections made per page)

dimensions. Subjects were able to observe the confederate proofreading from the beginning of the tenth page (i.e., 9 pages completed) to half way down the fourteenth page (i.e., 13.5) while detecting an average of 4.5 mistakes per page.

In all three videos, subjects were only able to see the hand of the confederate performing the proofreading task. Confederates with hands that minimized gender specificity were chosen to avoid a potential confounding with the independent variable due to sex differences between the subjects and social comparison targets. The videos were randomly assigned to one of the three monitors to avoid position effects. Large page numbers and red circles served as indicators of Quantity and Quality, respectively (see Appendix A).

Control Group. Subjects in the control group were not provided with social comparison information. During the five minute rest period they were simply asked to relax (see Appendix E for a full protocol).

Purpose Of Appraisal.

Reward Group. After the five minute rest period, subjects in this group were given the performance appraisal sheets, instructed on how to fill them out, and told that the top three performers would be receiving an additional extra-credit point towards their psychology grade (i.e., a positive outcome).

Performance Feedback Group. After the five minute rest period, subjects in this group were given the performance appraisal sheets, instructed on how to fill them out, and told that a research team would be analyzing their self-appraisals in order to generate feedback regarding their perceptual performance. Hence, although this purpose offered no reward or positive outcome, it was still self-relevant to the subjects.

Dependent Measures

Self-Ratings. Performance appraisal forms (see Appendix B) were used to assess subjects' self-ratings of performance. Subjects were asked to rate their performance on four dimensions: Quantity; Quality; Effort; and, Conscientiousness. All dimensions were rated on a nine point graphic rating scale anchored by 1 = (worse than anyone else), 3 = (below average), 5 = (average), 7 = (above average), 9 = (better than anyone else). Quantity was defined as the total number of pages proofread. Quality was defined as the average number of mistakes detected per page. Effort was defined as the amount of energy directed towards the task. Conscientiousness was defined as the motivation to do well. Since the first two indices may have reduced leniency due to their quantitative nature, the latter two served as analogous subjective indices (Quantity~Effort, Quality~Conscientiousness). Subjects were then asked to compute their Overall Performance Score by summing their scores on the four indices.

Objective Performance Measures. Objective performance measures were calculated for the dimensions of Quantity and Quality. Quantity was determined by calculating the total number of pages (rounding to the quarter-page) proofread during the 15 minute period. The Quality score was computed by subtracting the number of improper corrections (i.e., false positives) made from the number of proper corrections (i.e., true hits) made for every page of text, then summing these values and dividing by the number of pages proofread. Both of these objective performance measures (Quantity and Quality) were then placed in an appropriate stanine of the norm group in order to obtain objective performance scores (see Appendix J).

Leniency. Leniency was determined by comparing the mean scores on each of the four dimensions (Quantity, Quality, Effort, and Conscientiousness) as well as the overall performance score (OPS) for each condition. Consistent with Somers and Birnbaum's (1991) operational definition of leniency, leniency was said to exist if the mean OPS or dimension score of the condition was significantly greater than that of another condition.

Accuracy. Accuracy scores play an important role as dependent measures in analyzing the determinants of accuracy and error in performance appraisal (see Zedeck & Cascio, 1982). Cronbach (1955) proposes that accuracy can be broken down into four separable and independent component accuracy scores (Sulsky & Balzer, 1988). However, because the present study involved a single rater (i.e., the self-appraiser), only the two components of accuracy that are applicable in a single rater situation were utilized to operationalize accuracy.

The first component of accuracy operationalized was elevation (E^2) (Cronbach, 1955). Elevation refers to the accuracy of the average rating, over all dimensions (i.e., Quantity and Quality) (Murphy & Cleveland, 1991). In the present study elevation was defined as the square root of the following:

$$E^2 = (x - t)^2$$

where x , and t , = mean observed score and mean true score over all dimensions. For this component, perfect accuracy was said to be achieved if

the self-appraiser's average observed score was equal to the average of the true scores.

The second component of accuracy operationalized was dimensional accuracy (DA^2) (Hauenstein & Alexander, 1991). Dimensional accuracy was used because of the single ratee (i.e., the self-appraiser) situation.

Conceptually, dimensional accuracy is analagous to Cronbach's (1955) differential accuracy in that it measures the accuracy with which each self-appraiser evaluates him/herself on each dimension. In the present study dimensional accuracy was defined as the square root of the following:

$$DA^2 = 1/n \sum [(x_j - x.) - (t_j - t.)]^2$$

where n = number of dimensions, x_j and t_j = observed score and true score on dimension j , and $x.$ and $t.$ = mean observed score and mean true score over all dimensions. For this component, perfect accuracy was said to be achieved if both a correlation of positive one between a rater's observed scores and the true scores existed, and a rater's variance for his/her ratings was equal to the variance of the true scores (Hauenstein & Alexander, 1991).

Free Recall of Social Comparison Information. To date, no research in the performance appraisal domain has examined the thought processes involved in utilizing comparative information in the self-appraisal process. In order to assess this self-appraisal process, the performance appraisal sheets (see Appendix B) contained the open-ended item "Please describe how you came up with your ratings". Two undergraduate raters trained in the experimental protocol recorded the number of comparative responses to this

item (e.g., "I proofread more than the person in the video") in addition to recording the total number of responses. Interrater agreement for these two measures was .90 and .89, respectively. Any discrepancies between the two raters were resolved by the experimenter. The number of comparative responses as well as the percentage of relative responses (i.e., (number of comparative responses/number of total responses) * 100) were used in all subsequent analyses.

Utilization of Social Comparison Information. The performance appraisal sheets contained items which subjects were asked to rate based on their degree of influence on the performance evaluation process. The item "Evaluation of other students' proofreading abilities" assessed the utilization of social comparison information. Subjects were asked to rate this item based on a five point graphic rating scale anchored by 1 = (not at all), 3 = (somewhat influenced), and 5 = (totally influenced).

Manipulation Check

The item "Potential for extra credit" on the performance appraisal sheet assessed the efficacy of the purpose of appraisal manipulation. If the manipulation was effective, subjects responses to this item would be higher in the reward condition than in the performance feedback condition. Subjects were asked to rate these items based on a five point graphic rating scale anchored by 1 = (not at all), 3 = (somewhat influenced), and 5 = (totally influenced).

RESULTS

Manipulation Checks

To check on the validity of experimental manipulations, the various manipulation checks were analyzed using 2 (sex) X 2 (purpose of appraisal) X 3 (social comparison information) analyses of variance. Summary tables for these analyses are presented in Appendix K.

The purpose of appraisal manipulation was assessed by subjects' responses to the item "Please indicate the degree to which each of the following influenced your performance appraisal: Potential for extra-credit". As anticipated, the main effect of purpose of appraisal emerged as the only significant source of variance $F(1, 168) = 77.06, p < .01$ (see Table K-1). This purpose of appraisal effect indicated that subjects in the reward condition were more influenced by the potential for extra-credit ($M = 3.08, SD = 1.20$) than subjects in the performance feedback condition ($M = 1.72, SD = .85$). Therefore, the purpose of appraisal manipulation was highly effective.

When interpreting distributional errors, such as leniency, there is an underlying assumption that there is no variation, from worker to worker, in terms of actual performance (Murphy & Balzer, 1989). For, if one worker performed at a higher level than another worker, the fact that he or she produced higher self-ratings does not substantiate the existence of leniency. To ensure that random assignment to groups controlled for performance differences among experimental conditions, 2 X 2 X 3 analyses of variance were performed on the mean scores of Quantity and Quality.

These analyses indicated that social comparison information was not a significant source of variance for either Quantity $F(2, 168) = .45, p > .05$, or

Quality $F(2, 168) = 2.44, p > .05$ (see Table K-2). In other words, there was no significant difference in the mean Quantity score among the control group ($M = 4.25, SD = 2.30$), the behavioral observation inference group ($M = 4.50, SD = 2.12$), and the structured information inference group ($M = 4.63, SD = 2.35$); and, no significant difference in the mean Quality score among the control group ($M = 4.60, SD = 1.64$), the behavioral observation inference group ($M = 5.18, SD = 1.70$), and the structured information inference group ($M = 5.15, SD = 1.61$).

These analyses also indicated that purpose of appraisal was not a significant source of variance for either Quantity $F(1, 168) = .93, p > .05$, or Quality $F(1, 168) = 1.65, p > .05$ (see Table K-2). Therefore, there was no significant difference in the mean Quantity score between the performance feedback group ($M = 4.62, SD = 2.20$) and the reward group ($M = 4.30, SD = 2.31$); and, no significant difference in the mean Quality score between the performance feedback group ($M = 4.82, SD = 1.55$) and the reward group ($M = 5.13, SD = 1.77$).

One unexpected result of these analyses was the emergence of a sex effect as a marginally significant source of variance for Quantity $F(1, 168) = 3.09, p = .08$, and a significant source of variance for Quality $F(1, 168) = 6.13, p < .05$ (see Table K-2). These analyses revealed that, for Quantity, males ($M = 4.17, SD = 2.26$) performed at a slightly lower level than females ($M = 4.76, SD = 2.21$); and, for Quality, males ($M = 4.67, SD = 1.75$) performed at a lower level than females ($M = 5.28, SD = 1.52$). Hence, there appears to be performance differences in Quantity and Quality as a function of sex, with the females outperforming the males in both areas. Because of these

performance differences, it was decided to include Quantity score as a covariate in all subsequent analyses of leniency which included sex as an independent variable and Quantity and/or Effort (the dimension analogous to Quantity) self-ratings as dependent variables. Likewise, it was decided to include Quality score as a covariate in all subsequent analyses of leniency which included sex as an independent variable and Quality and/or Conscientiousness (the dimension analogous to Quality) self-ratings as dependent variables. Finally, it was decided to include both scores (Quantity and Quality) as covariates in all analyses dealing with the Overall Performance Score, due to its composition of both Quantity and Quality self-ratings.

Leniency

Leniency Hypotheses 1 and 2 were analyzed using 2 (sex) X 2 (purpose of appraisal) X 3 (social comparison information) analyses of covariance.

Leniency Hypothesis 1. Leniency Hypothesis 1 predicted that leniency in the no-information condition would be greater (i.e., self-ratings would be significantly higher) than leniency in the behavioral observation inference condition, which would be greater than leniency in the structured information inference condition.

For the Quantity dimension, the 2 X 2 X 3 ANCOVA yielded an unexpected significant sex by purpose of appraisal by social comparison information interaction $F(2, 167) = 3.97, p < .05$ (see Table L-1). Because this second-order interaction was statistically significant, a series of simple interaction effects analyses were conducted (see Pedhauer & Schmelkin, 1991). These analyses revealed that there was a significant, first-order sex by

social comparison information interaction in the reward condition $F(2, 167) = 4.48, p < .05$ (see Figure L-2), but not in the performance feedback condition $F(2, 167) = 1.73, p > .05$ (a significant sex by purpose of appraisal interaction also emerged, and will be discussed under Leniency Hypothesis 2).

Subsequent simple-simple effects analyses indicated that, in the reward condition, there was a significant effect for social comparison information for males $F(2, 83) = 8.12, p = .001$, but not for females $F(2, 83) = .13, p > .05$. For males, a-priori contrasts showed that, while self-ratings in the no-information condition ($M = 6.73, SD = 1.49$) were significantly higher than self-ratings in the behavioral observation inference condition ($M = 5.07, SD = 1.71$), $t(42) = 2.69, p = .01$, self-ratings in the behavioral observation inference condition were not significantly higher than self-ratings in the structured information inference condition ($M = 4.73, SD = 1.87$), $t(42) = .54, p > .05$.

For the Quality dimension, the 2 X 2 X 3 ANCOVA yielded the anticipated main effect of social comparison information $F(2, 167) = 12.25, p < .001$ (see Table L-4). A-priori contrasts confirmed that self-ratings in the no-information condition were higher ($M = 6.03, SD = 1.50$) than self-ratings in the behavioral observation inference condition ($M = 5.27, SD = 1.40$), $t(177) = 2.89, p < .005$. However, self-ratings in the behavioral observation inference condition were not significantly higher than self-ratings in the structured information inference condition ($M = 5.38, SD = 1.46$), $t(177) = -.44, p > .05$.

For the Effort dimension, the 2 X 2 X 3 ANCOVA did not yield the expected main effect of social comparison information $F(2, 167) = 2.41, p > .05$. Self-ratings were not significantly different among the no-information condition ($M = 6.48, SD = 1.24$) the behavioral observation inference condition

($M = 6.35$, $SD = 1.07$), and the structured information inference condition ($M = 6.01$, $SD = 1.24$). Therefore, the hypothesis was not supported for this dimension.

For the Conscientiousness dimension, the 2 X 2 X 3 ANCOVA revealed a significant sex by social comparison information interaction $F(2, 167) = 4.10$, $p < .05$ (see Table L-6). Subsequent simple effects analyses indicated that there was a significant effect for social comparison information for both males $F(2, 174) = 3.66$, $p < .05$, and females $F(2, 174) = 5.03$, $p < .01$.

For males, a-priori contrasts showed that, while self-ratings in the no-information condition ($M = 6.93$, $SD = 1.05$) were significantly higher than self-ratings in the behavioral observation inference condition ($M = 6.17$, $SD = 1.05$), $t(87) = 2.56$, $p < .05$, self-ratings in the behavioral observation inference condition were not significantly higher than self-ratings in the structured information inference condition ($M = 6.60$, $SD = 1.13$), $t(87) = -1.45$, $p > .05$. Indeed, male self-ratings in the structured information inference condition were higher than self-ratings in the behavioral observation inference condition. This latter fact appears to have facilitated the emergence of this disordinal interaction (see Figure L-7).

For females, a-priori contrasts indicated that self-ratings in the no-information condition ($M = 7.10$, $SD = .96$) were not significantly higher than self-ratings in the behavioral observation inference condition ($M = 6.70$, $SD = 1.02$), $t(87) = 1.49$, $p > .05$. But, self-ratings in the behavioral observation inference condition were significantly higher than self-ratings in the structured information inference condition ($M = 6.20$, $SD = 1.13$), $t(87) = 1.87$, $p = .06$.

For the Overall Performance Score, the 2 X 2 X 3 ANCOVA yielded the anticipated main effect of social comparison information $F(2, 166) = 21.40, p < .001$ (see Table L-8). A-priori contrasts confirmed that the Overall Performance Score in the no-information condition was higher ($M = 25.50, SD = 2.88$) than the Overall Performance Score in the behavioral observation inference condition ($M = 23.25, SD = 3.55$), $t(177) = 3.73, p < .001$. However, the Overall Performance Score in the behavioral observation inference condition was not significantly higher than the Overall Performance Score in the structured information inference condition ($M = 23.10, SD = 3.45$), $t(177) = .248, p > .05$.

Leniency Hypothesis 2. Leniency Hypothesis 2 predicted that self-ratings in the reward condition would be more lenient than self-ratings in the performance feedback condition.

As previously discussed, for the Quantity dimension, the 2 X 2 X 3 ANCOVA yielded an unexpected significant sex by purpose of appraisal by social comparison information interaction $F(2, 167) = 3.97, p < .05$ (see Table L-1). Simple interaction analyses revealed that there was a significant, first-order sex by purpose of appraisal interaction in the behavioral observation inference condition $F(1, 167) = 4.97, p < .05$ (see Figure L-3), but not in the no-information condition $F(1, 167) = 1.73, p > .05$, or in the structured information inference condition $F(1, 167) = 1.34, p > .05$.

Simple-simple effects analyses indicated that, in the behavioral observation inference condition, there was a significant effect for purpose of appraisal for females $F(1, 55) = 9.78, p < .005$, but not for males $F(1, 55) = .12, p > .05$. This effect confirmed that, for females in the behavioral observation

inference condition, self-ratings in the reward condition ($M = 5.73, SD = 1.13$) were significantly higher than self-ratings in the performance feedback condition ($M = 5.00, SD = 1.13$). Additionally, a main effect for sex emerged which indicated that, within the behavioral observation inference condition, male self-ratings in the performance feedback condition ($M = 5.67, SD = 1.35$) were significantly higher than female self-ratings in the performance feedback condition ($M = 5.00, SD = 1.13$) $F(1, 55) = 4.68, p < .05$ (see Figure L-3).

For the Quality dimension, the $2 \times 2 \times 3$ ANCOVA produced the expected main effect of purpose of appraisal $F(1, 167) = 5.37, p < .05$ (see Table L-4), with self-ratings in the reward condition ($M = 5.84, SD = 1.56$) being significantly higher than self-ratings in the performance feedback condition ($M = 5.28, SD = 1.35$).

For the Effort dimension, the $2 \times 2 \times 3$ ANCOVA yielded the expected main effect of purpose of appraisal $F(1, 167) = 3.69, p = .056$ (see Table L-5). This effect confirmed that self-ratings in the reward condition were higher ($M = 6.46, SD = 1.29$) than self-ratings in the performance feedback condition ($M = 6.11, SD = 1.08$).

For the Conscientiousness dimension, the $2 \times 2 \times 3$ ANCOVA revealed a main effect for purpose of appraisal $F(1, 167) = 4.56, p < .05$ (see Table L-6). Self-ratings in the reward condition ($M = 6.81, SD = 1.12$) were significantly higher than self-ratings in the performance feedback condition ($M = 6.42, SD = 1.13$).

For the Overall Performance Score, the $2 \times 2 \times 3$ ANCOVA yielded the anticipated main effect of purpose of appraisal $F(1, 166) = 10.96, p < .001$ (see Table L-8), indicating that the Overall Performance Score in the reward

condition was higher ($M = 24.69$, $SD = 3.74$) than the Overall Performance Score in the performance feedback condition ($M = 23.21$, $SD = 3.02$).

Leniency Hypothesis 3. Leniency Hypothesis 3 predicted that there would be a dimension by purpose of appraisal interaction such that there would be a smaller purpose of appraisal effect on the quantifiable dimensions (i.e., Quantity and Quality) than on the ambiguous dimensions (i.e., Effort and Conscientiousness). This hypothesis was analyzed using 2 (sex) X 2 (purpose of appraisal) X 3 (social comparison information) X 2 (dimension type: quantifiable, ambiguous) mixed-factor analyses of variance, with dimension type serving as a within-subjects factor. Separate 2 X 2 X 3 X 2 mixed ANOVAs were run for the "Quantity" dimensions (i.e., Quantity vs. Effort) and the "Quality" dimensions (i.e., Quality vs. Conscientiousness) (see Table L-11).

For the "Quantity" dimensions, a significant sex by purpose of appraisal by social comparison information by dimension type interaction emerged $F(2, 168) = 3.26$, $p < .05$ (see Table L-11). This four-way interaction was not unexpected, due to the previous finding that there was a significant sex by purpose of appraisal by social comparison information interaction for the Quantity dimension $F(2, 167) = 3.97$, $p < .05$ (see Table L-1), but not for the Effort dimension $F(2, 167) = 1.50$, $p > .05$ (see Table L-5). Likewise, the emergence of a significant sex by social comparison information by dimension type interaction was not surprising, given that previous simple-simple effects analyses had revealed a significant sex by social comparison information interaction for the Quantity dimension, whereas this interaction did not surface for the Effort dimension (see Leniency Hypothesis 1). The fact

that there were multiple effects for the Quantity dimension, and only one marginally significant effect for the Effort dimension (see Figure L-12) makes these results all the more difficult to interpret.

However, given that the aforementioned significant three-way interaction did not involve purpose of appraisal, as well as given the fact that the pattern of the four-way interaction was such that it did not compromise interpretation of purpose of appraisal, it is still possible to determine if there was a significant dimension type by purpose of appraisal interaction. As such, simple effects analyses were conducted to determine if there was a larger purpose of appraisal effect for the ambiguous dimension (i.e., Effort) than for the quantifiable dimension (i.e., Quantity). These simple effects analyses confirmed that the purpose of appraisal effect size was significant for the ambiguous dimension $F(1, 168) = 3.85, p = .05, \eta^2 = .02$, and was not significant for the quantifiable dimension $F(1, 168) = 1.52, p = .22, \eta^2 = .009$. Thus, these findings provide support for Leniency Hypothesis 3. Of additional note, a main effect for dimension type emerged $F(1, 168) = 29.15, p < .001$ (see Table L-11), indicating that self-ratings on the ambiguous dimension (i.e., Effort) ($M = 6.28$) were significantly higher than self-ratings on the quantifiable dimension (i.e., Quantity) ($M = 5.56$).

For the "Quality" dimensions, the expected dimension type by purpose of appraisal interaction did not emerge $F(1, 168) = .52, p > .05$ (see Table L-11). Therefore, Leniency Hypothesis 3 was not supported. However, there was a main effect for dimension type $F(1, 168) = 73.08, p < .001$ (see Table L-11), indicating that self-ratings on the ambiguous dimension (i.e.,

Conscientiousness) ($M = 6.62$) were significantly higher than self-ratings on the quantifiable dimension (i.e., Quality) ($M = 5.56$).

Accuracy

For Accuracy Hypotheses 1 and 2, elevation and dimensional accuracy were used as dependent measures in separate 2 (sex) X 2 (purpose of appraisal) X 3 (social comparison information) analyses of variance.

Accuracy Hypothesis 1. Accuracy hypothesis 1 predicted that self-ratings (i.e., Quantity and Quality) in the structured information inference condition would be more accurate than self-ratings in the behavioral observation inference condition, which would be more accurate than self-ratings in the no-information condition.

For elevation, an unanticipated sex by social comparison information interaction emerged $F(2, 168) = 3.50, p < .05$ (see Table L-13, Figure L-14). Simple effects analyses showed a significant main effect for social comparison information for both males $F(2, 174) = 22.39, p < .001$, and females $F(2, 174) = 6.57, p < .005$. Additionally, a main effect for sex emerged which indicated that males were significantly less accurate than females in the no-information condition $F(1, 174) = 8.67, p < .005$. This latter effect appears to be what drove the interaction.

A-priori contrasts revealed that, for males, self-ratings in the behavioral observation inference condition ($M = .85, SD = .72$) were significantly more accurate than self-ratings in the no-information condition ($M = 2.18, SD = 1.23$), $t(87) = 5.65, p < .001$. However, self-ratings in the structured information inference condition ($M = .97, SD = .69$) were not

significantly more accurate than self-ratings in the behavioral observation inference condition $t(87) = -.49, p > .05$.

For females, a-priori contrasts revealed that self-ratings in the behavioral observation inference condition ($M = 1.00, SD = .60$) were significantly more accurate than self-ratings in the no-information condition ($M = 1.53, SD = 1.09$), $t(87) = 2.61, p < .05$. However, self-ratings in the structured information inference condition ($M = .75, SD = .55$) were not significantly more accurate than self-ratings in the behavioral observation inference condition $t(87) = 1.23, p > .05$.

For dimensional accuracy, the $2 \times 2 \times 3$ ANOVA yielded the expected social comparison information effect $F(1, 168) = 4.62, p < .05$ (see Table L-13). A-priori contrasts confirmed that self-ratings in the no-information condition were less accurate ($M = 1.01, SD = .96$) than self-ratings in the behavioral observation inference condition ($M = .84, SD = .63$), $t(177) = 1.97, p < .05$. However, self-ratings in the structured information inference condition were not significantly more accurate ($M = .71, SD = .56$) than self-ratings in the behavioral observation inference condition, $t(177) = .99, p > .05$.

Accuracy Hypothesis 2. Accuracy hypothesis 2 predicted that self-ratings in the performance feedback condition would be more accurate than self-ratings in the reward condition.

For elevation, the $2 \times 2 \times 3$ ANOVA yielded the expected social comparison information effect $F(1, 168) = 27.34, p < .001$ (see Table L-13), indicating that self-ratings in the reward condition were less accurate ($M = 1.42, SD = 1.01$) than self-ratings in the performance feedback condition ($M = 1.00, SD = .90$).

For dimensional accuracy, the 2 X 2 X 3 ANOVA yielded an unexpected sex by purpose of appraisal effect $F(1, 168) = 4.77, p < .05$ (see Table L-13, Figure L-15). Simple effects analyses showed a significant effect of purpose of appraisal for females $F(1, 176) = 4.77, p < .05$, but not for males $F(1, 176) = .72, p > .05$. This effect revealed that, for females, self-ratings in the performance feedback condition ($M = .67, SD = .56$) were more accurate than self-ratings in the reward condition ($M = 1.01, SD = .84$). Moreover, a main effect for sex emerged, indicating that males were significantly less accurate ($M = 1.00, SD = .79$) than females ($M = .67, SD = .56$) in the feedback condition.

Utilization of Social Comparison Information (SCI)

Utilization of SCI Hypotheses 1, 2, and 3 were analyzed using 2 (sex) X 2 (purpose of appraisal) X 3 (social comparison information) analyses of variance.

Utilization of SCI Hypothesis 1. Utilization of SCI Hypothesis 1 predicted that, in a free recall measure assessing the self-appraisal process, subjects in the structured information inference condition would record more comparative statements than subjects in the behavioral observation inference condition, who would record more comparative statements than subjects in the no-information condition. For this hypothesis, number of comparative remarks and percentage of comparative remarks ($\# \text{ comparative remarks} / \text{total } \# \text{ of remarks} * 100$) served as dependent measures in separate 2 X 2 X 3 ANOVAs.

For number of comparative remarks, support for the hypothesized effect was provided by a significant main effect for social comparison information $F(2, 168) = 11.10, p < .001$ (see Table L-18). A-priori contrasts

revealed that this effect resulted from subjects in the structured information inference condition recording more comparative remarks ($M = 1.72, SD = 1.38$) than subjects in the behavioral observation inference condition ($M = 1.20, SD = 1.39$) ($t(177) = 2.28, p < .05$), who recorded more comparative remarks than subjects in the no-information condition ($M = .65, SD = .90$), $t(177) = 2.42, p < .05$.

For percentage of comparative remarks, the $2 \times 2 \times 3$ ANOVA revealed a significant main effect for social comparison information $F(2, 168) = 4.12, p < .001$ (see Table L-18). A-priori contrasts confirmed that subjects in the structured information inference condition recorded a higher percentage of comparative remarks ($M = 38.55, SD = 31.94$) than subjects in the behavioral observation inference condition ($M = 21.24, SD = 24.34$) ($t(177) = 3.81, p < .001$), who recorded a higher percentage of comparative remarks than subjects in the no-information condition ($M = 10.24, SD = 15.60$), $t(177) = 2.42, p < .05$.

Utilization of SCI Hypothesis 2. Utilization of SCI Hypothesis 2 predicted that, in response to the self-appraisal item "Evaluation of other students' proofreading abilities", ratings would be higher in the structured information inference condition than in the behavioral observation inference condition, which would be higher than ratings in the no-information condition.

This hypothesis was supported by the emergence of a significant social comparison information main effect $F(2, 168) = 45.96, p < .001$ (see Table L-19). A-priori contrasts confirmed that subjects in the structured information inference condition rated this item higher ($M = 3.18, SD = 1.07$) than subjects in the behavioral observation inference condition ($M = 2.85, SD = .94$), $t(177) =$

1.98, $p < .05$). Furthermore, subjects in the behavioral observation inference condition rated this item significantly higher than subjects in the no-information condition ($M = 1.67$, $SD = .73$), $t(177) = 7.04$, $p < .001$.

Utilization of SCI Hypothesis 3. Utilization of SCI Hypothesis 3 predicted that ratings for the self-appraisal item "Evaluation of other students' proofreading abilities" would be higher in the performance feedback condition than ratings in the reward condition.

This hypothesis was supported by the emergence of a significant social comparison information main effect $F(1, 168) = 5.25$, $p < .05$ (see Table L-19) which indicated that subjects in the performance feedback condition rated this item higher ($M = 2.72$, $SD = 1.07$) than subjects in the reward condition ($M = 2.41$, $SD = 1.16$).

DISCUSSION

The goal of the present study was to test the effects of contextual factors on the quality of self-appraisals. More specifically, the current research examined the effects of social comparison information and purpose of appraisal on the leniency and accuracy of self-appraisals of performance. Although the hypotheses were not supported as originally stated, several conclusions are possible based on the pattern of obtained results.

The results clearly indicate that, in general, people are sensitive to social comparison information. Moreover, as proposed by the basic tenets of social comparison theory, it appears as though this information is utilized in the self-appraisal of performance and has a positive effect on self-appraisal quality by decreasing leniency while increasing accuracy. It was hypothesized that the more helpful social comparison information was for calibrating perceptions of performance, the more it would be utilized during the self-appraisal process. This hypothesis was supported by the finding that self-appraisals were more influenced by the evaluation of others' abilities when social comparison information was made available; and, when available, the more helpful it was for calibrating perceptions of performance. Furthermore, the free recall measure, which assessed the self-appraisal process, revealed that both number of comparative remarks and percentage of comparative remarks were highest when comparison information was more helpful for calibrating perceptions of performance.

As such, in terms of memory processing, the more helpful the comparative information, the more it is used in cognitive processing. However, as depicted by the patterns of leniency and accuracy which emerged,

this ordering effect does not manifest itself in self-ratings. These patterns suggest a trend for self-ratings to be more lenient and less accurate when social comparison information is not made available; however, the quality of the comparison information (i.e., how helpful it is for calibrating perceptions of performance) does not appear to affect self-appraisal leniency and accuracy. Although there were some exceptions, the provision of social comparison information appears to reduce leniency and increase accuracy while the quality of the information does not appear to be influential.

This latter finding is interesting relative to the finding that the ordering effect emerged on the free recall measure. One possibility for the lack of an ordering effect for the leniency and accuracy measures has to do with the use of judgment data. Research on information-processing in performance appraisal has suggested that of the three types of measures used in information-processing research (i.e., encoding, retrieval, and judgment), judgment data (e.g., self-ratings) provide the least insight into how performance information is processed by raters (Foti & Hauenstein, 1993). Without some type of encoding or retrieval measures, mediating processes can only be inferred from judgment data. This reliance on inference has implications for research on the effects of contextual factors on self-appraisal quality. For example, if the current research had been interpreted based solely on the pattern of self-ratings (i.e., judgment data), it would have been concluded that the quality of comparison information had no influence on the self-appraisal process. Yet, it could be that self-ratings were not as sensitive a measure as the free recall, resulting in their failure to discriminate among the social comparison conditions. The results from the free recall

paint a different picture regarding the role of quality of social comparison information in the self-appraisal process. Hence, the role of comparison information quality should not be discounted until this discrepancy between self-ratings and free recall is investigated. Further, information-processing research is warranted which examines the cause of this discrepancy.

Results also indicate that self-raters are sensitive to the purpose of appraisal. As hypothesized, when the purpose of self-appraisal is performance feedback, self-appraisals tend to be less lenient and more accurate than when the purpose is linked to a reward for the self-rater. With two exceptions, self-ratings were more lenient and less accurate when conducted for a reward purpose. Moreover, it appears as though purpose of appraisal has an impact on the degree of influence social comparison has in the self-appraisal process. Indeed, subjects in the performance feedback condition indicated that they were more influenced by the evaluation of others' abilities than subjects in the reward condition. This suggests that purpose of appraisal may moderate the utilization of social comparison information.

Given the above results, one may inquire as to what the causal agents are underlying the leniency (and inaccuracy) of self-appraisals. With regard to contextual factors and what causes their effects on self-appraisal quality, four theoretical perspectives are particularly applicable.

The first perspective, which will be called the lack-of-information perspective, views self-appraisal leniency as caused primarily by the lack of comparative information available to the self-raters (Cardy & Dobbins, 1994). Self-raters are perceived as motivated to form an accurate self-appraisal, yet

not having enough comparative information to do so. Support for this perspective is grounded in research comparing self-ratings when social comparison information is available to self-ratings when social comparison information is not available. The results of this research (as well as the current research) usually show that leniency is higher (and accuracy lower) when comparative information is not available (e.g., Farh & Dobbins, 1989; Mabe & West, 1982). The current data were consistent with this view in that leniency was higher when comparison information was unavailable.

However, this lack-of-information perspective appears to be too simplistic. Most jobs are deeply embedded in a social environment saturated with comparative information yet self-appraisal leniency continues to be a pervasive problem. Indeed, the current study illustrated that people can transform limited comparative information (e.g., behavioral observation information) into comparative performance standards equally as well as those formed from the use of detailed, quantified information. As such, it is likely that the majority of organizational settings have adequate levels of social comparison information to make sound performance judgments. Why then does self-appraisal leniency continue to surface?

Another perspective, that of rendering-bias (Banks & Murphy, 1985), offers a second explanation. Perhaps the most popular perspective on leniency is that individuals are capable of accurate *private* self-appraisal (via the use of social comparison information, etc.); however, they are motivated to distort at the *public* appraisal level (Bernardin & Beatty, 1984; Saal & Knight, 1988). From this rendering-bias perspective, leniency is caused by motivational pressures weighing on the self-rater as he or she fills out the

performance appraisal form. This view would counter that of the lack-of-information perspective by arguing that there is enough social comparison information in the work environment to form accurate self-impressions; but, because of contextual pressures at the time of formal self-appraisal, this accurate self-impression is ignored and self-ratings are inflated. The current data also may be used to support this perspective. The emergence of a purpose of appraisal effect suggests that people were motivated by this contextual pressure (i.e., potential for extra-credit) to inflate their ratings. However, it is unknown if this finding represents the inflation of accurate self-impressions.

A third perspective, that of encoding-bias (Hauenstein, 1992), offers a more comprehensive explanation regarding the cause(s) of leniency. Hauenstein (1992) stresses that information-processing is complex and susceptible to many biases. As such, the rendering-bias perspective, which focuses on distortion only at the point of formal appraisal, is lacking. The encoding-bias perspective views contextual factors as influential in biasing the selective attention to and encoding of performance information as well as the rendering of the formal appraisal. For example, if the self-rater knows that, somewhere down the road, self-appraisals of performance will be used to calculate raises, he or she may selectively attend to positive performance incidents. Hence, by the time there is a formal self-appraisal, the self-rater may have developed an inflated private appraisal that he or she genuinely believes to be accurate. In turn, the public appraisal is an honest representation of the private appraisal. In terms of the current research, it is possible that self-raters who were providing ratings for the purpose of gaining

an extra-credit point selectively attended to information that supported a high self-appraisal (e.g., remembering how poorly someone else performed or recalling one's own positive performance incidents) which led to an inflated private appraisal and, in turn, an inflated public appraisal.

The large body of research on the self-enhancement motive and its effects on the processing of social comparison information (see Suls and Wills, 1991, for a review) provides an excellent framework from which to couch the encoding bias model. As proposed by Goethals, Messick, and Allison (1991), research on social comparison after Festinger's seminal paper in 1954 has emphasized the role of self-enhancement in the processing of social comparison information. Goethals et al., (1991) delineate a number of key extensions of social comparison theory since its original formulation. Of key importance is the role of downward social comparison and the need for self-enhancement. People often engage in social comparison in order to validate their positive assessment of their abilities. Moreover, people prefer comparison information to reflect well on the self. A second extension is that people often prefer not to engage in social comparison if comparison information is dissonant with high self-esteem. A third extension is that social comparison is often forced. A fourth extension is that people use attribution processes to interpret social comparison information. For example, comparison information that threatens self-esteem (e.g., noticing that others are performing at a higher level) is attributed to the environment (e.g., they have a better sales territory). Conversely, comparison information that maintains self-esteem (e.g., noticing that others are performing at a lower level) is attributed to the person (e.g., they are really bad workers). In short,

people will seek out information that is self-enhancing; avoid information that threatens self-esteem; and, if they can't avoid threatening information will downplay its significance through attribution.

These key extensions led Goethals et al., (1991) to propose two different types of social comparison: *realistic social comparison* and *constructive social comparison*. Realistic social comparison entails self-appraisal based on using actual information about social reality. This is analogous to the social comparison outlined by Festinger (1954). It involves the actual comparison with others' abilities in an attempt to place one's own ability on a continuum. This involves both upward and downward comparison in an attempt to achieve accurate appraisal of one's own performance. According to Goethals et al., (1991), people are motivated to engage in realistic social comparison when they are motivated to support a positive self-appraisal or, at least initially, when they are forced to by salient comparison information. However, if this forced information has dissonant qualities, over time people may switch to constructive social comparison.

Constructive social comparison entails self-appraisal based on "'in the head' social comparison based on guess, conjecture, or rationalization concerning social reality, often believed, and often self-serving" (Goethals et al., 1991, p. 154). This type of social comparison is engaged when people want to devise self-esteem-maintaining perceptions of social reality. With regard to the encoding-bias perspective, it is quite possible that people, in an attempt to maintain self-esteem, will selectively attend to comparative performance information that validates their high self-concept (e.g., focusing on others' poor performances). In turn, the encoding of performance incidents is biased

towards making the self look better. After time, this positive image of the above-average performing self may actually be believed because of all the stored positive performance incidents that back it up. This has serious implications for the assumption that social comparison information reduces leniency and increases accuracy. For, if esteem-maintenance does bias the selective attention and encoding of social comparison information, then the utilization of comparative information does not guarantee that self-ratings will be less lenient and more accurate as traditionally assumed. Indeed, self-raters may simply *believe* their ratings to be more accurate because they are more data based (e.g., they have plenty of incidents of others' poor performance to back up their claim).

A fourth perspective, that of expectation of validation, views leniency as a product of the self-rater's belief that the self-appraisal will not be validated. Indeed, Mabe and West (1982) suggest that one way of improving self-appraisal quality is to include instructions that self-ratings will be compared with other criterion measures. In light of the current investigation, it is possible that the social comparison information in and of itself did not reduce leniency and increase accuracy. Instead, it could be that, because the comparison information was *given* to the subjects by an authority figure, the self-raters became aware that there were criteria (e.g., performance distributions) against which their performance could be validated. That is, if the experimenter knew how everyone else performed it would be less likely that he or she would accept an inflated self-appraisal.

This perspective has major implications for the social comparison information effects on self-appraisal quality. Traditionally, the provision of

social comparison information is assumed to be sufficient for reducing leniency. But, without the expectation of validation it is possible that social comparison information may not be effective. The current study was unable to test this possibility directly since the source of the comparison information was not manipulated. In order to examine this possibility, it would be necessary to conduct a study where some subjects serendipitously absorbed comparative information (i.e., not being requested to do so by the experimenter) while other subjects were presented with comparative information from "higher up" (e.g., experimenter, supervisor, etc.). If self-ratings were less lenient in the latter group, the role of threat of validation in the reduction of self-appraisal leniency would be supported.

Because interpretation of the effects of contextual factors vary as a function of which perspective is adopted, further research in the area of motivational as well as perceptual processes underlying self-appraisal is warranted. While testing the validity of these perspectives was beyond the scope of the current investigation, the generated results do suggest ways to reduce self-appraisal leniency while increasing accuracy.

Irrespective of the "cause" of leniency, it is evident that the provision of social comparison information by an authority figure facilitates the reduction of leniency. Moreover, social comparison information may be more effective if presented at the time of appraisal. The results of the current study also suggest that the manner in which social comparison information is presented (i.e., observational versus written summaries) does not affect self-appraisal leniency. However, this latter finding must be interpreted with

caution; for, there are three possible explanations for why the hypothesized ordering effects did not emerge among the social comparison conditions.

One possibility, which was already discussed, is the reliance on judgment data (i.e., self-ratings) to operationalize leniency and accuracy. It could be that the self-ratings were not a sensitive enough measure to discriminate among social comparison conditions in terms of leniency and accuracy.

A second possibility has to do with the presentation of the social comparison information. Although the two types of comparison information were qualitatively different, the context in which they were presented was similar. Basically, the subjects were forced to attend to the comparison information under severe time constraints. As suggested by Goethals et al., (1991), people are motivated to engage in realistic social comparison when they are forced to by salient comparison information. However, if this forced information has dissonant qualities, *over time* people may switch to constructive social comparison. These constraints may have caused both groups to be equally motivated to engage in realistic social comparison which, in turn, resulted in their being no differences between them with regard to leniency and accuracy. However, with no time constraints, differences between the effects of comparison information absorbed observationally over time and comparison information presented in summary form at the time of appraisal may surface. For example, if workers are left to observe their co-workers performances and realize that they are not as good a performer as they thought, they may, over time, switch to constructive social comparison thereby selectively attending to and encoding

performance incidents that maintain self-esteem. Conversely, if comparison information is presented at the time of self-appraisal, self-raters may be forced to engage in realistic social comparison. Longitudinal research needs to be conducted to address this possibility.

A third possibility for why the ordering effects did not emerge has to do with expectation of validation. As discussed earlier, it is possible that it is the expectation of validation rather than the social comparison information per se which affects self-appraisal quality. As such, if the perceived expectation of validation was similar for both groups, an ordering effect would not be expected.

It is also evident that the purpose of self-appraisal has an affect on the leniency and accuracy of self-ratings. The current research suggests that self-appraisals are best suited for non-reward purposes such as feedback, employee development, and research. Self-appraisals conducted for reward purposes (e.g., administrative purposes) tend to be more lenient and inaccurate; therefore, rendering them less diagnostic in terms of differentiating among performance levels.

Future Research

The research provided here offers a strong starting point for conducting research on the effects of contextual factors on self-appraisal quality. The next step involves approaching a causal explanation for why these effects occur. Of primary significance is research which addresses the issue of whether social comparison information in and of itself is sufficient to affect self-appraisal quality or whether it must be coupled with an expectation of validation. Research is also needed which examines how comparison information is

cognitively processed as well as how self-raters are motivated to use this information once processed. It is important to delineate which of the three perspectives (lack-of-information, rendering bias, or encoding bias) provides the best explanation for why and how contextual factors affect self-appraisal quality. Future research must also search for ways to encourage realistic social comparison as opposed to constructive social comparison. Finally, longitudinal research should focus on the optimal time to present comparison information as well as when and if individuals switch from realistic to constructive social comparison.

References

- Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. Personnel Psychology, *38*, 335-345.
- Bernardin, H. J., & Beatty, R. W. (1984). Performance appraisal: Assessing human behavior at work (pp. 167-171). Boston, MA: Kent.
- Bernardin, H. J., & 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*, 197-211.
- Bernardin, H. J., Orban, J., & Carlyle, J. (1981). Performance ratings as a function of trust in appraisal and rater individual differences. Proceedings of the 41st annual meeting of the Academy of Management, 311-315.
- Bies, R. J., & Shapiro, D. L. (1988). Voice and justification: Their influence on procedural fairness judgments. Academy of Management Journal, *31*(3), 676-685.
- Boyle, K. A., & Klimoski, R. J. (1993). Improving the quality of self-assessments: The role of accountability. Paper presented at the Eighth Annual Conference of the Society for Industrial and Organizational Society, San Francisco, California.
- Cairns, R. B., & Green, J. A. (1979). How to assess personality and social patterns: Observations or ratings? In R. B. Cairns (Ed.), The analysis of social interactions: Methods, issues, and illustrations (pp. 209-226).
- Campbell, D., & Lee, C. (1988). Self-appraisal in performance evaluation: Development vs. evaluation. Academy of Management Review, *13*, 302-314.

- Cardy, R. L., & Dobbins, G. H. (1994). Performance appraisal: Alternative perspectives. Cincinnati, OH: Southwestern Publishing Co.
- Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applied Psychology, 74, 130-135.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". Psychological Bulletin, 52, 177-193.
- DeNisi, A. S., & Shaw, J. B. (1977). Investigation of the uses of self-reports of abilities. Journal of Applied Psychology, 62, 641-644.
- Dipboye, R. L., & Pontbriand, R. (1981). Correlates of employee reactions to performance appraisal and appraisal systems. Journal of Applied Psychology, 66(2), 248-251.
- Dobbins, G. H., Farh, J. L., & Werbel, J. D. (1993). The influence of self-monitoring on inflation of grade point average for research and selection purposes. Journal of Applied Social Psychology.
- Fahr, J., & Dobbins, G. (1989). Effects of comparative performance information on the accuracy of self-ratings and agreement between self- and supervisor ratings. Journal of Applied Psychology, 74(4), 606-610.
- Farh, J., & Werbel, J. D. (1986). Effects of purpose of appraisal and expectation of validation on self-appraisal leniency. Journal of Applied Psychology, 71(3), 527-529.
- Farh, J., Werbel, J. D., & Bedeian, A. G. (1988). An empirical investigation of self-appraisal-based performance evaluation. Personnel Psychology, 41, 141-156.

- Festinger, L. (1954). A theory of social comparison processes. Human Relations, *7*, 117-140.
- Folger, R., & Greenberg, J. (1985). Procedural justice: An interpretive analysis of personnel systems. In K. Rowland and G. Ferris (Eds.), Research in personnel and human resources management (Vol. 3, pp. 141-183). Greenwich, CT: JAI Press.
- Foti, R. J., & Hauenstein, N. M. A. (1993). Processing demands and the effects of prior impressions on subsequent judgments: Clarifying the assimilation/contrast debate. Organizational Behavior and Human Decision Processes, *56*, 167-189.
- Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls and T. Wills (Eds.), Social comparison: Contemporary theory and research. Hillsdale, NJ: Lawrence Erlbaum.
- Greenberg, J. (1987). A taxonomy of organizational justice theories. Academy of Management Review, *12*, 9-22.
- Greenberg, J., & Folger, R. (1983). Procedural justice, participation and the fair process effect in groups and organizations. In P. B. Paulhus (Ed.), Basic group processes (pp. 235-256). New York: Springer-Verlag.
- Hall, J. A. (1984). Nonverbal sex differences: Communication accuracy and expressive style. Baltimore, MD: Johns Hopkins University Press.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, *41*, 43-62.

- Hauenstein, N. M. A. (1992). An information-processing approach to leniency in performance judgments. Journal of Applied Psychology, 77, 1-9.
- Hauenstein, N. M. A., & Alexander, R. A. (1991). Rating ability in performance judgments: The joint influence of implicit theories and intelligence. Organizational Behavior and Human Decision Processes, 50, 300-323.
- Heneman, H. G. III. (1974). Comparisons of self and superior ratings of managerial performance. Journal of Applied Psychology, 58, 638-642.
- Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. Personnel Psychology, 39, 811-826.
- Hoffman, C., Nathan, B. R., & Holden, L. M. (1991). A comparison of validation criteria: Objective versus subjective performance measures and self- versus supervisor ratings. Personnel Psychology, 44, 601-619.
- Holzbach, R. (1978). Rater bias in performance ratings: Supervisory, self, and peer ratings. Journal of Applied Psychology, 63, 579-588.
- Hurley, J. R. (1988). Interpersonal correlates of individual's discrepant ratings by self and peers. Journal of Social Psychology, 128, 653-665.
- Jellison, J. M., & Arkin, R. M. (1977). Social comparison of abilities. Journal of Personality and Social Psychology, 31, 430-436.
- Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. Journal of Applied Psychology, 59, 445-451.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

- Latham, G. P., & Wexley, K. N. (1981). Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley.
- LaVoie, J. C., & Hodapp, A. F. (1987). Children's subjective ratings of their performance on a standardized achievement test. Journal of School Psychology, 25, 73-80.
- Lin, R., Farh, J., Dobbins, G. H., Doyle, T., & Spaulding, L. (1992). Rating anchors and self-assessments of ability: A field study. Paper presented at the Seventh Annual Conference of the Society for Industrial and Organizational Psychology, Montreal, Quebec.
- Lind, E. A., Kurtz, S., Musante, L., Walker, L., & Thibault, J. (1980). Procedural and outcome effects on reactions to adjusted resolution of conflicts of interest. Journal of Personality and Social Psychology, 39, 643-653.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review of meta-analysis. Journal of Applied Psychology, 67, 280-296.
- Maki, R. H., & Swett, S. (1987). Metamemory for narrative text. Memory and Cognition, 15, 72-83.
- Maroney, B. P., & Buckeley, M. R. (1992). Does research in performance appraisal influence the practice of performance appraisal? Regretfully not! Public Personnel Management, 21(2), 185-196.
- Mayo, C., & Henley, N. M. (1981). Gender and non-verbal behavior. New York: Springer-Verlag.
- 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, 147-156.

- Mohrman, A. M., & Lawler, E. E. (1983). Motivation and performance appraisal behavior. In F. Landy, S. Zedek, & J. Cleveland (Eds.), Performance measurement and theory. Hillsdale, NJ: Erlbaum.
- Meyer, H. (1980). Self-appraisal of job performance. Personnel Psychology, *33*, 291-295.
- Muchinsky, P. M. (1986). Personnel selection methods. In C. L. Cooper & I. T. Robertson (Eds.), International Review of Industrial and Organizational Psychology. New York: John Wiley and Sons.
- Mumford, M. D. (1983). Social comparison theory and the evaluation of peer evaluation: A review and some applied implications. Personnel Psychology, *36*, 867-881.
- Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. Journal of Applied Psychology, *74*, 619-624.
- Murphy, K. R., & Cleveland, J. N. (1991). Performance appraisal: An organizational perspective. Needham Heights, MA: Allyn/Bacon.
- Offerman, L. R., & Gowing, M. K. (1990). Organizations of the future: Changes and challenges. American Psychologist, *45*(2), 95-108.
- Pedhauzer, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Lawrence Erlbaum.
- Rakestraw, T. L., & Weiss, H. M. (1981). The interaction of social influences and task experience on goals, performance, and performance satisfaction. Organizational Behavior and Human Performance, *27*, 326-344.
- Saal, F. E., & Knight, P. A. (1988). Industrial/organizational psychology: Science and practice (pp. 121-125). Pacific Grove, CA: Brooks/Cole.

- Schmitt, N., & Noe, R. A. (1986). Personnel selection and equal employment opportunity. In C. L. Cooper & I. T. Robertson (Eds.), International Review of Industrial and Organizational Psychology. New York: John Wiley and Sons.
- Sharon, A., & Bartlett, C. (1969). Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 22, 252-263.
- Shepperd, B. H. (1985). Justice is no simple matter: Case for elaborating our model of procedural justice. Journal of Personality and Social Psychology, 49, 953-962.
- Shrauger, J., & Osberg, T. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. Psychological Bulletin, 90, 322-351.
- Somers, M. J., & Birnbaum, D. (1991). Assessing self-appraisal of job performance as an evaluation device: Are the poor results a function of method or methodology? Human Relations, 44, 1081-1091.
- Suls, J. M., & Wills, T. A. (1991). Social comparison: Contemporary theory and research. Hillsdale, NJ: Lawrence Erlbaum.
- Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73, 497-506.
- Sundstrom, E., Demeuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. American Psychologist, 45(2), 120-133.
- Thornton, G. (1980). Psychometric properties of self-appraisals of executive performance. Personnel Psychology, 33, 263-271.

- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. Psychological Review, 66, 297-333.
- Williams, W. E., & Seiler, D. A. (1973). Relationship between measures of effort and job performance. Journal of Applied Psychology, 57, 49-54.
- Zedeck, S., & Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of appraisal. Journal of Applied Psychology, 67, 752-758.

APPENDICES

APPENDIX A:
PROOFREADING TEST SAMPLE

Karen Rusa was a 30-year-old married woman and the mother of four children. Although she had been having anxiety-related problems for a number of years, she had never sought professional help prior to this time. During the preceding three months, she had become increasingly depressed; her family physician finally suggested that she seek psychological services.

For the past several months Karen had been experiencing intrusive, repetitive thoughts that centered around her children's safety. She frequently found herself imagining that a serious accident had occurred; she was unable to put these thoughts out of her mind. On one such occasion she imagined that her son, Alan, had broken his leg playing football at school. There was no reason to believe that an accident had occurred, but Karen brooded about the possibility until she finally called the school to see if Alan was all right. Even after receiving their reassurance that he had not been hurt, she described herself as being somewhat surprised when he later arrived home unharmed. Karen also noted that her daily routine was seriously hampered by an extensive series of counting rituals that she performed throughout each day. Specific numbers had come to have a special meaning to Karen; she found that her preoccupation with these numbers was interfering with her ability to perform everyday activities. One example was grocery shopping. Karen believed that if she selected the first item (e.g., a box of cereal) on the shelf, something terrible would happen to her oldest child. If she selected the second item, some unknown disaster would befall her second

APPENDIX B:
PERFORMANCE APPRAISAL FORMS

PERFORMANCE APPRAISAL FORM

Please rate yourself on the following dimensions:

1. Quantity of Proofreading (total number of pages completed)

1	2	3	4	5	6	7	8	9
worse than anyone else		below average		average		above average		better than anyone else

2. Quality of Proofreading (average number of mistakes detected per page)

1	2	3	4	5	6	7	8	9
worse than anyone else		below average		average		above average		better than anyone else

3. Effort (amount of energy directed towards the task)

1	2	3	4	5	6	7	8	9
worse than anyone else		below average		average		above average		better than anyone else

4. Conscientiousness (motivation to do well)

1	2	3	4	5	6	7	8	9
worse than anyone else		below average		average		above average		better than anyone else

Overall Performance Score (sum of first four dimension ratings) _____

In the space provided below, please describe in detail how you came up with your ratings.

Please indicate the degree to which each of the following influenced your performance appraisal:

1. Evaluation of your proofreading ability

1	2	3	4	5
not at all		somewhat influenced		totally influenced

2. Evaluation of other students' proofreading abilities

1	2	3	4	5
not at all		somewhat influenced		totally influenced

3. Potential for extra-credit

1	2	3	4	5
not at all		somewhat influenced		totally influenced

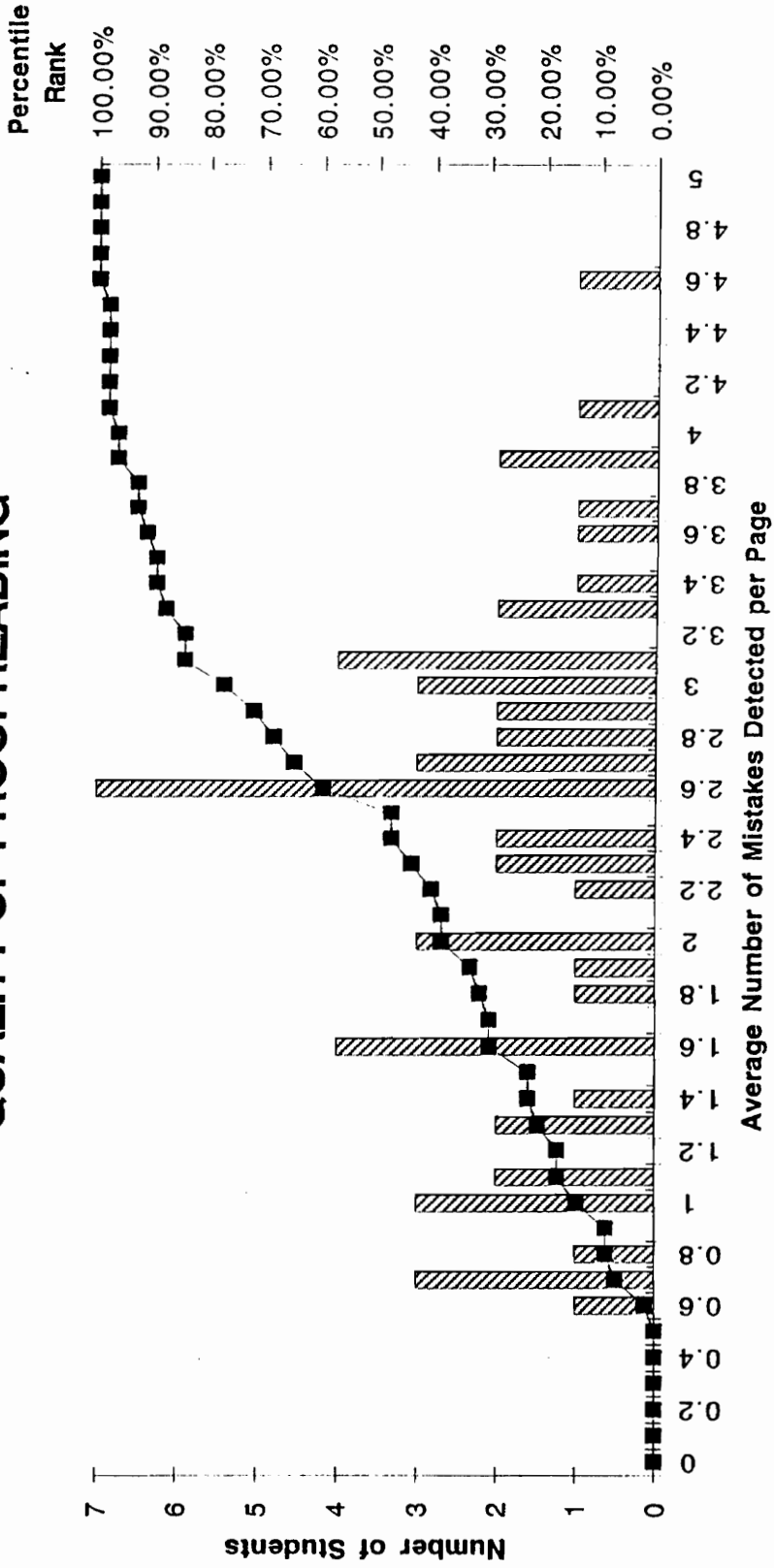
**APPENDIX C:
PILOT STUDY DATA**

QUANTITY (total # of pages completed)	
Sample Size	57
Mean	9.53
Median	9.50
Standard Deviation	2.01
Minimum	5.50
Maximum	16.0

ADJ. QUALITY (avg. (# proper - # improper corrections per page))	
Sample Size	57
Mean	2.33
Median	2.60
Standard Deviation	0.97
Minimum	0.60
Maximum	4.60

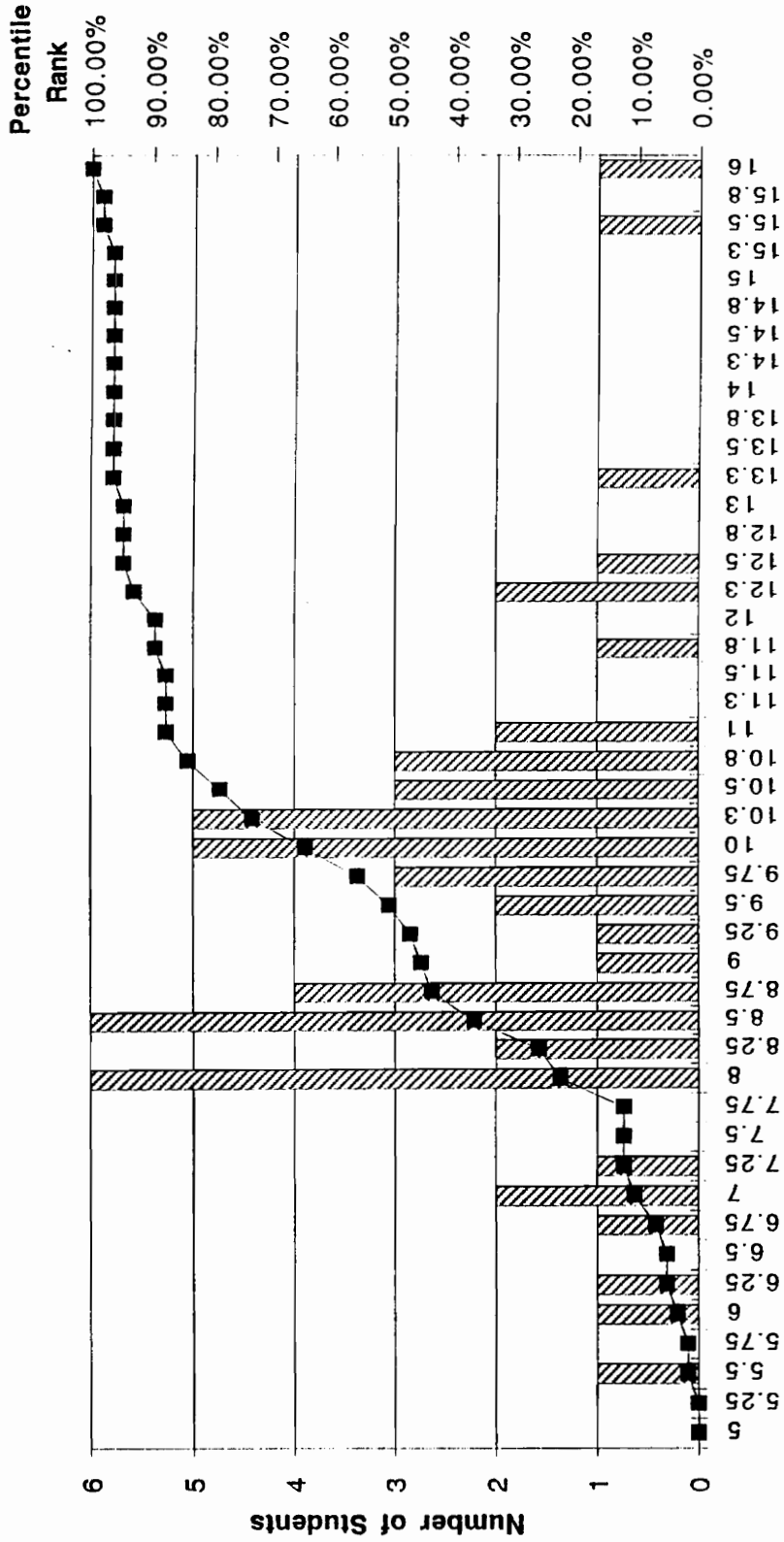
APPENDIX D:
NORMATIVE DATA SHEETS

QUALITY OF PROOFREADING



Average Performance Score = 2.3 mistakes

QUANTITY OF PROOFREADING



Total Number of Pages Completed

Average Performance Score = 9.5 pages

APPENDIX E:
EXPERIMENTAL PROTOCOL: CONTROL GROUP

CONTROL GROUP

Hello, my name is _____. Thank you for agreeing to participate in this study. The purpose of this study is to research human perceptual performance. In order to collect the necessary data, you will be asked to perform a short proofreading task.

I will begin this session by asking you to complete an extra-credit form and a consent form required by the University for any project utilizing college students. Next, you will be asked to perform a 15 minute proofreading task. You will then take a 5 minute break. Finally, you will be asked to complete a short questionnaire regarding your performance.

HAND OUT EXTRA CREDIT AND CONSENT FORMS
WAIT FOR SUBJECT TO COMPLETE, AND COLLECT FORMS
HAND OUT TEST (FACE DOWN) AND RED PEN
REVIEW DIRECTIONS

The directions for the proofreading task are as follows. Do not turn the test over until I say "Begin". You are to circle all misspelled words with the red pen. Do **NOT** correct the mistakes. Do not circle proper names. When I give the signal to stop, put an X next to the last line you read. Are there any questions? You may begin.

COLLECT TEST IN 15 MINUTES
MAKE SURE SUBJECT MARKED LAST SENTENCE READ

I will now ask you to relax for 5 minutes. At the end of this period you will be asked to complete a performance appraisal form.

AFTER 5 MINUTES, HAND OUT PERFORMANCE APPRAISAL FORM
 Please complete this performance appraisal form.

IF IN REWARD CONDITION:

The top three performers will receive an additional extra credit point towards their psychology grade.

IF IN PERFORMANCE FEEDBACK CONDITION:

A research team will be analyzing your self-appraisals in order to generate feedback regarding your perceptual performance.

BOTH CONDITIONS:

If you have any questions while completing the form, don't hesitate to ask them. You will have 15 minutes to complete the form. You may begin.

COLLECT FORM WHEN SUBJECT IS DONE OR AFTER 15 MINUTES
MAKE SURE PAGES 1 AND 3 ARE COMPLETE
ADMINISTER DEBRIEFING
DISMISS SUBJECTS

APPENDIX F:
EXPERIMENTAL PROTOCOL:
BEHAVIORAL OBSERVATION INFERENCE GROUP

BEHAVIORAL OBSERVATION INFERENCE GROUP

Hello, my name is _____. Thank you for agreeing to participate in this study. The purpose of this study is to research human perceptual performance. In order to collect the necessary data, you will be asked to perform a short proofreading task.

I will begin this session by asking you to complete an extra-credit form and a consent form required by the University for any project utilizing college students. Next, you will be asked to perform a 15 minute proofreading task. You will then take a 5 minute break. Finally, you will be asked to complete a short questionnaire regarding your performance.

HAND OUT EXTRA CREDIT AND CONSENT FORMS
WAIT FOR SUBJECT TO COMPLETE, AND COLLECT FORMS
HAND OUT TEST (FACE DOWN) AND RED PEN
REVIEW DIRECTIONS

The directions for the proofreading task are as follows. Do not turn the test over until I say "Begin". You are to circle all misspelled words with the red pen. Do **NOT** correct the mistakes. Do not circle proper names. When I give the signal to stop, put an X next to the last line you read. Are there any questions? You may begin.

COLLECT TEST IN 15 MINUTES
MAKE SURE SUBJECT MARKED LAST SENTENCE READ

I will now ask you to relax for 5 minutes. During this five minute period I would like you to watch videos of other students participating in the study. These students have been proofreading for 10 minutes. You are to watch them complete the task for the next five minutes. Do you have any questions? At the end of this period you will be asked to complete a performance appraisal form.

RANDOMLY INSERT VIDEOS IN VCRs AND PUSH PLAY
AFTER 5 MINUTES, HAND OUT PERFORMANCE APPRAISAL FORM
 Please complete this performance appraisal form.

IF IN REWARD CONDITION:

The top three performers will receive an additional extra credit point towards their psychology grade.

IF IN PERFORMANCE FEEDBACK CONDITION:

A research team will be analyzing your self-appraisals in order to generate feedback regarding your perceptual performance.

BOTH CONDITIONS:

If you have any questions while completing the form, don't hesitate to ask them. You will have 15 minutes to complete the form. You may begin.

COLLECT FORM WHEN SUBJECT IS DONE OR AFTER 15 MINUTES
MAKE SURE PAGES 1 AND 3 ARE COMPLETE
ADMINISTER DEBRIEFING
DISMISS SUBJECTS

APPENDIX G:
EXPERIMENTAL PROTOCOL:
STRUCTURED INFORMATION INFERENCE GROUP

STRUCTURED INFORMATION INFERENCE GROUP

Hello, my name is _____. Thank you for agreeing to participate in this study. The purpose of this study is to research human perceptual performance. In order to collect the necessary data, you will be asked to perform a short proofreading task.

I will begin this session by asking you to complete an extra-credit form and a consent form required by the University for any project utilizing college students. Next, you will be asked to perform a 15 minute proofreading task. You will then take a 5 minute break. Finally, you will be asked to complete a short questionnaire regarding your performance.

HAND OUT EXTRA CREDIT AND CONSENT FORMS
WAIT FOR SUBJECT TO COMPLETE, AND COLLECT FORMS
HAND OUT TEST (FACE DOWN) AND RED PEN
REVIEW DIRECTIONS

The directions for the proofreading task are as follows. Do not turn the test over until I say "Begin". You are to circle all misspelled words with the red pen. Do **NOT** correct the mistakes. Do not circle proper names. When I give the signal to stop, put an X next to the last line you read. Are there any questions? You may begin.

COLLECT TEST IN 15 MINUTES
MAKE SURE SUBJECT MARKED LAST SENTENCE READ

I will now ask you to relax for 5 minutes. During this time I would like you to examine these two graphs.

HAND OUT GRAPHS

These graphs represent the performances of other student in the study. These graphs provide three types of information regarding the Quantity and Quality of other students' performances. The shaded bars represent how many other students performed at a given level. For example, as can be seen on the Quantity graph, 5 students proofread 10 pages.

POINT OUT ON GRAPH

The black boxes represent percentile rank. For example, as can be seen on the Quantity graph, proofreading 10 pages is equivalent to the 82nd percentile. This means 82% of students proofread 10 pages or less.

POINT OUT ON GRAPH

Finally, the average performance score is given at the bottom of the sheet. Do you have any questions? Please do not write on the graphs. At the end of this period you will be asked to complete a performance appraisal form.

AFTER 5 MINUTES, HAND OUT PERFORMANCE APPRAISAL FORM

Please complete this performance appraisal form.

IF IN REWARD CONDITION:

The top three performers will receive an additional extra credit point towards their psychology grade.

IF IN PERFORMANCE FEEDBACK CONDITION:

A research team will be analyzing your self-appraisals in order to generate feedback regarding your perceptual performance.

BOTH CONDITIONS:

If you have any questions while completing the form, don't hesitate to ask them. You will have 15 minutes to complete the form. You may begin.

COLLECT FORM WHEN SUBJECT IS DONE OR AFTER 15 MINUTES

MAKE SURE PAGES 1 AND 3 ARE COMPLETE

ADMINISTER DEBRIEFING

DISMISS SUBJECTS

APPENDIX H:
CONSENT FORM

INFORMED CONSENT FORM

TITLE OF EXPERIMENT: Perceptual Performance

EXPERIMENT # _____ .

1. PURPOSE OF EXPERIMENT:

You are invited to participate in a study about perceptual performance. This study involves experimentation for the purpose of increasing knowledge in the area of perception.

2. PROCEDURES TO BE FOLLOWED IN THE STUDY:

To accomplish the goals of the study, you will be asked to perform a short proofreading task. After completing this task you will be asked to complete a short questionnaire. The experiment will last less than one hour.

3. ANONYMITY OF SUBJECTS AND CONFIDENTIALITY OF RESULTS:

The results of this study will be kept strictly confidential. At no time will the researchers release your results to anyone without your written consent. The information you provide will have your name removed and only a subject number will identify you during analyses and any write-up of the research.

4. DISCOMFORTS AND RISKS FROM PARTICIPATING IN THE STUDY:

There are no apparent risks to you from participating in this study.

5. EXPECTED BENEFITS:

Your participation in the project will provide information that may be helpful for increasing our understanding of perceptual performance. Further benefits will be explained at the end of the study.

6. FREEDOM TO WITHDRAW:

You are free to withdraw from participation in this study at any time without penalty.

7. EXTRA CREDIT:

For participating in the study you will receive 1 extra credit point.

8. USE OF RESEARCH DATA:

The information from this research may be used for scientific or educational purposes. It may be presented at scientific meetings and/or published and reproduced in professional journals or books, or used for any other purpose that Virginia Tech's Department of Psychology considers proper in the interest of education, knowledge, or research.

9. APPROVAL OF RESEARCH:

This research project has been approved by the Human Subjects Committee of the Department of Psychology and by the Institutional Review Board of Virginia Tech.

10. SUBJECT'S PERMISSION:

I have read and understand the above descriptions of the study. I have had an opportunity to ask questions and have had them all answered. I hereby acknowledge the above and give my voluntary consent for participating in the study.

I further understand that if I participate I may withdraw at any time without penalty.

I understand that should I have any questions regarding this research and its conduct, I should contact any of the persons named below.

PRIMARY RESEARCHER: Steven C. Morgan

PHONE: 231-6279

FACULTY ADVISOR: Neil M. A. Hauenstein

PHONE: 231-5716

CHAIR, HSC: Robert J. Harvey

PHONE: 231-7030

CHAIR, IRB: Ernest Stoudt

PHONE: 231-6077

SUBJECT'S SIGNATURE: _____ DATE: _____

**APPENDIX I:
CONFIRMATION OF DEBRIEFING**

CONFIRMATION OF DEBRIEFING

There was more to this study than you were told at the beginning. We do not tell participants the full purpose at the beginning because this might affect the way people act, and would not be an indication of everyday life.

The purpose of the study was not to research perceptual performance. The true purpose of the study was to study the effects of contextual factors on the accuracy and leniency of self-appraisals of performance. One of the contextual factors examined was the availability of social comparison information. Basically, the experimenters were interested in whether or not subjects compared their performance to the performance of other subjects when rating performance. Another contextual factor examined was purpose of appraisal. In order to manipulate purpose of appraisal, some subjects were told that they would be receiving feedback regarding their performance. Other subjects were told that the top three performers would receive an additional extra credit point. Indeed, the top three performers among **all** participants will receive one additional extra credit point towards their Intro. to Psychology grade. Furthermore, **all** participants will have access to feedback regarding their performance at the conclusion of the study.

PLEASE READ THE FOLLOWING AND SIGN BELOW:

I understand that the purpose of the study was to study the effects of contextual factors on the accuracy and leniency of self-appraisals of performance. The experimenters were interested in whether the availability of social comparison information and the purpose of appraisal had any effects on the accuracy and leniency of self-appraisals of performance.

Most importantly, I realize that I will be notified at the conclusion of the study if I received an additional extra credit point. Furthermore, if I am interested in feedback regarding my performance, I realize that I may contact the experimenter and he will provide me with the information. This information will be available in about 10 weeks. Also, my identity will remain strictly confidential. All data is coded and analyzed by number.

I understand that any discussion of this research with other students in my class could lead to contamination of the results of this research, and I agree to refrain from discussing this research for at least 3 months.

SIGNED _____

DATE _____

APPENDIX J:
OBJECTIVE SCORE DERIVATION

OBJECTIVE SCORE DERIVATION

OBJECTIVE PERFORMANCE MEASURE

TRUE SCORE (Stanine)

Quantity (# of pages completed)

0 - 6	1
6.25 - 7	2
7.25 - 8	3
8.25 - 8.5	4
8.75 - 9.75	5
10 - 10.25	6
10.5 - 12	7
12.25 - 15.25	8
15.5 +	9

Adj. Quality (avg. (# proper - # improper corrections per page))

0 - .6	1
.7 - .9	2
1 - 1.5	3
1.6 - 2.2	4
2.3 - 2.6	5
2.7 - 3	6
3.1 - 3.5	7
3.6 - 4	8
4.1 +	9

**APPENDIX K:
SUMMARY TABLES FOR MANIPULATION CHECK ITEMS**

Table K-1

ANOVA Table for Purpose of Appraisal Manipulation Check: Subjects' Ratings of "Potential for extra-credit"

Source	DF	Mean Square	F-Ratio
Subject Sex (S)	1	.02	.02
Purpose of Appraisal (POA)	1	82.7	77.06 *
Social Comparison Info (SCI)	2	1.27	1.18
S X POA	1	.80	.75
S X SCI	2	1.49	1.39
POA X SCI	2	2.96	2.75
S X POA X SCI	2	.00	.00
Residual	168	1.07	

* $p < .01$

Table K-2

ANOVA Table for Ability Checks: Objective Performance Scores for Quantity and Quality

Source	DF	T1		T2	
		Mean Square	F-Ratio	Mean Square	F-Ratio
Subject Sex (S)	1	15.61	3.09*	16.2	6.13 **
Purpose of Appraisal (POA)	1	4.67	.93	4.36	1.65
Social Comparison Info (SCI)	2	2.27	.45	6.44	2.44
S X POA	1	.67	.13	3.76	1.42
S X SCI	2	6.44	1.27	2.22	.84
POA X SCI	2	5.87	1.16	2.37	.90
S X POA X SCI	2	3.91	.77	2.84	1.07
Residual	168	5.05		2.64	

Note. T1 = Objective Quantity Score; T2 = Objective Quality Score

* $p = .08$

** $p < .05$

APPENDIX L:
SUMMARY TABLES FOR DEPENDENT VARIABLES

Table L-1

ANCOVA Table for Self-Ratings of Performance: Quantity

Source	DF	Mean Square	F-Ratio
<u>Covariate</u>			
Quantity Score	1	125.57	101.06 **
<u>Between Subjects Effects</u>			
Subject Sex (S)	1	.00	.00
Purpose of Appraisal (POA)	1	6.49	5.22*
Social Comparison Info (SCI)	2	10.72	8.63 **
S X POA	1	1.28	1.03
S X SCI	2	1.66	1.34
POA X SCI	2	.48	.68
S X POA X SCI	2	4.93	3.97 *
Residual	167	1.24	

* $p < .05$ ** $p < .001$

REWARD CONDITION

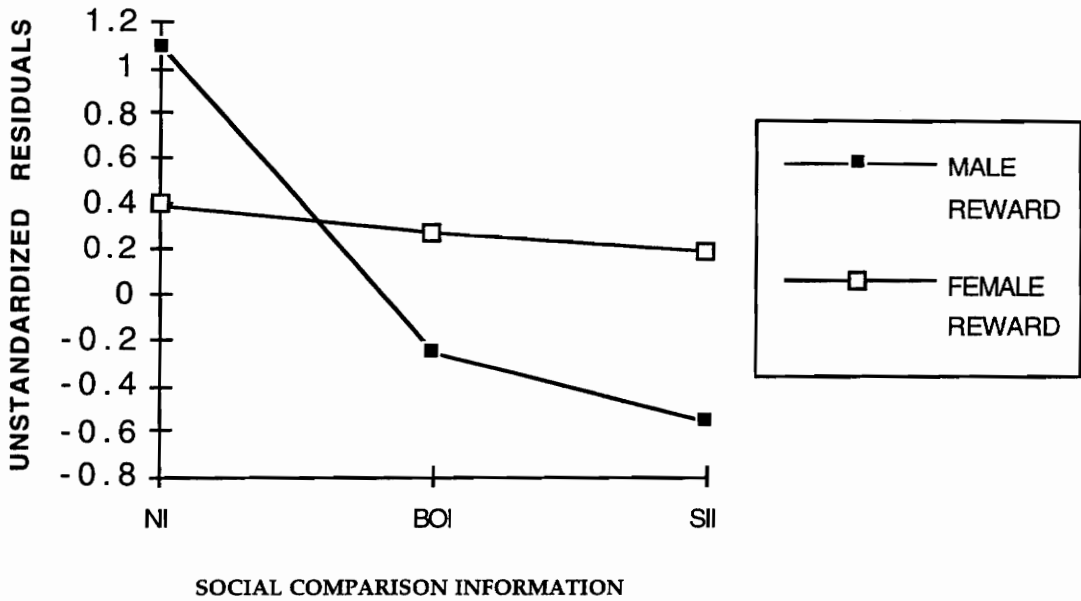


Figure L-2. Plot of unstandardized residuals of Quantity self-ratings as a function of sex and social comparison information. Unstandardized residuals reflect the mean variance unaccounted for using Quantity scores to predict Quantity self-ratings. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information Inference.

BEHAVIORAL OBSERVATION INFERENCE CONDITION

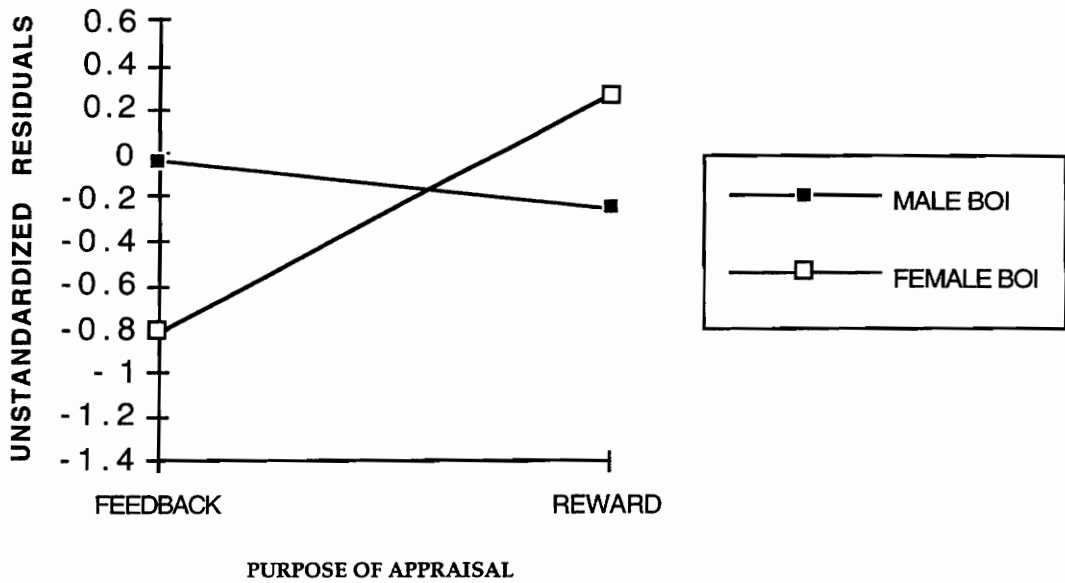


Figure L-3. Plot of unstandardized residuals of Quantity self-ratings as a function of sex and purpose of appraisal. Unstandardized residuals reflect the mean variance unaccounted for using Quantity scores to predict Quantity self-ratings. BOI = Behavioral Observation Inference.

Table L-4

ANCOVA Table for Self-Ratings of Performance: Quality

Source	DF	Mean Square	F-Ratio
<u>Covariate</u>			
Quality Score	1	87.45	58.20 **
<u>Between Subjects Effects</u>			
Subject Sex (S)	1	.07	.05
Purpose of Appraisal (POA)	1	8.06	5.37*
Social Comparison Info (SCI)	2	18.41	12.25 **
S X POA	1	.15	.10
S X SCI	2	.04	.43
POA X SCI	2	3.95	2.63
S X POA X SCI	2	.61	.41
Residual	167	1.50	

* $p < .05$ ** $p < .001$

Table L-5

ANCOVA Table for Self-Ratings of Performance: Effort

Source	DF	Mean Square	F-Ratio
<u>Covariate</u>			
Quantity Score	1	.56	.40
<u>Between Subjects Effects</u>			
Subject Sex (S)	1	.54	.38
Purpose of Appraisal (POA)	1	5.14	3.69 *
Social Comparison Info (SCI)	2	3.37	2.41
S X POA	1	.44	.32
S X SCI	2	2.75	1.96
POA X SCI	2	.30	.22
S X POA X SCI	2	2.09	1.50
Residual	167	1.40	

* $p = .056$

Table L-6

ANCOVA Table for Self-Ratings of Performance: Conscientiousness

Source	DF	Mean Square	F-Ratio
<u>Covariate</u>			
Quality Score	1	13.06	12.17 **
<u>Between Subjects Effects</u>			
Subject Sex (S)	1	.01	.01
Purpose of Appraisal (POA)	1	4.89	4.56 *
Social Comparison Info (SCI)	2	9.70	9.03 ***
S X POA	1	1.48	1.38
S X SCI	2	4.39	4.10 *
POA X SCI	2	.94	.88
S X POA X SCI	2	1.87	1.74
Residual	167	1.30	

* $p < .05$ ** $p < .01$ *** $p < .001$

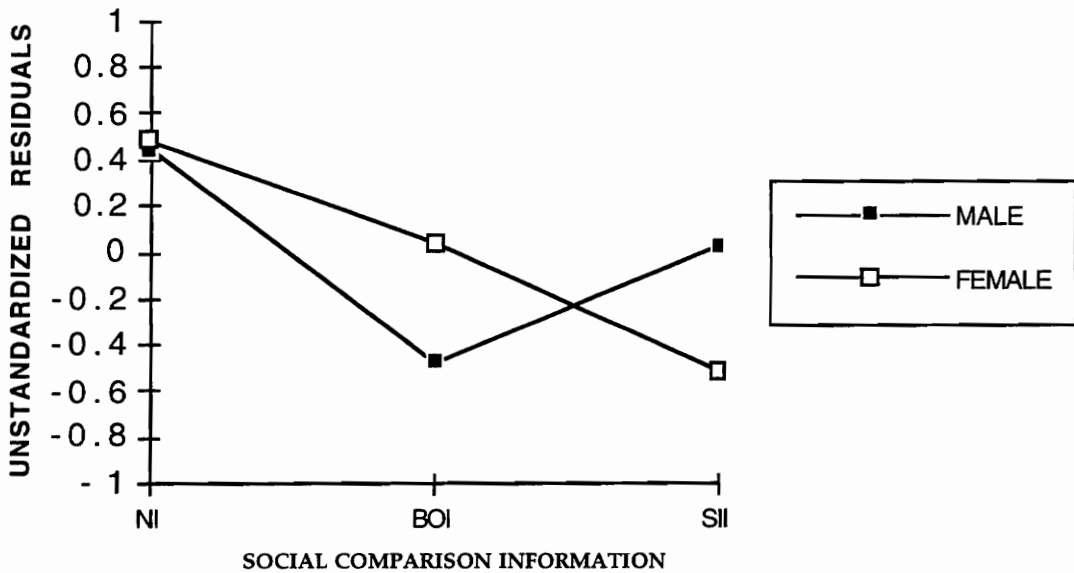


Figure L-7. Plot of unstandardized residuals of Conscientiousness self-ratings as a function of sex and social comparison information. Unstandardized residuals reflect the mean variance unaccounted for using Quality scores to predict Conscientiousness self-ratings.. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information Inference.

Table L-8

ANCOVA Table for Self-Ratings of Performance: Overall Performance Score

Source	DF	Mean Square	F-Ratio
<u>Covariates</u>			
Quantity Score	1	111.40	30.01 *
Quality Score	1	227.64	13.62 *
<u>Between Subjects Effects</u>			
Subject Sex (S)	1	.36	.05
Purpose of Appraisal (POA)	1	83.15	10.96*
Social Comparison Info (SCI)	2	162.36	21.40 *
S X POA	1	.38	.05
S X SCI	2	10.06	1.33
POA X SCI	2	15.82	2.09
S X POA X SCI	2	11.58	1.53
Residual	166	7.56	

* $p < .001$

Table L-9

Descriptive Statistics for Self-Ratings broken down by Social Comparison Information

Dimension	<u>Social Comparison Information</u>		
	NI	BOI	SII
Quantity			
M	5.97	5.37	5.35
SD	1.30	1.36	1.60
Quality			
M	6.03	5.27	5.38
SD	1.50	1.40	1.46
Effort			
M	6.48	6.35	6.02
SD	1.24	1.07	1.24
Conscientiousness			
M	7.02	6.43	6.40
SD	1.00	1.18	1.14
Overall Performance Score			
M	25.50	23.25	23.10
SD	2.88	3.55	3.45

Note. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information Inference.

Table L-10

Descriptive Statistics for Self-Ratings broken down by Purpose Of Appraisal

Dimension	<u>Purpose of Appraisal</u>	
	Reward	Performance Feedback
Quantity		
M	5.69	5.43
SD	1.57	1.32
Quality		
M	5.84	5.28
SD	1.56	1.35
Effort		
M	6.46	6.11
SD	1.29	1.08
Conscientiousness		
M	6.81	6.42
SD	1.12	1.13
Overall Performance Score		
M	24.69	23.21
SD	3.74	3.02

Table L-11

ANOVA Table for Dimension Self-Ratings

Source	DF	"Quantity" <u>Dimensions</u>		"Quality" <u>Dimensions</u>	
		Mean Square	F-Ratio	Mean Square	F-Ratio
<u>Between Subjects Effects</u>					
Subject Sex (S)	1	2.50	1.46	2.50	1.33
Purpose of Appraisal (POA)	1	8.10	4.73 *	20.54	10.92 ***
Social Comparison Info (SCI)	2	9.17	5.35 *	17.17	9.12 ***
S X POA	1	1.11	.65	.04	.02
S X SCI	2	.26	.15	.97	.52
POA X SCI	2	1.41	.82	1.39	.74
S X POA X SCI	2	5.25	3.06 *	2.94	1.56
Residual	168	1.71		1.88	
<u>Within Subjects Effects</u>					
Dimension (D)	1	46.94	29.15 ***	100.28	73.08 ***
S X D	1	.40	.25	.40	.29
POA X D	1	.18	.11	.71	.52
SCI X D	2	1.70	1.06	.29	.21
S X POA X D	1	.01	.01	.90	.66
S X SCI X D	2	7.76	4.82 **	2.86	2.08
POA X SCI X D	2	.22	.14	.97	.71
S X POA X SCI X D	2	5.25	3.26 *	.97	.71
Residual	168	1.61		1.37	

Note. "Quantity" dimensions reflect Quantity vs. Effort self-ratings;
 "Quality" dimensions reflect Quality vs. Conscientiousness self-ratings.

* $p < .05$

** $p < .01$

*** $p < .001$

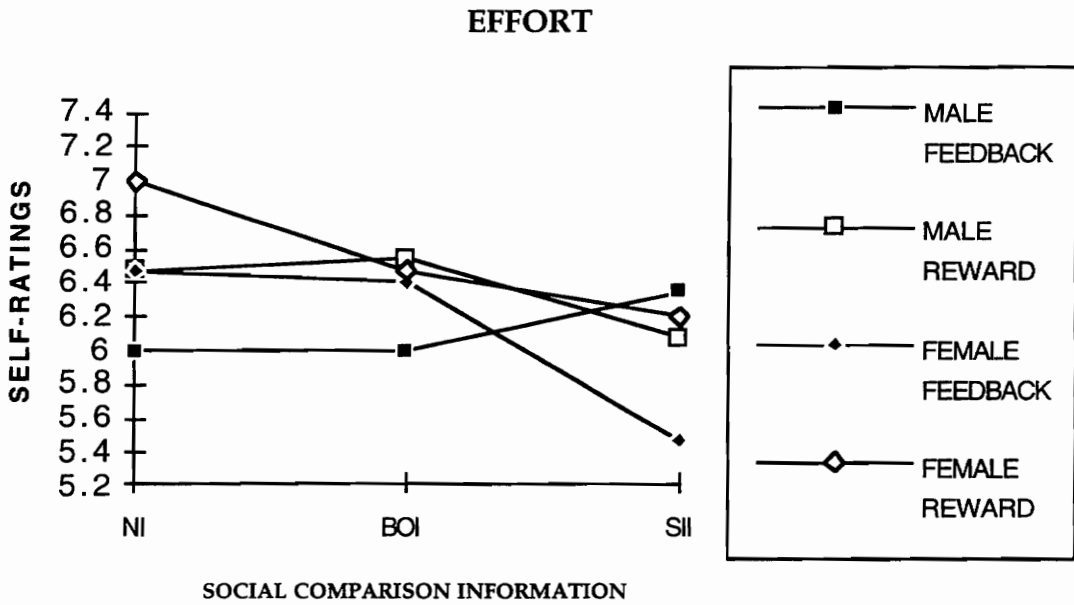
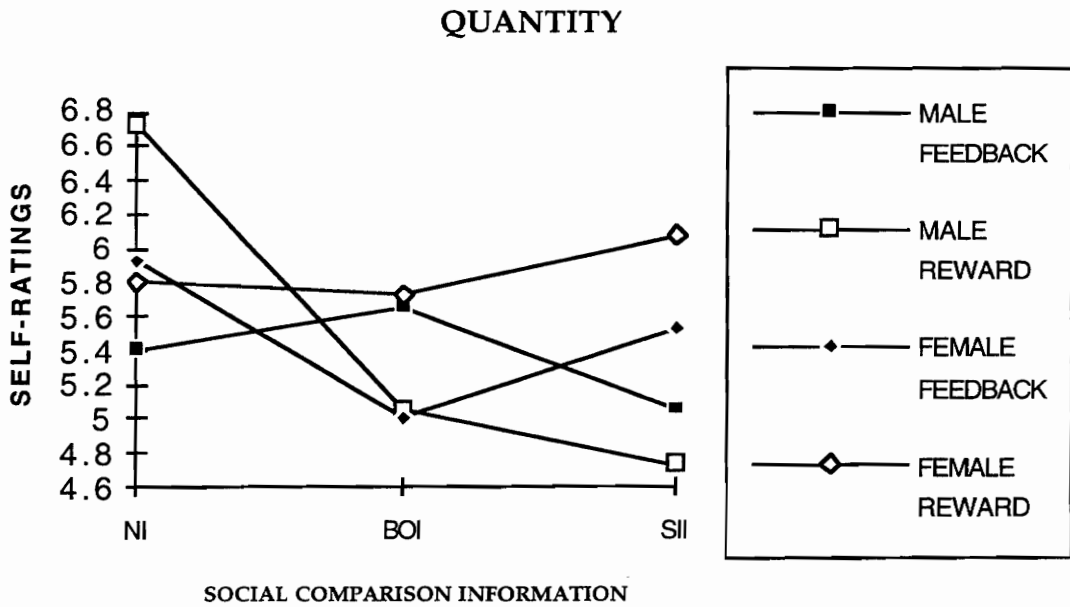


Figure L-12. Plots of self-ratings as a function of sex, purpose of appraisal, and social comparison information. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information Inference.

Table L-13

ANOVA Table for Accuracy Measures

Source	DF	E		DA	
		Mean Square	F-Ratio	Mean Square	F-Ratio
Subject Sex (S)	1	2.57	3.74 *	.40	.75
Purpose of Appraisal (POA)	1	8.24	12.00 ***	.50	.93
Social Comparison Info (SCI)	2	18.76	27.34 ***	2.49	4.62 **
S X POA	1	.61	.89	2.57	4.77 **
S X SCI	2	2.40	3.50 **	.51	.94
POA X SCI	2	1.17	1.70	.51	.94
S X POA X SCI	2	.35	.51	.51	.94
Residual	168	.69		.54	

Note. E = Elevation; DA = Dimensional Accuracy.

* $p = .055$

** $p < .05$

*** $p < .001$

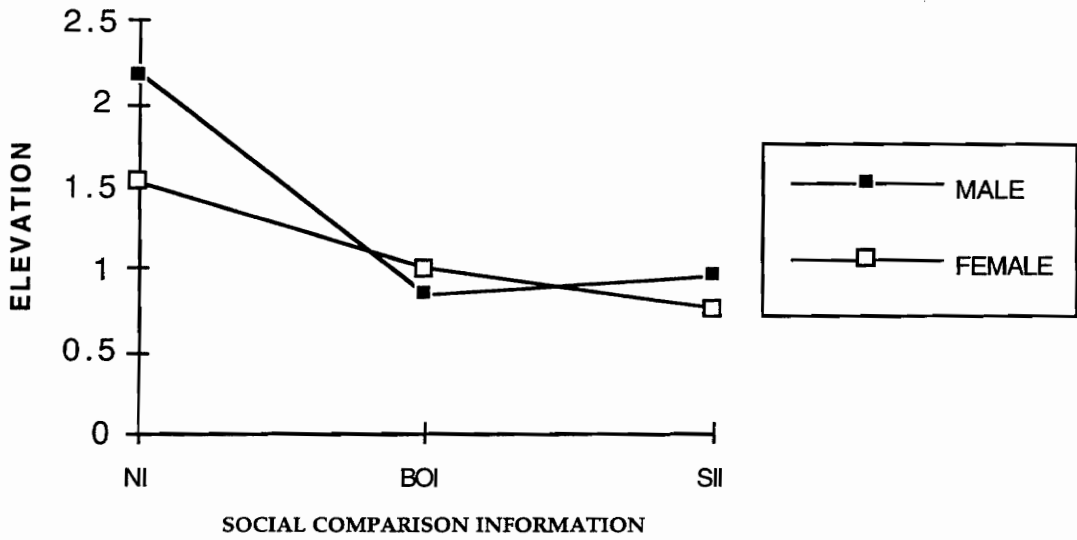


Figure L-14. Plot of Elevation accuracy measure as a function of sex and social comparison information. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information inference. Note: smaller values reflect greater accuracy.

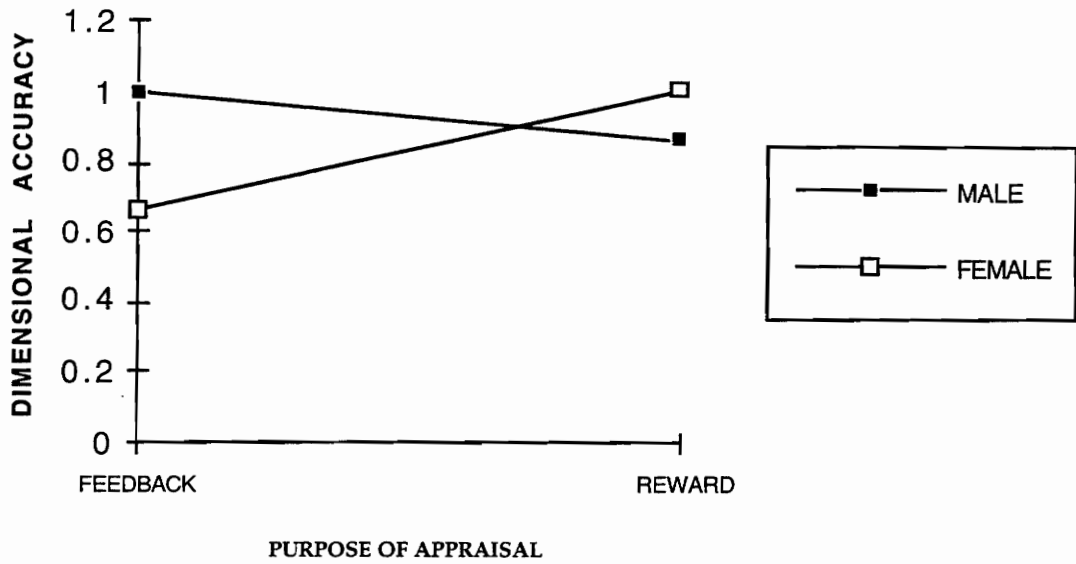


Figure L-15. Plot of Dimensional Accuracy measure as a function of sex and purpose of appraisal. Note: smaller values reflect greater accuracy.

Table L-16

Descriptive Statistics for Accuracy Measures broken down by Social Comparison Information

Accuracy Measure	<u>Social Comparison Information</u>		
	NI	BOI	SII
Elevation			
M	1.85	.93	.86
SD	1.20	.66	.63
Dimensional Accuracy			
M	1.11	.84	.71
SD	.96	.63	.56

Note. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information inference. Smaller values reflect greater accuracy

Table L-17

Descriptive Statistics for Accuracy Measures broken down by Purpose Of Appraisal

Accuracy Measure	<u>Purpose of Appraisal</u>	
	Reward	Performance Feedback
Elevation		
M	1.43	1.00
SD	1.02	.90
Dimensional Accuracy		
M	.94	.83
SD	.80	.70

Note. Smaller values reflect greater accuracy.

Table L-18

ANOVA Table for Utilization of SCI Measures: Number of Comparative Remarks and Percentage of Comparative Remarks

Source	DF	NCR		PCR	
		Mean Square	F-Ratio	Mean Square	F-Ratio
Subject Sex (S)	1	.80	.52	2586.22	4.12 *
Purpose of Appraisal (POA)	1	.36	.23	428.71	.68
Social Comparison Info (SCI)	2	17.07	11.10 **	1225.28	19.50 **
S X POA	1	5.00	3.25	27.73	.04
S X SCI	2	2.15	1.40	102.93	.16
POA X SCI	2	.74	.48	197.56	.31
S X POA X SCI	2	1.55	1.01	263.44	.42
Residual	168	1.54		627.17	

Note. NCR = Number of Comparative Remarks; PCR = Percentage of Comparative Remarks.

* $p < .05$

** $p < .001$

Table L-19

ANOVA Table for Utilization of SCI Measures: Subjects' Ratings of
"Evaluation of other students' proofreading abilities"

Source	DF	Mean Square	F-Ratio
Subject Sex (S)	1	.00	.00
Purpose of Appraisal (POA)	1	4.36	5.25 *
Social Comparison Info (SCI)	2	38.12	45.96 **
S X POA	1	1.09	1.31
S X SCI	2	1.12	1.35
POA X SCI	2	.11	.13
S X POA X SCI	2	1.37	1.66
Residual	168	.83	

* $p < .05$

** $p < .001$

Table L-20

Descriptive Statistics for Utilization of SCI Measures broken down by Social Comparison Information

SCI Measure	<u>Social Comparison Information</u>		
	NI	BOI	SII
NCR			
M	.65	1.20	1.72
SD	.90	1.39	1.38
PCR			
M	10.24	21.24	38.55
SD	15.60	24.35	31.94
Evaluation of Others			
M	1.67	2.85	3.18
SD	.73	.94	1.07

Note. NI = No-Information; BOI = Behavioral Observation Inference; SII = Structured Information inference; NCR = Number of Comparative Remarks; PCR = Percentage of Comparative Remarks.

Table L-21

Descriptive Statistics for Utilization of SCI Measures broken down by Purpose
Of Appraisal

SCI Measure	<u>Purpose of Appraisal</u>	
	Reward	Performance Feedback
Evaluation of Others		
M	2.41	2.72
SD	1.16	1.07

VITA

STEVEN C. MORGAN

PERSONAL INFORMATION

Date of Birth: April 4, 1969

Marital Status: Married

Business Address: Academic Assessment Program
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061-0157

Business Phone: (703) 231-4581

Business FAX: (703) 231-4522

Home Address: 247 Fairfax Road
Blacksburg, VA 24060

Home Phone: (703) 552-7428

EDUCATION

B. S. The University of Pittsburgh, distinction, 1992
Major field of study: Psychology

Minor fields of study: Sociology and Communications

M. S. Virginia Polytechnic Institute and State University, 1994
Major field of study: Industrial/Organizational Psychology

**THESIS: Effects of Context on the Leniency, Accuracy, and
Utility of Self-Appraisals of Performance: Social Comparison
Information and Purpose of Appraisal.**

Major advisor: Dr. Neil M. A. Hauenstein

ACADEMIC EXPERIENCE

1993 - 1994 Lab Coordinator - Department of Psychology, VPI.
Responsible for maintaining two computer labs,
coordinating lab scheduling, loading software, offering
technical support as needed.

1992 - 1993 Graduate Lab Instructor - Department of Psychology, VPI.
Responsible for teaching two Introductory to Psychology
undergraduate labs. Designed and conducted lectures,
designed and graded exams and essays.

APPLIED EXPERIENCE

- 1994 - present Research Analyst - Academic Assessment Program, VPI. Responsible for evaluating teaching/learning initiatives and other University-implemented projects. Responsible for disseminating the findings of research projects via campus-wide newsletter. Analyzed and presented research findings to department and university representatives. Served as computer network liason and trainer.
- 1994 Consultant - The Childrens' Nest Daycare Center, Blacksburg, VA. Conducted job analyses and incumbent interviews. Wrote comprehensive job descriptions and performed other miscellaneous human resource functions as requested.
- 1993 Research Analyst - Outcome Assessment Project, Department of Psychology, VPI. Responsible for collecting, classifying and analyzing data regarding career paths of Psychology graduates.
- 1991 - 1992 Intern - Office of Human Resources, University of Pittsburgh. Responsible for designing performance appraisal system. Designed and implemented Walk-In screening procedure. Wrote job descriptions for the university. Processed applications and conducted interviews. Represented the university at career fairs.

RELEVANT GRADUATE COURSES

Personnel	Organizational Staffing
Organizational Psychology	Work and Motivation
Research Methods	Psychological Measurement
Statistics for Social Science I, II	GTA Training Workshop
Quantitative Topics	Social Psychology
Psychometrics	Multiple Regression
Industrial Psychology	

PROFESSIONAL AFFILIATIONS

American Psychological Association
Society of Industrial/Organizational Psychology

Psychological Society

COMPUTER SKILLS

Experience using SPSS, SAS, and JMP statistical packages. Experience using Word, Excel, Pagemaker, and Power Point in both IBM and Macintosh environments. Familiar with MS-DOS, Windows, and Macintosh systems.

A handwritten signature in black ink, appearing to read "A. C. M. J.", with a long horizontal line extending to the right.