

The Influence of Discrepant Perceptions of Performance and Amount of Performance
Information on Future Performance Judgments

by

Paige P. Porter

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State

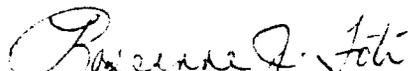
University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

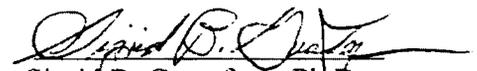
IN

PSYCHOLOGY

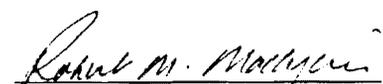
APPROVED:


Roseanne J. Foti, Ph.D., Chair


Neil Hauenstein, Ph.D.


Sigrid B. Gustafson, Ph.D.


Danny Axson, Ph.D.


Robert Madigan, Ph.D.

May 15, 1996

Blacksburg, Virginia

Keywords: Rating Discrepancy, Amount of Performance Information, Self-appraisals,
Performance Appraisal, 360-degree Feedback

THE INFLUENCE OF DISCREPANT PERCEPTIONS OF PERFORMANCE AND
AMOUNT OF PERFORMANCE INFORMATION ON FUTURE PERFORMANCE
JUDGMENTS

by

Paige P. Porter

Roseanne J. Foti, Chairperson

Psychology

(ABSTRACT)

The purpose of the present study was to investigate the effect of amount of performance information (API) and discrepant self-appraisal information on the modification of supervisor's original performance ratings. Confidence in original ratings was expected to mediate the relationship between API and modification of ratings after begin presented with discrepant self-appraisal information. The results indicated that API was not related to modification of original ratings. Moreover, API was unrelated to confidence in original ratings. However, as expected there was a negative relationship between confidence and modification of ratings, although the correlation was not significant. Additional analyses indicated that supervisors in the low API condition became more confident in their final performance ratings after viewing self-ratings than those in the high API condition. The implications of the findings of this study are discussed.

ACKNOWLEDGMENTS

I have several people to thank for contributing to the successful completion of this project. First, I would like to give my most sincere thanks to my advisor and friend, Dr. Roseanne Foti. Her support and guidance allowed me to make this a quality project and to complete it in a timely manner. I also want to extend thanks to the other members of my dissertation committee, Dr. Sigrid Gustafson, Dr. Neil Hauenstein, Dr. Danny Axsom, and Dr. Robert Madigan. Each of their suggestions and input have contributed to the success of this endeavor.

My sincerest gratitude also goes out to the Center for Creative Leadership for funding this research and for allowing the use of their “Looking Glass, Inc.” simulation. I owe the success of the execution of the Looking Glass simulation to my dedicated research assistants, Kevin Basik, Adina Hertzberg, and Hue Brombeck. My best friends Dan Russell and Anne Holloway also lended a hand in the execution of the simulations and, more importantly, were tirelessly supportive of me throughout the duration of this project. I am sincerely grateful to both of them.

I could not imagine completing this project with out the help, encouragement, and unwavering support of my parents, Paul and Diane Porter. They have been a constant source of inspiration during the trials and successes of this process as well as my entire life; for this I owe them my sincerest gratitude.

Finally, I would like to extend my deepest thanks to my best friend and biggest fan, Phil Wolf. His laughter and optimism helped me get through this project on a daily basis. His light-hearted and jovial attitude helped me keep this project in perspective and I will always be grateful to him for his support during this time.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
<u>Chapter</u>	<u>Page</u>
INTRODUCTION.....	1
LITERATURE REVIEW.....	7
Self-Appraisal Use and Benefits.....	7
Discrepancy.....	10
Causes of Discrepancy.....	13
Effects of Discrepancy.....	17
Amount of Performance Information.....	23
METHOD.....	32
Pilot Study.....	32
Participants.....	33
Materials.....	34
Design and Procedure.....	35
Independent Variables.....	37
Dependent Variables.....	39
Additional Variables.....	41

<u>Chapter</u>	<u>Page</u>
RESULTS.....	42
Manipulation Checks.....	42
Hypothesis 1.....	45
Hypothesis 2.....	46
Hypothesis 3.....	47
Additional Analyses.....	48
DISCUSSION.....	53
Manipulation Checks.....	53
Hypothesis 1.....	56
Hypothesis 2.....	58
Hypothesis 3.....	62
Additional Analyses.....	65
Limitations.....	69
Conclusions.....	70
REFERENCES.....	76

<u>Chapter</u>	<u>Page</u>
APPENDICES.....	84
A. Organizational Chart.....	84
B. Adapted Skillscope.....	86
C. Informed Skillscope.....	92
D. Participant’s Schedule of Events.....	95
E. Summary Rating Form.....	98
F. Manipulation Check Questionnaires.....	100
G. Tables.....	106
VITA.....	117

INTRODUCTION

The present study investigates the relationship between raters' amount of performance information, discrepant self-ratings from the ratee, and subsequent modifications of raters' performance judgments. Several researchers have observed the existence of self-supervisor rating discrepancy, although few studies have focused on what effect these discrepant self-ratings have on supervisors' later perceptions of the ratee. It is proposed in this study that amount of performance information will affect raters' confidence in their initial ratings which will, thereby, affect the degree to which they modify their initial ratings when presented with discrepant self-ratings.

This topic is interesting and important to investigate because of the increasing use of 360-degree feedback systems. These systems incorporate judgments from supervisors, customers/clients, peers, and selves in order to gain a more fair and comprehensive description of one's performance (Tornow, 1993). Self-appraisals, specifically, have received much attention in the performance appraisal literature (Cawley & Levy, 1995; Fahr, Werbel, & Bedeian, 1988; Folger, Kanovsky, & Cropanzano, 1992; Harris & Schaubroek, 1988; Porter & Foti, 1995; Schrader & Steiner, 1995; Thornton, 1968). The use of self-appraisals for evaluation is common because they are thought to increase perceived fairness by the employee and to facilitate discussion between the supervisor and subordinate on the employee's performance (Fahr et al., 1988; Folger et al., 1992). Self-appraisals are also helpful to raters because they provide performance information beyond that which they have directly

observed (Murphy & Cleveland, 1991). Therefore, self-appraisals are often used as a source of performance information which may be combined with other's ratings to result in a "final" performance judgment (Zalesny, 1990).

This task proves to be challenging because many researchers have reported finding significant differences between self and supervisor ratings (Ashford, 1989; Harris & Schaubroek, 1988; Mabe & West, 1982; Thornton, 1968; 1980; Williams & Levy, 1992). Specifically, managers generally have been found to evaluate themselves more favorably than their supervisors (Ashford, 1989; Harris & Schaubroek, 1988; Steele & Ovalle, 1984; Thornton, 1968; Yu & Murphy, 1993). However, some studies have found self-ratings to be lower than supervisor ratings (Cawley & Levy, 1995; Farh & Dobbins, 1989; Farh, Dobbins, & Cheng, 1991; Heneman, 1974; Nilsen & Campbell, 1993; Porter & Foti, 1995; Williams & Levy, 1992). Therefore, supervisors are likely to be in situations in which their subordinates' self-ratings are different from their own and will have to decide how to use this information. Throughout this paper, the term "overrating" (or overrater) will be used to describe situations in which subordinate ratings are higher than supervisor ratings. Likewise, "underrating" (or underrater) describes situations in which subordinate ratings are lower than supervisor ratings. "In-agreement" rating describes situations when there is little or no discrepancy between supervisor and subordinate ratings (Atwater & Yammarino, in press; Nilsen & Campbell, 1993).

With the growing use of self-assessments in organizational settings and the high likelihood that self-supervisor ratings will be discrepant, it is important to investigate the effect of these discrepancies on later judgments made by the supervisor. Therefore, one focus of the present study is the effect of over, under, and in-agreement rating on supervisors' modification of initial performance ratings. Although research identifying self-supervisor discrepancies is extensive, few researchers have investigated the effects of these discrepancies on subsequent performance judgments. Moreover, the results of studies investigating these effects have yielded inconsistent results (Blakely, 1993; Ferris, Judge, Rowland, & Fitzgibbons, 1994; Porter & Foti, 1995; Thornton, 1968).

For example, Thornton (1968) found that managers who rated themselves higher than their supervisors were considered least promotable based on their prior success in the organization. Similarly, using a sample of managers in a hospital, Porter and Foti (1995) found that overraters received lower subsequent performance ratings and lower ratings of likelihood of promotion while underraters received more favorable subsequent performance ratings and higher ratings of likelihood of promotion. Ferris et al. (1994) found that job-focused tactics, those behaviors used to appear more competent at one's job, decreased supervisor affect towards the employee and were negatively related to performance ratings. Overrating may be considered a job-focused tactic which may actually hinder supervisor's performance perceptions. All of these studies show the same relationship between over and underrating and performance

judgments and used actual raters in an organizational setting who had worked with their subordinates for at least three months.

However, in laboratory studies, both Blakely (1993) and Porter and Foti (1995) found the opposite relationship between over and underrating and subsequent judgments. That is, overraters received higher subsequent performance ratings and underraters received lower subsequent performance ratings after the supervisor viewed their self-ratings. In both studies, the supervisors had limited performance information and were evaluating the subordinate on an in-basket task with which the raters had little or no experience.

Research on the effect of amount of performance information on performance ratings has been limited, although a few studies have investigated this relationship (Favero & Ilgen, 1989; Fried, Tieg, & Bellamy, 1992; Heneman & Wexley, 1983). For example, Fried et al. (1992) found that supervisors attempted to avoid conducting a performance appraisal when they had only worked with their subordinate for a short period of time. These authors suggest that the raters had little confidence in the accuracy of their ratings and, therefore, felt their ratings would be difficult to justify (Fried et al., 1992).

The purpose of the present research is to clarify under what conditions raters are likely to adjust or modify their initial ratings when presented with discrepant self-ratings from their subordinate. It is predicted that the amount of performance information obtained by a rater affects the extent to which discrepant self-ratings made

by his/her subordinate will impact subsequent performance judgments. Specifically, the process variable predicted to mediate this relationship is rater's confidence in his or her initial ratings. Paese and Snizek (1991) and Oskamp (1965) found that amount of information was positively related to confidence in a raters's judgments. Therefore, rater confidence is expected to reduce the extent to which subordinates' over or underestimation impacts ensuing performance ratings made by the rater.

Specifically, it is predicted that the greater amount of performance information obtained by the rater, the more confidence he/she will have in his/her ratings and the less likely he/he will be to modify his/her initial ratings after reviewing a discrepant self-assessment from the subordinate. To test this prediction the amount of performance information and over, under, and in-agreement rating by the ratee will be manipulated to determine their effects on raters' modifications of initial performance judgments.

The amount of performance information (high or low) between supervisors and subordinates will be manipulated using the University Edition - Looking Glass, Inc. leadership development simulation. In this simulation, all participants are given management positions and several supervise other participants. It is expected that more interaction between supervisors and managers will provide the supervisor with a greater amount of knowledge about the manager's performance. Although discrepancy between self-other ratings is prevalent in organizations, the number of over, under, and in-agreement raters in a laboratory study is difficult to predict. Therefore, supervisors

will receive bogus self-ratings so that the effect of over, under, and in-agreement rating may be investigated.

Literature Review

SELF-APPRAISAL USE AND BENEFITS

Self-appraisals are a useful instrument in the evaluation and development of employees in organizations. Although research has suggested that the psychometric characteristics of self-appraisals limits their accuracy and utility (Fox & Dinur, 1988; Thornton, 1980), several researchers have found that the use of self-appraisals has many benefits to the employee, supervisor, and organization as a whole. For example, Bassett and Meyer (1968) found that the use of self-appraisals decreased defensiveness in performance appraisal interviews. Self-appraisals also increase communication and an employee's sense of control which leads to greater acceptance of the appraisal process (Farh et al., 1988). In fact, Farh et al. (1988) found that 70% of the employees thought an appraisal system which incorporated self-appraisals was more fair, more accurate, and generally preferred it to a previous system in which no self-appraisal information was collected. The self-appraisal based system also benefitted the supervisors in that they felt less defensive during the performance meetings and felt more justified in making their final decisions (Farh et al., 1988). Farh et al. (1988) mention that self-appraisal based systems are especially advantageous for jobs in which employees work alone or are not directly supervised (e.g., managerial positions).

Moreover, Dipboye and dePontbriand (1981) found that although the type of feedback received is an important determinant of the employee's perceptions of the performance appraisal system, participation in the feedback session, explicitly set goals,

and the extent to which dimensions were job-related were all positively related to perceptions of the performance appraisal system. Therefore, it seems that open communication and a feeling of contribution into the evaluation are ways in which self-appraisals can enhance the performance evaluation process.

Aside from improving perceptions of the appraisal process, evidence suggests that self-appraisals are related to subsequent job performance (Bassett & Meyer, 1968; Meyer, Kay, & French, 1985). For example, Fox and Dinur (1988) found low but significant correlations between self-ratings and subsequent performance in a two-year military training course. There was also less evidence of halo in the self-ratings than in the commander and peer ratings of the target individual.

Another benefit of self-ratings is that they force individuals to reflect on their perceptions of their performance. Popular today are 360-degree feedback systems which incorporate ratings from supervisors, selves, peers, customers, and subordinates so that several aspects of one's performance may be considered. These systems assume that self-awareness is enhanced by knowledge of discrepancies between self and other-perceptions. Thus, a heightened self-awareness is assumed to be essential for improving performance as a leader and has become an integral part of management development programs (Tornow, 1993).

Essentially self-ratings give the employee a chance to contribute information to the evaluation process that may not otherwise be considered. Self-ratings can be especially helpful when supervisors are at a different location or simply do not have the opportunity

to observe the employee's performance. Thus, self-ratings can supplement the supervisor's limited performance information. Although individuals often use self-ratings to promote their most positive attributes, individuals do have an advantage over others because they have knowledge of their actual behaviors and normally are aware of the results of their efforts. As Murphy and Cleveland (1995) discuss, no one individual will have a complete picture of another's performance. Therefore, the use of appraisal systems which include several perspectives (including the self) increases the chance for accurate appraisals. Even organizations who feel that peer and subordinate ratings run counter to the organizational hierarchy can benefit from using at least one other source of information - the self (Murphy & Cleveland, 1995).

One frequent argument against the use of self-ratings is their tendency to be discrepant from other ratings (Harris & Schaubroeck, 1988; Mabe & West, 1982). However, individuals with different perspectives are likely to observe distinct aspects of performance and evaluate them differently (Borman, 1974). This phenomenon may not necessarily be a disadvantage for using self-appraisals, but instead, an opportunity to increase the breadth of performance information collected about an employee. Thus, collecting information from people with different vantage points may serve to improve the accuracy of performance appraisals, in addition to enhancing the employee's perceived fairness of the appraisal system (Folger, et al., 1992).

In 1974, Borman investigated whether more agreement among raters would occur for raters whose work relationship with the ratee was similar. The dimensions to be used

in the rating were developed independently by groups at the same level in the organization. One group consisted of peers and the other of supervisors of the target position. However, peers and supervisors rated the target individual on all the dimensions developed. Borman (1974) found higher inter-rater agreement for peers and supervisors on dimensions that their own group developed. This study provides evidence that raters at different levels in the organization are likely to observe different facets of job performance. As such, discrepancy among ratings does not mean the ratings are not valid. Instead, a discrepancy could exist because some dimensions are simply easier for one rater to evaluate than the other. Further, discrepancies among individual's ratings may become an impetus for organizational change by raising awareness of managerial shortcomings and/or qualities (Tornow, 1993). The following section will discuss the evidence, the possible causes, and effects of discrepant self-other performance ratings.

DISCREPANCY

The existence of discrepancies between self and other ratings has been extremely well documented in the performance appraisal literature (Ashford, 1989; Cawley & Levy, 1995; Farh & Dobbins, 1989; Farh et al., 1991; Harris & Schaubroeck, 1988; Holzbach, 1978; Kacmar, Carlson, Wright, & McMahan, 1995; Mabe & West, 1982; Nilsen & Campbell, 1993; Parker, Taylor, Barrett & Martens, 1959; Thornton, 1968; VanVelsor, Taylor, & Leslie, 1993; Yammarino & Atwater, 1993; Yu & Murphy, 1993). This research has consistently found evidence for overrating. That is, self-ratings are generally

more lenient than other ratings, especially supervisor ratings (Ashford, 1989; Farh & Dobbins, 1989; Harris & Schaubroek, 1988; Holzbach, 1978; Kacmar et al., 1995; Mabe & West, 1982; Nilsen & Campbell, 1993; Parker, et al., 1959; Thornton, 1968; VanVelsor et al., 1993; Yammarino & Atwater, 1993; Yu & Murphy, 1993). However several studies have documented the existence of underrating (occurs when subordinates rate themselves lower than their supervisor) (Cawley & Levy, 1995; Farh & Dobbins, 1989; Heneman, 1974; Nilsen & Campbell, 1993; Porter & Foti, 1995; Williams & Levy, 1992). For example, Farh and Dobbins (1989) found that Chinese workers rated their job performance lower than their supervisors did. These authors suggested that this “modesty bias” was caused by the collectivist nature of the Chinese culture and they warn against embracing the notion that self-ratings are always lenient. Yu and Murphy (1993) attempted to replicate these findings with a larger sample of Chinese workers. They found no evidence of modesty bias in either the Chinese or American sample of workers. It was suggested that culture may not determine whether employees will over or underrate themselves and that other factors should be investigated (Yu & Murphy, 1993).

In fact, researchers have found other characteristics that affect the leniency of self-ratings such as type of job (blue or white collar), age, tenure, and self-esteem (Harris & Schaubroek, 1988; Shore & Thornton, 1986). Nilsen and Campbell (1993) argue that over or underestimation of one’s attributes is a stable personality characteristic. They asked managers to complete the Campbell Leadership Index (CLI) which measures several personality attributes (e.g., dependability, affability, resilience). Three to five other

employees with whom the target manager worked were also asked to use the CLI to rate the target manager. Discrepancies between the self and other ratings were assessed. One month later this process was repeated. The mean correlation between discrepancies on all personality attributes was .71 indicating that the nature of the discrepancy was stable over time. Participants also completed the Executive Success Profile (ESP) which measured managerial skill. The nature of discrepancies was constant across the measures of personality (CLI) and managerial skills (ESP). Therefore, it is likely that supervisors will be faced with self-ratings from subordinates who have the tendency to always over or underrate their performance. Other causes of self-other rating discrepancy are discussed below.

Thus, research suggests that self-other ratings of performance are prone to be discrepant. It is difficult to assess whether the self or supervisor is accurate. It is likely that both of them are accurate to some extent. However, in an organizational setting, whether the supervisor is objectively “right” or “wrong” is not of critical importance (Ashford, 1989). Typically, supervisors are in control of their subordinates’ futures to the extent that their judgments affect organizational rewards such as promotion, raises, developmental opportunities, etc. Therefore, it is important for employees to be aware of how others view them because those are the judgments that are likely to determine the employee’s career advancement (Ashford, 1989).

Discrepancy, although a natural outcome of performance appraisals employing multiple sources, can create negative feelings in an appraisal interview. Thus, to the

extent that the employee's perceptions match his/her supervisor's, the negative consequences of discrepant perceptions may be attenuated. It is advantageous for an employee to be self-aware and/or to not over or underestimate their performance because if employee's perceptions of their understanding of task requirements is inaccurate, they are unlikely to seek out the appropriate information to complete the task successfully (Nilsen & Campbell, 1993). In a later section, a comprehensive discussion of the effects of discrepant perceptions of performance on the employee, supervisor, and organization as a whole is provided. First, the literature citing various causes of discrepant self-supervisor ratings is reviewed.

Causes of Discrepancy

Reasons for discrepant perceptions and evaluations of performance are widespread. They include personal characteristics of the rater and ratee, motivational forces acting on the rater at the time of rating, the appraisal purpose, the culture of the organization, and several more. Several studies have investigated why being at a different level in the organizational hierarchy than the target person so often contributes to a discrepancy in ratings (Cawley & Levy, 1995; Wohlers & London, 1989; Yammarino & Atwater, 1993). For example, Wohlers and London (1989) hypothesized that certain managerial characteristics were inherently more difficult to evaluate because they were difficult to observe. As predicted, they found less agreement among co-workers on characteristics that were considered "difficult to rate" than on other characteristics. Interestingly, they found that agreement was also related to the relevance of the

characteristic to the rater's job. Thus, because of their position in the organization, certain characteristics were more familiar to the raters which increased rater agreement on those characteristics.

In another study, Cawley and Levy (1995) hypothesized that same level raters (e.g., peers, selves) would agree on individual competency ratings as well as overall ratings of performance of a target individual, but that different level raters would agree on competency ratings only. They suggest that raters at different levels of the organization will weight characteristics differently to arrive at one's overall level of performance. This hypothesis was confirmed indicating that employees at the same level of an organization place importance on similar characteristics when considering the overall performance of an individual. This finding has been supported by other research as well (Klimoski & London, 1974; Parker et al., 1959; Schmitt, Noe, & Gottshalk, 1986).

For example, Schmitt et al., (1986) had school teachers (subordinates) and administrators (supervisors) evaluate principals on several dimensions. The principals also evaluated themselves on the same dimensions. Overall judgments were regressed on each performance dimension rating for each group of raters (self, subordinates, and supervisors). The authors found significant differences in the regression equations indicating that each group of raters weighted the dimensions differently to arrive at their overall evaluation. Zammuto, London, and Rowland (1982) also found differences on weights applied to performance dimensions. They found that ratees stressed skill and

technical competence whereas their supervisors put more emphasis on the outcomes and results of their employees' effort.

A large body of research has focused on what variables affect the extent to which self-other ratings will differ. Steele and Ovalle (1984) suggested that self-raters may intentionally use a different criteria when rating their own performance. They found that when employees were asked to rate themselves based on formal or informal feedback from their supervisors ("Your supervisor believes..."), their ratings were much more congruent to their supervisors ratings than their original self-ratings ("Base your ratings on your view of your own performance."). This study suggests that when prompted to think about how others view them, ratees are able to adjust their ratings to be more congruent with other's ratings.

In a laboratory study, Farh and Dobbins (1989) found that objective information on other's performance reduced discrepancy between self and supervisor ratings. The correlation between self and supervisor ratings for participants who received this social comparison information was .42 versus .13 for participants in the control condition. Additional analyses indicated higher interrater agreement among experimental condition participants' estimates of average performance than estimates made by participants in the control group suggesting that participants were able to combine several individual's objective performance to create a standard by which to compare themselves.

Comparison standards also affected discrepancy in a study conducted by Schrader and Steiner (1995). These researchers had supervisors rate employees using five different

comparison standards: 1) ambiguous (no specific instructions), 2) internal (rating based on employee's improvement since the last evaluation), 3) absolute (rating based on a specific goal), 4) relative (rating based on relation to other's performance), 5) multiple (combination of the other four). Correlations were .26, .43, .50, .43, and .55 for ambiguous, internal, absolute, relative, and multiple, respectively. Thus, using any comparison standard other than "ambiguous" greatly improved self-supervisor agreement.

As mentioned above, organizational characteristics can also affect self-other rating discrepancy. Williams and Levy (1992) found that the degree to which individuals agreed with standards set by the organization, understood the appraisal system, and knew explicitly what was expected of them contributed to what these authors termed "perceived system knowledge" (PSK). They predicted a positive relationship between an employee's PSK and self-supervisor rating congruence. In fact, there was a strong positive relationship between high PSK and self-supervisor rating congruence and a moderate negative relationship between low PSK and self-supervisor rating congruence. This study suggests that organizations can improve self-supervisor rating congruence by educating their employees about general performance standards, the appraisal system, and specific requirements for successful performance in their jobs.

Clearly, performance rating discrepancies exist and are caused by a variety of variables. The focus of the present study is to investigate the effects of these pervasive discrepancies on subsequent judgments made by the supervisor. The next section surveys

literature that has examined what effects performance rating discrepancies can have on the relationships between the individuals involved and on the organization as a whole.

Effects of Discrepancy

The majority of research investigating the effect discrepancy between self-supervisor ratings has found negative consequences for the overrater, the supervisor, and the organization (Atwater & Yammarino, in press; Ferris et al., 1994; Giacalone, 1985; Godfrey, Jones, & Lord, 1986; Nilsen & Campell, 1993; Pearce & Porter, 1986; Thornton, 1968; Wayne & Ferris, 1990). For example, Pearce and Porter (1986) found that negative and satisfactory ratings made by a supervisor resulted in negative feelings by the subordinate towards the organization and the performance appraisal system. They suggested that even satisfactory ratings injure one's self-concept because the majority of people tend to consider themselves better than average.

A number of years ago, Thornton (1968) had managers and their supervisors evaluate the manager's performance. Thornton discusses the fact that in upper-level management, objective measures of performance are often difficult to obtain. As such, there is a heavy reliance on performance ratings made by supervisors to measure managerial performance. The results of this study indicated that the managers tended to overrate their performance and there was considerable disagreement between managers' and supervisors' performance ratings. However, in some instances the supervisors made higher ratings than the managers on specific performance categories. Later, the manager's supervisor and the supervisor's supervisor rated each manager's promotability which was

based on previous performance, current status, and future potential in the organization.

Overall, the managers who overrated their performance were considered least promotable (Thornton, 1968). This study provides evidence that subsequent judgments can be affected by self-appraisal information even after the formal appraisal process is completed.

Research on self-promotion further elucidates the consequences of overrating (Godfrey et al., 1986; Ferris et al., 1994; Kacmar, Delery, & Ferris, 1992; Wayne & Ferris, 1990). Overrating may be considered a form of self-promotion in that lenient self-ratings present a favorable impression of one's competencies on the job to a supervisor. Godfrey et al. (1986) investigated what effect self-promotion techniques (portraying competence) and ingratiation techniques (seeking likability) during an unstructured conversation had on likability and competence of the self-promoter or ingratiator. Likability and competence ratings were compared to those taken from a previous conversation in which no attempt was made to self-promote or ingratiate. They found that ingratiators gained in likability ratings without losing in perceived competence ratings. However, self-promoters' likability scores declined without a compensatory increase in perceived competence. This study shows that, in unstructured conversations, it is difficult to enhance others perceptions of competence without losing likability (Godfrey et al., 1986).

Similarly, Wayne and Ferris (1990) investigated the influence of three types of impression management tactics on performance and likability ratings of the target individual. Job-related tactics were those in which the employee presents positive aspects

of his/her job performance (e.g., taking credit for positive events, coming to work early and staying late). These behaviors seem to be comparable to favorable self-ratings in that they seek to enhance perceptions of job performance. Self-focused tactics were defined to be behaviors that portrayed the target individual to be a “nice and polite” person, whereas supervisor-focused tactics (comparable to ingratiation) involved praising the supervisor and doing favors for him or her. The results indicated that job-related tactics were negatively related to performance ratings and self-focused and supervisor-focused tactics were not related to performance ratings in a sample of bank employees (Wayne & Ferris, 1990). Thus, it is likely that lenient self-ratings or self-ratings that are greater than supervisor’s ratings may be perceived as self-promoting and may, under some conditions, result in negative consequences for the employee.

One of these conditions was discovered by Giacalone in 1985 in a laboratory study. He suggested that, although their intended purpose is enhancing others’ perceptions, entitlements (claims of responsibility for positive events), may actually hinder others’ perceptions and prevent the entitler from gaining organizational rewards. Giacalone (1985) predicted that entitlers would be considered immodest and to lack humility which would lead to negative reactions. However, the results suggested that entitlements were virtually ignored unless they were disconfirmed by a credible source. In which case, negative ratings of the entitler were observed. Giacalone (1985) concluded that entitlers should be sure that others cannot discount their entitlements by making unverifiable entitlements and/or by investigating how much others know about the positive

outcome for which they are taking credit. Thus, it is possible that overraters whose supervisors have evidence that their ratings are not justified, are more likely to create negative perceptions. These negative perceptions may translate into lower performance and/or promotability ratings, thereby stifling the overrater's career progress.

In fact, Ferris et al. (1994) specifically studied the influence of supervisor-focused and job-focused tactics on performance ratings. These authors found that job-focused tactics, those behaviors used to appear more competent at one's job, decreased supervisor affect towards the employee and were negatively related to performance ratings. Supervisor-focused tactics (e.g., ingratiation behaviors) increased supervisor affect and led to favorable performance ratings. Thus, evidence suggests that self-promotion may have negative consequences for employees. To the extent that overraters are perceived to be self-promoters, the same results are likely to occur for them.

However, Blakely (1993) conducted a lab study investigating the effects of relatively higher and lower self-ratings on the change in supervisor's performance judgments and raise increases. Blakely (1993) manipulated the direction of the self-ratings relative to the supervisor's ratings and whether the subordinate would contest the supervisor's ratings or not (high or low contest condition). The participants in this study were undergraduate business majors who were randomly chosen (by flipping a coin) to take the role of a supervisor or subordinate. The subordinate completed a managerial decision-making task (in-basket) with which they had no prior experience. Meanwhile, the supervisor looked over the in-basket exercise and was trained on how to conduct a

performance appraisal interview. Upon completion of the in-basket exercise, the supervisor looked over the completed in-basket and rated the subordinate's responses. After viewing the bogus self-ratings of the subordinate, the supervisor was able to change his/her initial performance ratings and made salary increase recommendations (Blakely, 1993).

The results indicated a main effect of direction of performance rating discrepancy on supervisors' final performance ratings. Eleven of the 24 supervisors whose subordinates rated themselves unfavorably relative to supervisors' ratings lowered their initial ratings. Four of the 24 supervisors in the relatively favorable self-rating condition raised their initial ratings. However, of the four supervisors who raised their ratings, all of these were told that the subordinate would contest their performance ratings (high contest condition) if they were discrepant from the subordinate's self-ratings. Therefore, it is likely that the supervisors raised their ratings to avoid an unpleasant appraisal interview rather than because they believed the self-ratings provided additional performance information (Blakely, 1993). Supervisors in the low contest condition did not change their initial ratings when presented with the subordinate's relatively high self-ratings. Moreover, eight of the 11 supervisors who lowered their ratings were in the low contest condition indicating that the relatively low self-ratings provided the supervisor with relevant additional performance information (Blakely, 1993).

The results of this study are similar to those found by Porter and Foti (1995). In this study, participants evaluated an in-basket completed by a fictional manager.

Subsequently, participants received bogus self-ratings that were either higher or lower than the participant's ratings and were then able to re-evaluate the fictional manager.

Participants' final ratings in this study, as in Blakely's (1993) study, were influenced by the self-ratings of the fictional manager. That is, the higher the self-ratings, the higher the final performance ratings made by the participant. Likewise, relatively low self-ratings were positively related to lower final performance ratings.

However, the results of these studies should be interpreted with caution. It is important to realize that the presentation of performance information in a laboratory study is not equivalent to the type of performance information typically available to supervisors in an organizational setting. The participants in these studies were undergraduate college students with little or no experience with evaluating performance on an in-basket exercise. Therefore, their judgments may have been easily changed due to the lack of confidence in their original ratings.

The research on the effects of overrating on subsequent performance judgments is still unclear. Several of the studies described have suggested that relatively high self-ratings cause negative performance judgments, although this has not been explicitly tested except in a laboratory setting where little performance information was available. The present study seeks to directly investigate the effects of discrepancy on supervisor's subsequent performance judgments. Based on Blakely (1993) and Porter and Foti (1995) it seems that limited performance information leads to greater acceptance of a subordinate's self-ratings indicated by a change in performance ratings towards the

subordinate's self-ratings. However, research by Godfrey et al. (1986), Ferris et al. (1994), Thornton (1968), Van Velsor et al. (1993), and Wayne and Ferris (1990) suggests that overrating may cause negative reactions and, therefore, negative performance ratings when the supervisor is familiar with the subordinate's work performance.

Thus, it is logical to conclude that the extent to which a supervisor is influenced by a subordinate's self-ratings may be dependent upon his/her confidence in their initial ratings. Amount of performance information available is one variable that has been shown to affect one's confidence in his/her initial ratings (Paese & Sniezek, 1991). The following section discusses research investigating the effect of amount of performance information and confidence on performance judgments.

AMOUNT OF PERFORMANCE INFORMATION

The amount of performance information obtained for an employee would seem to effect subsequent performance judgments, although very little research on the these effects or the process by which this would occur has been conducted (Judge & Ferris, 1993; Ferris et al., 1994). Supervisors in actual organizations often collect little performance information from their subordinates because other requirements of the job prevent them from adequately observing their subordinate's performance and/or because they have limited interaction with subordinate due to the nature of their positions (Judge & Ferris, 1993; Rothstein, 1990; Zalesny, 1990). In fact, many performance evaluations are based on a small sample of actual observed behaviors which contributes to the unreliability and inaccuracy of performance ratings (Heneman & Wexley, 1983; Favero & Ilgen, 1989;

Murphy & Cleveland, 1991; Rothstein, 1990). For example, Rothstein (1990) found that opportunity to observe was strongly related to reliability of performance ratings.

Moreover, Greenberg (1986) found that rater familiarity with the ratee's work led to perceptions of procedural justice in a sample of middle managers from various organizations.

In their model, Judge and Ferris (1993) consider two variables that are related to those in the present study; opportunity to observe and supervisor's inference of subordinate's self-rating of job performance. They predicted that opportunity to observe performance would result in lenient performance ratings because supervisors are motivated to seek out positive information about their subordinates. They contend that supervisors want to search for positive information because they want to avoid giving negative evaluations and potential conflict with their subordinate. Therefore, the more opportunity to observe performance, the more opportunity supervisors have to collect positive information (Judge & Ferris, 1993). Judge and Ferris (1993) also predicted that the supervisor's inference of a positive self-rating of the subordinate's job performance would positively influence performance ratings. Both of these predictions were confirmed; however, the study was non-experimental so the causal relationship between opportunity to observe and inference of subordinate's self-rating on performance ratings is still unclear. Moreover, in this study, no data was collected to elucidate the interactive effect of opportunity to observe and supervisor's inference of subordinate's self-ratings on performance ratings; these relationships were considered individually. Judge and Ferris

(1993) suggest that these variables in particular deserve further attention in future research.

A few studies provide evidence that accurate ratings are related to the amount of performance information obtained and that raters are aware of this relationship. In a laboratory study, Heneman and Wexley (1983) found that the fewer the number of critical incidents observed, the less accurate raters were in recalling the occurrence of the critical incident. This effect was only observed when rating was delayed three weeks after the participants observed the performance. Moreover, Favero and Ilgen (1989) also found that amount of performance information collected was positively related to rating accuracy. Fried et al. (1992) provided evidence to suggest that raters are aware of this relationship between information and accurate ratings. They investigated the variables associated with performance appraisal avoidance and found that supervisors were more likely to avoid conducting performance appraisals when they had only worked with the subordinate for a fairly short time. These authors suggest that the supervisors did not feel confident that their ratings were accurate and, therefore, felt less equipped to defend their judgments (Fried et al., 1992).

Empirical evidence for the relationship between amount of information and confidence in judgments is provided by Oskamp (1965) and by Paese and Snizek (1991). Oskamp (1965) found that an increase in amount of diagnostic information provided to raters increased their confidence in judgments made by clinical psychologists. However, amount of practice was confounded with amount of information in this study. Paese and

Sniezek (1991) attempt to consider these effects on confidence separately. They define confidence to be “the stated probability that one’s previous judgment is correct” (p.100) and investigated the influence of amount of relevant information on confidence in judgments. In Study 1, amount of relevant information was a between-subjects factor and was manipulated by giving participants limited information for which to base concurrent and predictive judgments of baseball players’ earned run averages (ERA). The authors did not find significant differences in confidence levels presumably because participants were not aware of the amount of information they could have received relative to what they did receive (Paese & Sniezek, 1991). Therefore, in a second experiment, they manipulated amount of information in a within-subjects design. They found that confidence was positively related to amount of information regardless of whether the information was directly relevant to the judgments they were required to make or not (Paese & Sniezek, 1991). Interestingly, confidence was not related to accuracy in either studies by Paese and Sniezek (1991) or by Oskamp (1965).

Confidence in ratings has also been found to be related to the degree to which raters may be persuaded to change their ratings (Zalesny, 1990). With the growing use of multiple rater instruments, how information from multiple raters is combined is an important research question. Zalesny (1991) contends that the amount of rater confidence will affect his or her influence on others when making a consensus judgment. Likewise, the greater the confidence the rater has in his/her judgment, the less likely that rater will be influenced by others to change his/her ratings. Zalesny (1991) predicted that this would

occur regardless of whether the raters were able to discuss their ratings or not (no interaction with the other raters). In her study, student teachers watched a 30-minute videotape of a teacher performing a lecture and rated the teachers' performance on instructional methods, classroom management, interpersonal relationships, professional qualities, and overall performance. Participants rated the importance for effective teacher performance of each of these dimensions and their confidence level on each of their ratings. Subsequently, participants were given performance and confidence ratings of the teacher in the videotape supposedly from student teachers from other districts. Participants were asked to integrate the ratings and make a combined rating of the teacher. Finally, participants evaluated the teacher again on all performance dimensions, made confidence ratings, and were asked the extent to which they changed their personal evaluation of the teacher.

Confidence ratings were negatively related to perceived change in performance ratings for participants on instructional methods, although the correlations on all other dimensions were in the expected direction. In experiments 2 and 3 the raters were able to interact with each other to reach their consensus ratings. Again, the correlations between confidence and perceived change were in the expected direction and three out of the five rating categories were significant in each experiment. However, contrary to the hypothesis, there was no relationship between actual change in ratings and confidence in any of the experiments (Zalesny, 1991).

Several possible reasons could explain this finding. Confidence ratings were made on a 7-point scale (7 = very confident) and the mean on initial confidence ratings was 5.6 with a standard deviation of 1.2. It is not surprising that confidence levels were high because participants could focus all their attention on the teacher in the videotape. However, participants received written reasons for the other raters' responses in experiment 1 and could discuss others' reasons for their ratings in experiments 2 and 3. Thus, these persuasive arguments may have attenuated the effect of the participants' original confidence level. In fact, raters significantly lowered their performance ratings when other raters' evaluations were lower than their own and raised their rating, although not significantly, when other's ratings were higher than their own (Zelasny, 1990). Moreover, the participants knew that the other raters had seen the same 30 minute videotape and that all of the confidence ratings were based on the same limited amount of information. Thus, their own confidence level may not have been considered as important as persuasive arguments from other raters who could have easily seen behaviors that the participant missed.

In addition, the standard deviation of initial confidence ratings was 1.2. Thus, the nature of the task did not seem to produce a lot of variability on the confidence measure. The task in the present study should promote more variability in confidence ratings so that the relationship between confidence and other variables may be detected. Raters will receive performance information more realistically in that they will observe the ratee perform while also attempting to complete several other tasks and/or will interact with the

ratee to obtain performance information. Moreover, amount of interaction and observation time will be manipulated which, in turn, is expected to affect confidence levels.

As further evidence for this expected effect, Zelasny (1990) found that for performance dimensions for which there were very few directly observable behaviors, participants were more likely to be influenced than for those dimensions that were directly observable. This finding supports the idea that the more performance information available to the rater, the less likely he or she will be to change his original performance rating. Although confidence did not mediate this relationship in Zelasny (1990), the design of the present study will replicate a more realistic setting in which performance judgments take place. Moreover, participants in the present study will decide to what extent they will combine their ratings with one other source of information, the ratee's ratings, rather than with other raters' ratings.

To summarize, it seems that the amount of performance information collected by the supervisor is important for the accuracy of performance ratings (Heneman & Wexley, 1983; Favero & Ilgen, 1989), rater's motivation and confidence in his/her ratings (Fried et al., 1992; Harris, 1994; Paese & Snizek, 1990), and ratee perceptions of a fair and just performance evaluation (Greenberg, 1986). The present study seeks to contribute to the literature by investigating the effect of amount of performance information on the **modification** of supervisors' performance ratings when presented with discrepant self-appraisal information. It is predicted that supervisors' **confidence** in their original ratings

will mediate the relationship between amount of performance information and subsequent adjustments to those ratings (Oskamp, 1965; Paese & Sniezek, 1991; Zelasny, 1991).

To test these predictions, amount of performance information obtained by the supervisor and over, under, and in-agreement self-ratings will be manipulated in a laboratory study using a managerial simulation exercise. Graduate students and upper-level undergraduates in the College of Business, Engineering, Human Resources, and Psychology will participate in the University Edition - Looking Glass, Inc. simulation developed by the Center for Creative Leadership. This simulation was developed by industrial/organizational psychologists and gives participants the opportunity to realistically experience the job demands of a manager in a multilevel organization. Twenty participants will participate in the simulation at a time and will be randomly assigned to positions in the simulated organization. Seven of these participants will supervise two to three other participants resulting in 19 supervisor-subordinate pairs. Amount of performance information collected by the supervisor will be controlled by limiting the interaction between supervisors and subordinates during the simulation. At the completion of the simulation, subordinates and supervisors will complete an adaptation of SkillScope which is a performance appraisal measure created by the Center for Creative Leadership designed to measure attributes relevant to managerial positions. In order to manipulate over, under, and in-agreement rating, bogus self-ratings will be given to the supervisors before they have the opportunity to modify their original ratings.

Based on the combined results of studies investigating the effect of self-supervisor rating discrepancy on subsequent evaluative judgments (Blakely, 1993; Ferris et al., 1994; Harris, 1994; Porter & Foti, 1995; Wayne & Ferris, 1990) and the effect of amount of information on confidence ratings (Paese & Sniezek, 1991; Zelasny, 1990) the following is hypothesized.

H1: Amount of performance information will be negatively related to modification of supervisors' original ratings after viewing discrepant self-ratings.

H2: Confidence in original ratings will mediate the relationship between amount of performance information and modification of supervisor ratings after viewing discrepant self-ratings.

For the following hypotheses, positive modification scores represents an increase in the supervisors' ratings and negative modification scores represents a decrease in the supervisors' ratings after reviewing the self-ratings of their subordinates.

H3: It is predicted that amount of performance information (API) and discrepancy will interact to affect the supervisors' modification of their initial performance ratings. The specific nature of this interaction is predicted as follows.

H3a: Raters' modification scores in the low API/overrater condition will be greater than raters' modification scores in the high API/overrater condition.

H3b: Raters' modification scores in the low API/underrater condition will be lower than raters' modification scores in the high API/underrater condition.

H3c: Raters' modification scores in the low API/overrater condition will be higher than raters' modification scores in the low API/underrater condition or in the low API/agreement condition.

Based on the findings from the self-promotion literature, it is predicted that supervisors in the overrater condition will be inclined to modify their ratings downwards (Ferris et al., 1994; Giacalone, 1985; Wayne & Ferris, 1990).

H3d: Raters' modification scores in the high API/overrater condition will be lower than raters' modification scores in the high API/underrater condition or in the high API/agreement condition.

Method

Pilot Study

Before conducting the actual study, it was necessary to conduct a pilot study to be sure that supervisors' original ratings would be normally distributed so that discrepancy could be manipulated equally in the over and underrater conditions. Moreover, it was important to verify how different the self-ratings of subordinates needed to be for the supervisors to perceive them as discrepant from their own. Therefore, the pilot study was conducted to 1) find out the variability of supervisor's original ratings and, 2) to ensure that a 2-point discrepancy would be perceived as different to supervisors.

The simulation for the pilot study was conducted in the high API condition. Discrepancy was manipulated by creating self-ratings that were different by 2 points

whenever possible. To ensure that this amount of discrepancy was perceived as different, supervisors were asked the extent to which their subordinates self-ratings were different from their own. In the event that a 2 point discrepancy was not perceived as different, supervisors were also asked how different they would have felt their subordinate's self-ratings were if they were different by 3 points.

The results of the pilot study indicated that the distribution of supervisors' original ratings enabled the manipulation of discrepancy to be constant for the over and underrater conditions. Moreover, self-ratings that were discrepant by 2 points were perceived to be significantly different by the supervisors. Given these results, it was decided that a 2-point discrepancy would be used in the simulations conducted for the actual study. Therefore, the data collected in the pilot study was included in the analyses of the data from the remaining simulations and is reported in the description of the sample below.

Participants

One-hundred and forty university students participated in this study (59% male and 41% female). Sixty-eight percent were Masters of Business Administration graduate students. Thirteen percent were graduate students in various other programs and the remaining 19% were upper-level undergraduate students who were members of professional organizations on campus. The sample size for the analyses involving "raters" is 49. Each of these raters supervised two or three other participants depending on their position in the fictional organization. Volunteers were paid for their

participation and were treated in accordance with the “Ethical Principles of Psychologists and Code of Conduct” (American Psychological Association, 1992).

Materials

University Edition Looking Glass, Inc - The University Edition - Looking Glass, Inc. is a simulation developed by the Center for Creative Leadership to use as an educational tool for individuals seeking careers in management. Participants were each assigned management roles in a large fictitious organization and were given in-basket memos to read to prepare for the simulation. Those roles included a CEO, three Vice Presidents, nine Directors, and seven Plant Managers (see Appendix A - Organizational Chart). A series of problems were embedded in the in-basket memos which were as ambiguous and complex as those found in actual organizations.

Adapted SkillScope - SkillScope is a self-other evaluation instrument developed by the Center for Creative Leadership. This tool may be used to make judgments about behaviors that are necessary for effective performance on the Looking Glass simulation (e.g., time management). The format of SkillScope was adapted for the purposes of this research and was used by all participants to evaluate themselves and for supervisors to evaluate their subordinates (see Appendix B). SkillScope consists of 15 major performance dimensions and requires between 4 and 9 specific judgments to be made for each dimension. For use with the University Edition-Looking Glass and for the purposes of this study, ten of the performance dimensions were excluded. The remaining five performance dimensions were chosen because they measured areas of

performance that could be captured during the Looking Glass simulation (see Appendix B). These were time management, problem identification, decision making, planning and organizing, leadership, and an overall rating. The specific items relating to each dimension were used to define the performance dimensions. In its original form, SkillScope has only a two point rating scale. This was changed to a 9-point scale for purposes of this research.

Design and Procedure

A 2 (high and low amount of performance information) X 3 (overrater, underrater, and in-agreement rater) mixed model design was implemented for this study. Amount of performance information was a between-subjects factor and discrepancy was a within-subjects factor. Four high amount of performance information simulations and three low amount of performance information simulations were conducted. Participants were randomly assigned to conditions as well as to their role in the University Edition - Looking Glass simulation. Participants were given their in-baskets one or two days prior to the simulation and were instructed not to discuss the simulation with other participants before the simulation began.

Before the simulation started each participant signed an informed consent form and was given another one to keep (see Appendix C). During the simulation the experimenter distributed a schedule of activities, including the times the supervisors and subordinates were to meet with each other before the simulation began (see Appendix D). Those participants who were only involved in one meeting were asked

to sit at their desks and to look over their in-basket memos while the other meetings were taking place. Following the meetings, participants were told to begin the simulation. Those participants in the low amount of performance information condition were instructed that they could not communicate with their direct supervisor or subordinates during the simulation and that they must try to resolve the problems embedded in the simulation with their peers or by themselves. Participants in the high amount of performance information condition were told to meet with their supervisor and subordinates frequently throughout the simulation because this would result in the resolution of more problems.

Immediately following the simulation all participants sat at their desks and evaluated their performance. After all the self-ratings were complete, those individuals who did not supervise others were taken out of the room to fill out the manipulation check questionnaire and to be paid. The participants who did supervise others were asked to evaluate each of their subordinates' performance during the simulation using the adapted SkillScope. The experimenter collected the supervisors' ratings and explained that she was going to leave to sort the self-ratings, original ratings, summary rating forms, and final performance evaluations for each supervisor-subordinate pair. Participants were asked to complete the Wonderlic Personnel test as a filler task. During this time, a research assistant administered the 12-minute Wonderlic Personnel Test to the supervisors while the experimenter actually created the bogus self-ratings.

After the bogus self-ratings were complete, the experimenter returned and read the following to the subordinates. "I will be giving you the self-ratings of your subordinates. To the extent that you use these as additional sources of performance information is up to you. When you have finished looking over your subordinates' ratings record them on the summary rating form and re-evaluate your subordinate. For the purposes of the research it is important that the most accurate description of your subordinate's performance be captured." Supervisors were handed the self-ratings, their original ratings, the summary rating form, and the final performance evaluation form for each of their subordinates. The summary rating form required them to transfer their original performance ratings as well as the subordinates' self-ratings onto this form and add up the ratings given by each (see Appendix E). This activity forced supervisors to attend to their subordinates' self-ratings and to notice discrepancy if it existed. After making their final judgments, participants were given the manipulation check questionnaires (Appendix F). Finally, the participants were thoroughly debriefed by the experimenter in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 1992) as to the purpose of the study and the use of bogus self-ratings.

Independent Variables

Amount of Performance Information. To control the amount of performance information the supervisors obtained, the time allowed for supervisors to interact and observe their subordinates was monitored. In the low amount of performance

information condition, supervisors and subordinates met for 10 minutes before the simulation began and for 10 minutes at the end of the simulation. No other communication between supervisors and subordinates was permitted during the simulation and was controlled by physically separating supervisors from their subordinates to prevent interaction. Therefore, the Plant Manager and Vice Presidents were in one-half of the room and the Directors and the President were in the other half and were separated by a divider.

Participants in the high amount of performance information condition also met for 10 minutes before and at the end of the simulation. However, prior to the simulation, participants in this condition were also told that a great deal of interaction between supervisors and subordinates is encouraged and is related to how many problems can be successfully resolved. Moreover, work stations were set up such that each division of the company was grouped together. Therefore, there was no physical obstacles preventing observation or interaction for any of the supervisor-subordinate pairs.

Discrepancy. The adapted version of Skill Scope was used to present the self-appraisal information to the supervisor during the second phase of the experiment (Appendix B, Self-Appraisal Form). Subordinates were randomly assigned to over, under, or in-agreement rater conditions before the simulation began. To manipulate discrepancy in the over and underrater conditions, self-ratings were made to be two points higher and two points lower, respectively, than the supervisor's original rating on every

dimension. Discrepancy was a within-subjects factor so the five raters (per simulation) who supervised three subordinates were given bogus self-ratings that were higher (overrater), lower (underrater), and in-agreement with the raters' original ratings. The two supervisors with only two subordinates were given only higher and lower bogus self-ratings. Therefore, the sample size in the high amount of performance information condition was 28, 20, and 28 in the overrater, in-agreement rater, and underrater conditions, respectively. Because one fewer low amount of performance information condition was completed, the sample size for this condition was 21, 15, and 21 in the overrater, in-agreement rater, and underrater conditions respectively.

To ensure that the supervisors attended to the discrepant self-ratings, supervisors were asked to complete a summary rating form (see Appendix E).

Dependent Variables

Confidence Ratings. Supervisors were asked to respond on a 5-point likert-type scale how confident they were in each of the initial ratings they made. Confidence ratings were also collected on supervisor's final performance ratings (see Appendix B, Section II).

Performance Rating Modification. The degree and direction of supervisors' rating modification on the adapted SkillScope (see Appendix B, Section I) was calculated on each performance dimension and on the overall dimension. The supervisors' second performance ratings were subtracted from their original performance rating on each dimension rated so that a positive number indicated an

upwards modification. Likewise, a negative number indicated a downwards modification. These values were combined across dimensions and averaged to result in one value indicative of the degree and direction of the supervisors' modifications for each of his or her subordinates.

Manipulation Checks. All manipulation check items are found in Appendix F. To check the discrepancy manipulation, supervisors were asked to respond on a 5-point likert-type scale to what extent they felt their subordinates' self-ratings were discrepant from their own ratings. Supervisors were also asked whether their subordinates' ratings, overall, were higher, lower, or in-agreement with their ratings. To check the amount of information manipulation, each of the seven participants who supervise others were observed every 2 minutes during the one hour and 20 minutes of simulation time (the 10-minute meetings at the end were not included). An observer recorded with whom the supervisor was interacting. As an additional check, all participants were asked what percentage of their time was spent with their peers, subordinates (when applicable), supervisor, or alone (see Appendix F). These manipulation checks indicated the amount of time spent with the subordinate but did not necessarily indicate how much performance information was retrieved. Therefore, for the five dimensions used, supervisors indicated whether they observed behaviors which corresponded to the dimensions in the original version of SkillScope (see Appendix F). Supervisors indicated whether or not they observed behaviors related to each item corresponding to the performance dimension.

Additional Variables

Perceptions of Modification. Participants' perceptions of whether they modified their ratings or not were measured by asking "Did you change your rating of this subordinate after viewing his/her self-rating?" This question was followed up by asking participants to list reasons why they did or did not change their ratings to provide potential mediators to explore in future research in the event that confidence was not a full mediator.

Impression Change. As part of the first manipulation check questionnaire, participants were asked to rate their impression of each of their subordinates before they viewed their self-ratings and after they viewed their self-ratings. Participants responded on a 5-point scale (1 = poor performer, 5 = excellent performer).

Confidence in Final Performance Ratings. Confidence ratings on supervisors' final performance ratings were measured although no formal hypotheses were made about these ratings (see Appendix B, Performance Appraisal Form, Section II).

Perceived Fairness. Additional items were asked to all subordinates to measure their perceived fairness of the appraisal process. It was expected that subordinates in the low amount of performance information condition would consider the appraisal system to be less fair and to believe that their self-ratings were more accurate than their supervisors than those in the high amount of performance information condition. These items may be found in Appendix F.

Results

Manipulation Checks

Statistics describing the manipulation check variables are provided in Table 1 and are numbered in the order in which they are discussed (see Appendix G).

Manipulation Check #1. To check the discrepancy manipulation, participants' responses to "Rate how different you felt your subordinate's self-ratings of performance were from your original ratings of his or her performance" were analyzed with a 2 (high and low API) by 3 (underrater, in-agreement, overrater) ANOVA with repeated measures on the second factor (5-point scale: 1=not at all different, 5=extremely different). It was expected there would be a main effect for discrepancy in that the mean in the in-agreement condition would be significantly lower than the mean in both the overrater and underrater conditions. The results revealed that the interaction was not significant, $F(2,66) = 1.60$, *ns.*, but, as expected, the main effect for discrepancy was significant, $F(2,66) = 36.68$, $p < .05$. To determine which means were significantly different, planned comparisons were made with t-tests for paired samples. To ensure that the strength of the discrepancy manipulation was equivalent, the means in the overrater and underrater conditions were compared but no differences were expected. As expected, the mean in the in-agreement condition ($M=1.31$) was significantly different from the mean in the underrater condition ($M=3.17$), $t(34) = 7.87$, $p < .05$ and from the mean in the overrater condition

($M=3.08$), $t(34) = 7.00$, $p < .05$. The underrater and overrater means were not significantly different, $t(34) = .68$, ns .

Manipulation Check #2. As a further test of this manipulation, participants' responses to whether their subordinates' ratings, overall, were higher, lower, or in-agreement with their ratings was also analyzed with a 2 (high and low API) by 3 (underrater, in-agreement, overrater) ANOVA with repeated measures on the second factor (5-point scale: 1 = self-ratings were lower, 3 = self-ratings were the same, 5 = self-ratings were higher). It was expected that there would be a main effect for discrepancy and that the mean in the overrater condition would be significantly higher than the mean in the in-agreement condition which would be higher than the mean in the underrater condition. As predicted, the results of the analysis revealed no interaction, $F(2,64) = .55$, ns , but a significant main effect for discrepancy, $F(2,64) = 95.54$, $p < .05$. Subsequent paired t-tests revealed that mean in the overrater condition ($M=4.35$) was significantly higher than the mean in the in-agreement condition ($M=2.97$), $t(33) = -9.08$, $p < .05$, which was significantly higher than the mean in the underrater condition ($M=1.46$), $t(33) = -7.65$, $p < .05$.

Manipulation Check #3. To check the amount of performance information (API) manipulation, the number of times supervisors were observed interacting with each of their subordinates during the simulation was averaged. Because this mean was zero for the supervisor's in the low amount of performance information condition, the mean in the high amount of performance information condition was tested to see if it

was significantly different from zero with a one-sample t-test. The results indicated that the mean ($M=10.26$) was significantly different from zero, $t(27)=7.68$, $p < .05$.

Manipulation Check #4. To further check the amount of information manipulation, participants' mean response in the high and low API conditions to what percentage of their time was spent with their subordinates was compared with t-tests for independent groups. For each supervisor, the percentage of time spent with each subordinate was averaged. It was expected that this mean would be greater in the high API condition than in the low API condition. The mean percentage of time spent with subordinates in the low API condition ($M=13.66\%$) was significantly lower than in the high API condition ($M=21.83\%$), $t(31)=-2.76$, $p < .05$. The sample size was less than 49 in this analysis due to missing data.

Manipulation Check #5. As an additional check, the subordinates' mean reported percentage of time spent with supervisors was compared for the low and high API condition. Participants in the high API condition were expected to report that they spent significantly more time with their supervisors than those in the low API condition. In fact, the mean percentage of time spent with supervisors in the low API condition ($M=16.66$) was significantly less than in the high API condition ($M=25.78$), $t(138)=-3.92$, $p < .05$.

Manipulation Check #6. To establish evidence that high levels of interaction actually increased the amount of performance information obtained, a 2 (low and high API) by 3 (over, under, and in-agreement) ANOVA with repeated measures on the

second factor was performed on the number of behaviors observed by the supervisor. It was expected that there would be a main effect for amount of performance information; more behaviors observed by the supervisor were expected in the high API condition. The results indicated that there was no interaction or significant main effects. However, the main effect of amount of performance information approached significance, $F(1,33) = 2.71, p = .10$. As shown in Table 1, the means were lower in the low API condition in the underrater and in-agreement rater conditions but were practically equivalent in the overrater condition. Possible reasons for this result are presented in the discussion section below.

A correlation matrix of all of the proxy variables for amount of performance information is found in Table 2. As indicated, almost all of the correlations were significant except those involving number of behaviors observed.

Hypothesis 1

Hypothesis 1 predicted that amount of performance information would be negatively related to modification scores of supervisor's ratings after viewing discrepant self-ratings. Two modification scores were computed for each supervisor-subordinate pair: one was computed across the five performance dimensions and another was computed for the overall performance dimension only. Descriptive statistics on these variables are provided in Table 3 (see Appendix G). Given the similarity of the means and standard deviations of these scores, the modification scores for the overall dimension were used in all of the relevant analyses.

To test hypothesis 1, the raters' modification score for subordinates in the over and underrater conditions were averaged and the absolute value was calculated. This value was correlated with amount of performance information (low=1, high=2) but the relationship was not significant $r = .08$, ns. Because amount of performance information was coded as a dummy variable with very little variance, the manipulation check variables for amount of performance information were used as proxy variables to further test the hypothesized relationship. The correlation between the absolute value of the modification scores and the number of coded interactions, reported time spent with subordinates, reported time spent with supervisor, and number of behaviors observed was $-.03$ (N=49), $.03$ (N=42), $-.04$ (N=41), $-.08$ (N=49), respectively. None of these correlations were significant, therefore, hypothesis 1 was not supported.

Hypothesis 2

Hypothesis 2 predicted that confidence would mediate the relationship between amount of performance information and modification of supervisor ratings. However, the results of hypothesis 1 indicated that these two variables were not related in this data. Therefore, the mediational analysis was not testable. However, because hypothesis 2 also predicted a relationship between confidence and modification and confidence and amount of performance information, these relationships were investigated. Raters' confidence ratings in their original performance ratings were averaged across their subordinates in the overrater and underrater conditions. Statistics describing the confidence variables are in Table 4 (see Appendix G).

A correlation between amount of performance information and confidence in original ratings was conducted and a positive relationship was expected. However, the correlation was not significant, $r = .09$, ns. As mentioned previously, because amount of performance information was a dummy variable with minimal variance, all of the relevant manipulation check variables were also correlated with confidence. These correlations are provided in Table 2 (see Appendix G). Although most of the correlations were in the predicted direction, only number of behaviors observed was significantly positively related to confidence.

As in hypothesis 1, the absolute values of the average modification scores were used in the next analysis. Modification scores were correlated with confidence and the analysis indicated that, as predicted, confidence ratings were negatively related to modification of ratings although the correlation was not significant, $r = -.24$, $p = .09$. Thus, hypothesis 2 was not supported.

Hypothesis 3

Hypotheses 3a through 3d predicted an interaction between discrepancy and amount of performance information on raters' modification scores. To test hypotheses 3a through 3d, a 2 (low and high API) X 3 (over, under, in-agreement) ANOVA with repeated measures on the second factor was performed on the modification scores. The mean modification scores for each condition are found in Table 5 (see Appendix G). This hypothesis was not supported because the interaction between amount of performance information and discrepancy was not significant $F(2,94) = 1.43$, ns. The

results revealed a main effect for the discrepancy condition, $F(2,94)=27.95$, $p < .05$. To determine which means were significantly different, paired t-tests were conducted. The results indicated that the mean modification score in the underrater condition ($M=-.22$) was significantly lower than the mean in the in-agreement condition ($M=.007$), $t(48) = -2.88$, $p < .05$. The mean modification score in the in-agreement condition was also significantly lower than the mean in the overrater condition ($M=.35$), $t(48) = -5.1$, $p < .05$. These results mean that, regardless of the amount of performance information, participants modified their ratings downward in the underrater condition, did not change in the in-agreement condition, and modified their ratings upward in the overrater condition. This pattern of results was expected in the low amount of performance information condition only (Hypothesis 3c.)

Additional Analyses

Perceptions of Modification. A 2 (high and low amount of information) by 3 (over, under, and in-agreement) ANOVA with repeated measures on the second factor was performed on participants' responses to whether they changed their performance ratings after viewing their subordinates' self-ratings. Consistent with the results from hypothesis 3, the interaction was not significant, $F(2,64)=.29$, *ns*, and, the results indicated a main effect for discrepancy, $F(2,64)=4.89$, $p < .05$. To identify which means were significantly different, a paired t-test was conducted on each pair of means and revealed that participants responses in the underrater ($M=.48$) and overrater condition ($M=.50$) were both significantly greater than those in the in-agreement

condition ($M=.17$), $t(33)=2.48$, $p < .05$, and $t(34)=3.76$, $p < .05$, respectively.

The means in the over and underrater conditions were not significantly different, $t(47)=-.21$, ns. These results are congruent with the preceding analyses: participants changed their ratings in the direction of their subordinates' self-ratings in the over and underrater conditions and accurately reported that they changed their ratings in those conditions. Participants' reasons for why they did or did not change their ratings in the high and low API conditions may be found in Tables 6 and 7, respectively.

Change in Impression. Participants' responses to their impression of their subordinate after viewing his/her self-ratings was subtracted from their response to their impression of their subordinates before viewing their subordinates' self-ratings to yield a value indicative of the degree and direction of impression change for each subordinate. To test whether there were any differences in the amount of impression change, a two (high and low API) and 3 (over, under, and in-agreement) ANOVA with repeated measures on the second factor was performed. The results revealed no interaction $F(1,47)=1.29$, ns., or main effects for amount of performance information, $F(1,47)=1.83$, ns., or discrepancy $F(1,47)=3.57$, ns. The means for this variable in each condition are in Table 8.

Change in Confidence. To test whether participants in the low API condition increased their confidence ratings after viewing their subordinates' self-ratings more than raters in the high API condition a 2 (high and low API) by 2 (confidence in original and final ratings) ANOVA with repeated measures on the second factor was

performed. Before averaging across subordinates, two 2 (high and low API) by 3 (over, under, in-agreement) ANOVAs with repeated measures on the second factor were conducted on the mean first and second confidence ratings. The results indicated that there was no interaction or main effects. Therefore, the confidence ratings on the original performance ratings and the final performance ratings were averaged across subordinates for each rater. The results indicated a significant 2-way interaction, $F(1,46)=6.83, p < .05$. Subsequent paired t-tests indicated that the mean confidence ratings on the original performance ratings ($M=3.87$) were significantly lower than the mean confidence ratings on the final performance ratings for participants in the low API condition ($M=4.22$), $t(20)=-4.00, p < .05$. However, this difference in confidence ratings did not exist for participants in the high API condition, $t(26)=-1.51, ns.$, ($M = 4.01$ and $M = 4.10$ in the first and second confidence ratings, respectively). Therefore, confidence in ratings increased after viewing self-ratings for raters in the low API condition, but not in the high API condition.

Perceived Fairness. A multivariate ANOVA was conducted on the questions, responded to by all participants, which related to perceived fairness in order to detect differences between high and low API conditions. The results indicated a significant difference in these variables between high and low API condition $F(5,119)=4.43, p < .05$. The t-tests for independent samples for each item are shown in Table 9 (Appendix G). Subsequent regression analyses showed that participants' responses to "My supervisor collected enough information about my performance during the simulation to

make an accurate assessment” captured a significant amount of variance (49%) in perceived fairness, $F(1,138)= 136.8, p < .05$.

Re-analysis With Extreme Groups. Because amount of performance information as measured by number of interactions was a continuous variable, some supervisor-subordinate pairs had a relatively few number of interactions. Therefore, the manipulation of high amount of performance information may not have been as strong for some participants. Consequently, it was decided, a priori, to reanalyze the data using only those people in the high API condition whose mean number of interactions with their subordinates was five or greater. This value was chosen because five interactions indicates that the supervisors and subordinates interacted for at least 10 minutes more than those in the low API condition. Choosing five interactions as a cut-off also permitted a reasonable sample size for effects to be detected. Using this criteria, 8 people in the high API condition were excluded, resulting in a sample size of 20. The descriptive statistics on the modification scores in the full and partial sample may be found in Table 5 (see Appendix G).

For all of the manipulation checks and for hypothesis 1 and 2, the results using extreme groups were similar to those using the full sample. However, the analysis of hypothesis 3 resulted in a significant 2-way interaction between API and discrepancy on the raters’ modification scores, $F(2,78)=3.8, p < .05$. T-tests for independent samples were then conducted to test hypothesis 3a and 3b. Hypothesis 3a predicted that modification scores in the low API/overrater condition ($M=.317$) would be greater

than than those in the high API/overrater condition ($\underline{M} = .325$). This hypothesis was not supported, $t(39) = .08$, *ns*. Hypothesis 3b predicted that modification scores in low API/underrater condition ($\underline{M} = -.365$) will be lower than those in the high API/underrater condition ($\underline{M} = -.02$). This hypothesis was supported, $t(39) = -2.54$, $p < .05$.

To test hypothesis 3c and 3d paired sample t-tests were conducted on modification scores within each API condition. Hypothesis 3c predicted that within the low API condition, modification scores in the overrater condition ($\underline{M} = .325$) would be higher than those in the underrater ($\underline{M} = -3.65$) or in-agreement ($\underline{M} = .008$) conditions. Modification scores in the overrater condition were significantly higher than those in the underrater, $t(20) = -5.18$, $p < .05$, and in-agreement conditions, $t(20) = -4.26$, $p < .05$, providing support for this hypothesis. Hypothesis 3d predicted that within the high API condition, modification scores in the overrater condition ($\underline{M} = .317$) would be lower than those in the underrater ($\underline{M} = -.02$) or in-agreement conditions ($\underline{M} = .00$). Contrary to the hypothesis, modification scores in the overrater condition were significantly greater than those in the underrater, $t(19) = -2.5$, $p < .05$, and in-agreement conditions, $t(19) = -3.78$, $p < .05$.

Given the change in the results for hypothesis 3 with the partial sample, the impression change analysis was also conducted with the partial sample. The interaction between amount of performance information and discrepancy approached significance, $F(1,39) = 2.96$, $p = .09$. The means of impression change for each condition using the

partial sample are in Table 8. Subsequent paired t-tests were conducted to detect the nature of the interactive effect. The difference in the mean impression change in the low API/overrater condition ($M=.24$) and in the low API/underrater condition ($M=-.05$) approached significance, $t(20)=-1.83$, $p=.08$. None of the other means in the low API condition were significantly different. Within the high API condition, none of the means in impression change were significantly different. The means suggest that impression change was greater for participants in the low API / overrater condition than for participants in the high API / overrater condition. However, this difference did not occur in the underrater conditions.

The perceived fairness and change in confidence analyses were also performed again with the partial sample and provided the same results as those with the full sample. The relevance of these findings will be discussed below.

Discussion

Manipulation Checks

From the analysis of the manipulation checks, it is clear that the discrepancy manipulation was successful. Moreover, the mean number of interactions between supervisors and subordinates in the high API condition was significantly greater than the number of interactions (0) in the low API condition. This was supported by the coding of interactions, as well as by self-reported amount of time spent with supervisors and subordinates. However, to determine if number of interactions resulted in the transmission of a greater amount of performance information, participants used a

behavioral checklist to indicate how many performance-related behaviors were actually observed. Although the analysis of this manipulation check approached significance, the practical significance of the difference between these means is small, ($M=19.05$) and ($M=21.56$) in low and high API, respectively.

An explanation for this result is that the behaviors on the checklist were all worded positively (e.g., seeks information energetically) which was confounded with the number of behaviors actually observed. In fact, there is evidence to suggest that the number of behaviors observed is actually a proxy variable for performance rather than an accurate measure of the amount of information gathered by the supervisor. For example, the correlation between first performance ratings and the average number of behaviors observed across all subordinates was .34, $N=49$, $p < .05$. Therefore, a halo effect could have occurred in that raters with a positive impression of the ratee may have assumed that the ratee performed several of the other behaviors. It is likely that raters did not consider what they actually observed when checking off behaviors but, instead, what they believed the subordinate did based on their limited interaction. Moreover, this significant correlation also indicates that participants who felt their subordinate was unsuccessful did not check off many behaviors because they all represented positive characteristics. This means that raters may have collected more performance information than they reported due to the fact that all behaviors were worded positively.

Secondly, the pattern of correlations shown in the matrix in Table 2 demonstrates that number of behaviors observed is not significantly related to any of the other measures of amount of performance information. This provides further evidence that number of behaviors observed captured performance rather than amount of performance information. Taken together, these effects may account for the lack of significant difference in number of behaviors observed in low and high API conditions. However, based on the data collected on the number of actual interactions and the self-reported time spent with supervisors and subordinates, it is clear that supervisors and subordinates interacted more in the high API condition. Moreover, observers of the simulation noticed that supervisors and subordinates stayed on task and limited their conversations to topics related to the simulation. Therefore, it is logical to assume that more performance information was transmitted in the high API condition than in the low.

In future research, including negative behaviors on the behavioral checklist may help ameliorate the problem with the checklist in the present study. In this way, the amount of information collected, whether positive or negative, could be recorded. However, there is evidence from the social cognition literature that individual's overall impressions will cause them to check off all the behaviors that are congruent with their impressions regardless of whether they actually observed those behaviors or not (Cantor & Mischel, 1977). Coding the actual content of supervisor-subordinate interactions as

inobtrusively as possible may be the best way to measure how much performance information is transmitted to the supervisor.

Hypothesis 1

Hypothesis 1 predicted that amount of modification, in general, would be negatively related to amount of performance information. This hypothesis was unsupported largely due to the fact that raters in the high API condition changed their ratings in the direction of their overrating subordinates. An analysis of the reasons given by participants in the high API condition who changed their ratings reveals that the majority of the raters were unsure that they had obtained all the relevant information necessary to make accurate judgments (see Table 6). Therefore, they were willing to increase their rating. Because there was a lot of activity during the simulation, it is quite possible that the supervisor would not have been able to witness all of the behaviors of his/her subordinate. Therefore, the raters seemed to be willing to accept the ratees' high self-ratings. However, the same effect did not occur in the underrater condition. Raters in the high API condition did not change their ratings nearly as much for underraters as for overraters.

These results are inconsistent with those found by Blakely (1993). In Blakely's no-contest condition, the condition most similar to those in the present study, he found that raters in the underrater condition predominantly lowered their ratings to be more consistent with their subordinates' self-ratings. However, in the overrater condition, raters did not raise their ratings at all. Blakely (1993) concluded that the low self-

ratings provided informational value which caused raters to modify their ratings downwards, whereas raters in the overrater condition only changed their ratings when they thought the subordinate would contest discrepant ratings. Therefore, Blakely concluded that they only modified their ratings to avoid conflict with the subordinate.

Contrary to the findings of Blakely (1993), in the present study, raters in the high API condition only modified their ratings upward and did not modify their ratings downward, whereas in the low API condition raters' ratings changed towards the self-ratings of the subordinate regardless of the direction. Raters in the high API/underrater condition did not change their ratings perhaps because they were satisfied that what they viewed of the subordinate's performance was reflective of their original ratings and, therefore, were not persuaded by how the subordinate perceived his/her performance. In other words, the standards for their original ratings were met and information from the subordinate did not detract from the knowledge they already acquired. It is unclear why raters in the overrater condition changed their ratings in the present study but not in Blakely's study. One reason may be that Blakely's study involved an in-basket task which might have been less ambiguous than participants' performance during the Looking Glass simulation. Therefore, performance may have been able to be measured more objectively causing the raters to be unaffected by the subordinates' positive self-ratings. A more comprehensive explanation of the results of hypothesis 1 will be discussed in later sections.

Hypothesis 2

Hypothesis 2 predicted that amount of performance information would affect confidence which would then affect amount of modification in ratings. However, amount of performance information did not affect modification of ratings, so testing confidence as a mediator was unnecessary. Nonetheless, the relationship between these variables and confidence was investigated to clarify which parts of the hypothesized mediating relationship were unsupported. Number of behaviors observed was the only proxy variable for amount of performance information that was significantly related to confidence. Given the evidence that this variable measures performance rather than amount of performance information, it cannot be concluded that amount of performance information is related to confidence. It seems that the manipulation of this variable between-subjects was not strong enough to elicit significant differences in confidence.

This same result was found by Paese and Sniezek (1991). These researchers gave participants performance information (in the form of statistics) on minor league pitchers with which to make judgments about their current and future earned run averages (ERA's). In their first experiment, they manipulated amount of information as a between-subjects factor as was done in the present study. However, this manipulation did not cause differences in confidence between the two groups. Paese and Sniezek (1991) explained this finding by the fact that their participants were unaware "of the amount and quality of information potentially available to

them” (p.106). In their second experiment, amount of performance information was a within-subjects factor. Participants either received an increase in relevant performance information (relevant increase), a decrease in relevant performance information (relevant decrease), a increase in irrelevant information (irrelevant increase), or a decrease in irrelevant information (irrelevant decrease). The results suggested that confidence was positively related to amount of performance information and questions responded to after the experiment indicated that participants in the relevant information conditions were more confident in the judgments they made with relevant information than the baseline information only. This difference did not exist for participants in the irrelevant information conditions. Therefore, it seems that not only is amount of information important to increase confidence, but also the relevance of the additional information.

In planning the present experiment, it was expected that raters would be aware that they were not being provided with all the performance information they could have been due to the fact that during the simulation a movable divider separated the supervisors from the subordinates. It was clear to the raters that, during the simulation, no communication was allowed and that the subordinates were working to resolve as many problems and issues facing his/her division as possible. However, this manipulation may not have been effective because the supervisors may have thought that separating supervisors from subordinates was the way the simulation was always conducted and was just part of the challenge of the exercise. Moreover, the

experimenter asked them to make judgments on their subordinates' performance so they may have thought that the experimenter must have believed that they had enough information on which to make accurate assessments. In other words, just being given the power to make performance ratings may have made the participants confident that they could judge their subordinates' performance effectively.

Additional findings by Paese and Sniezek (1991) may provide another reason for this lack of difference in original confidence ratings. In one condition, after estimating players' current and future ERA's, participants were then asked to make a decision as to whether the player should be promoted to the major leagues, based on their prior judgments. Contrary to the researchers' predictions, judges who had to make a decision were not as overconfident as those who did not have to make a decision. The explanation provided for this finding was that the act of making a decision forced judges to think about the evidence supporting both alternatives. Support for this explanation is provided by Koriat, Lichtenstein, and Fischhoff (1980) who found that requiring participants to list reasons for and against an alternative before choosing it increased the appropriateness of their confidence levels. Moreover, those participants who did not have to provide disconfirming reasons for their chosen alternative were overconfident in their choices (Koriat et al., 1980).

Therefore, in the Paese and Sniezek study, those participants who were not forced to make a decision were likely to only consider confirming evidence for their judgments. Performance ratings in the present study could be considered analogous to

the ERA judgments. However, participants in the present study were not forced to make a decision (e.g., promotion or salary raise) based on their judgments. Thus, it is possible that raters' were overconfident in their judgments because they did not have to make the type of decision which would force them to think about disconfirming evidence (e.g., contradictory performance behaviors). Future research should investigate the relationship between amount of performance information and confidence when the rater needs to use their ratings to make an important decision that will affect the ratee.

The second part of the mediating relationship proposed in hypothesis 2 received partial support. Confidence did seem to be related to amount of modification of performance ratings, although the proportion of variance accounted for (6%) was quite small. Zalesny (1990) investigated this relationship and found that confidence was related to perceived change in ratings but not to actual change in ratings. In the present study, confidence in original ratings was unrelated to perceived change in ratings. Taken together, the results of the present study and those of Zalesny (1990) do not lend support to the idea that confidence is strongly related to modification of ratings. However, there was not much variability in confidence in either of these studies. Before concluding that confidence is unrelated to modification of ratings, it is necessary to manipulate confidence effectively so that a relationship may be detected if one exists. The weak relationship detected in the present study provides hope that this relationship

does exist, but future research should concentrate on producing more variability in confidence to provide more power to detect the relationship.

Hypothesis 3

Hypothesis 3 predicted an interaction between discrepancy and amount of performance information on the modification of performance ratings. Contrary to the hypotheses, a main effect for discrepancy was found such that supervisors tended to change their ratings in the direction of their subordinates' self-ratings. These results were expected in the low API condition only. One reason for this result is that amount of performance information was truly a continuous variable for those participants in the high amount of performance information condition. Participants were strongly encouraged to interact with their supervisor and subordinates but it was impossible to control interaction beyond that without negatively affecting the simulation. Therefore, it is possible that those raters who interacted very little with their subordinates contributed to the insignificant interaction because their pattern of responses may have been more similar to what was expected by participants in the low API condition. To investigate this possibility, those raters whose mean number of interactions across all their subordinates was less than 5 were excluded from the analysis and the results revealed a significant interaction. It seems that, as expected, excluding those raters from the analysis resulted in a reduction in the mean amount of modification in the overrater and underrater conditions. However, the predicted nature of the interaction was not entirely supported.

Hypothesis 3a predicted that raters' modification scores in the low API/overrater condition would be greater than those in the high API/overrater condition. This hypothesis was not supported. An explanation for this result was discussed in the section on hypothesis 1. Hypothesis 3b predicted that raters' modification scores in the low API /underrater condition would be lower than the raters' modification scores in the high API /underrater condition. This hypothesis was supported because raters modified their ratings downwards after viewing lower self-ratings in the low API condition but not in the high API condition. Raters in the low API/underrater condition were probably not as sure as raters in the high API/underrater condition that the sample of performance they saw was reflective of the subordinate's performance during the simulation, thereby causing them to change their ratings in the direction of the subordinates'.

The reanalysis also showed that the pattern of modification scores in the underrater, in-agreement, and overrater conditions was different in the high and low API conditions. Specifically, for the low API condition, the mean modification score in the overrater condition was greater than the in-agreement condition which was greater than the underrater condition, supporting hypothesis 3c.

Contrary to hypothesis 3d, the mean modification score in the overrater condition was greater than the in-agreement condition for the high API condition. However, the difference between the means in the in-agreement and underrater conditions was not significant. Hypothesis 3d predicted that, for the high API

condition, the mean modification score in the overrater condition would be lower than in the in-agreement condition which would be lower than the underrater condition.

Hypothesis 3d was predicted based on literature demonstrating a negative relationship between self-promotion tactics and supervisory affect and/or performance ratings (Ferris et al., 1994; Godfrey et al., 1986; Wayne & Ferris, 1990). It was assumed that self-ratings that were relatively more favorable than the supervisor's ratings would be perceived as self-promoting to the supervisors. Therefore, it was expected that this would cause negative perceptions of the ratee and result in a downwards modification of performance ratings. As shown in Table 6, only one person cited that they modified their ratings downward because the ratee thought so highly of him/herself. In fact, the results of the present study indicated the opposite result in that supervisors in the high API /overrater condition modified their ratings upwards.

A possible explanation for this result is provided by the findings of Giacalone (1985). Giacalone predicted that entitlements (those who claimed responsibility for positive events) would be considered immodest and to lack humility. These perceptions were then expected to lead to negative reactions towards the subordinate. However, the results revealed that raters did not pay attention to entitlements unless they were disconfirmed by a credible source (Giacalone, 1985). Thus, due to the short-term nature of the simulation, it is possible that raters in the present study were too busy to collect information that disconfirmed their subordinate's high ratings and, therefore, accepted them. It is likely that supervisors in actual organizations would have access to

more information that could discount entitlements made by their subordinates. Thus, having more performance information in this circumstance may negatively impact subordinates who inflate their self-ratings. The demands on the supervisor's time due to the nature of the task in this study may have precluded them from learning any disconfirming evidence. This, in turn, may have prevented them from perceiving the subordinate so negatively as to modify their ratings downwards.

Additional Analyses

The additional analyses performed help to clarify several of the findings from the tests of the formal hypotheses. For example, the change in impression results with the partial sample suggest that, although raters changed their ratings of their subordinates in the overrater conditions, the change in ratings was not related to a change in impression for those raters in the high API condition. It may be that the raters changed their ratings only to be fair to the ratee rather than because they truly believed the ratee's self-ratings. That is, from the raters' point of view, changing their ratings indicated that it was possible that the subordinate performed better than they had originally thought, but the change in impression results suggest that the raters were not entirely convinced. Conversely, the upwards modification of ratings by raters in the low API condition seemed to reflect more of a change in their impression of the ratee presumably because their original ratings were based on limited information. Therefore, a strong impression may not have been formed which allowed for a change in impression to occur.

Surprisingly, neither raters in the high or low API conditions changed their impression of their subordinates in the underrater conditions (this result was true for full and partial samples). This is especially unexpected for the raters in the low API because they actually modified their ratings downward more than did those in the high API condition (from partial sample results). Because the experimenter instructed the raters that it was important to obtain the most accurate picture of the subordinate's ratings, the raters may have trusted the subordinate's ratings because they did not obtain that much information about the person and, in trying to rate accurately, modified their ratings downward. However, the subordinate's self-ratings may not have been salient enough to change their original overall impression of the subordinate which was based on their actual interactions.

The interaction between low and high API and first and second confidence ratings provides more insight into the impact of the subordinates' self-ratings on the raters' final performance ratings. These results suggest that, although confidence in original ratings did not differ for participants in the high and low API conditions, participants in the low API condition reported higher confidence levels on their final performance ratings after viewing their subordinates' self-ratings. This effect is congruent with the other results from this study. Raters who received minimal performance information changed their ratings more, changed their impression (overrater condition only), and became more confident in their final ratings after receiving self-ratings from their subordinate. However, raters in the high API

condition did not change their impression of their subordinate nor did the self-ratings significantly improve their confidence in their final performance ratings.

It is interesting to compare these results with those of Paese and Sniezek (1991). As mentioned previously, these researchers found that global confidence ratings were not only affected by the amount of information, but also by the relevance of that information. It could be argued that for raters in the high API condition, the additional information provided in the form of subordinate's self-ratings was not considered relevant which would explain why they did not change their impression of the subordinate or their confidence in their final performance ratings. Those raters in the high API condition may not have considered their subordinates' self-ratings to be valuable sources of performance information because they may have felt that they personally had collected enough information to make an accurate assessment of their subordinates' performance. They may have changed their ratings just to be fair to the subordinate and/or because they could not think of any concrete reasons to justify their original ratings. Conversely, those raters in the low API condition may have considered their subordinate's self-ratings to be a relevant additional source of performance information and, therefore, felt more confident in their final ratings.

Finally, one effect of different amounts of performance information collected by supervisors was illustrated by the results of the items relating to the subordinates' perceived fairness of the appraisal process. As expected, subordinate's in the low API condition disagreed more with the statement "My supervisor collected enough

information about my performance to make an accurate assessment.” These subordinates also felt that their peers could have evaluated their performance better than their supervisors and that their self-ratings were probably more accurate than their supervisors’ ratings. It is likely that the responses to these questions accounted for the fact that subordinates in the high API condition also believed the appraisal process was more fair than did subordinates in the low API condition. In fact, subordinates’ responses to whether they felt their supervisor collected enough information about their performance captured a significant amount of variance in perceived fairness.

Taken together these results are congruent with previous findings (Greenberg, 1986; Landy, Barnes, & Murphy, 1978). Greenberg (1986) found that rater familiarity with ratee’s work was a procedural factor related to the perception of fair performance appraisals. Moreover, Landy et al. (1978) found a correlation between ratings of fairness and items measuring the supervisor’s knowledge of job performance. It seems evident that limited information on the subordinate’s performance greatly impacts subordinate’s perceived fairness of the appraisal and, consequently, is likely to impact their acceptance of feedback from their supervisors. Clearly, this lack of acceptance could seriously affect the extent to which an employee seeks to improve performance in the areas that his/her supervisor has criticized which could, in turn, impede the employee’s development and career progress.

Limitations

The present study has several limitations worth noting. For example, replicating supervisor-subordinate interactions in an organizational environment is clearly difficult to do in a laboratory study. However, this study attempted to simulate an organizational environment as well as possible by creating an office-like environment and by using a task in which participants were each assigned unique roles in a very realistic simulated organization. Although participants were involved in this task for approximately four hours (more than most studies involving performance appraisal), they still were not exposed to all the factors that affect supervisors' ratings in actual organizations.

For example, raters in this study were not in any way accountable for their ratings and there was no threat of future interactions with the subordinate about their performance ratings. This situation is certainly not representative of the way performance appraisals work in actual organizations. However, the focus of the study was to identify the factors which accounted for a change in supervisors' ratings when faced with discrepant self-appraisal information. When conducting research such as this, it is often necessary to strip away various other external variables to be able to identify the effect of one's variables of interest on the dependent variable. Once these relationships are established, investigating the effects of other variables may be a valuable endeavor.

Another limitation of this research is the way amount of performance information was manipulated. Manipulating amount of performance information as a between-subjects factor was ineffective in this study as well as in Paese and Sniezek (1991). In both studies, amount of performance information did not affect confidence levels in the predicted manner. Future research studying the effect of amount of performance information on confidence should attempt to manipulate this within-subjects to increase the chance that a relationship will be found.

Conducting a study with actual supervisors whose amount of interaction with two or more of their subordinates is significantly different may be a way to address the limitations of the present study described above. It is likely that the amount of performance information available to the rater would be related to confidence which may then be related to a change in ratings. Moreover, it may be possible to design a study in which supervisor's ratings would only be known to the experimenter so that accountability and other external factors would not confound the results.

Conclusions

The majority of the findings of this study have generated questions to be answered by future research before any general conclusions may be made. For example, the relationship between amount of performance information and confidence in original ratings and confidence and modification in ratings should be investigated further to understand how important confidence is in the performance appraisal process. Based on the results from hypothesis 3 and the reasons given by supervisors, the extent

to which supervisor's are concerned with being fair to the subordinate may be another variable that is related to the degree of rating modification. As was found in the high API /overrater condition, raters did not change their impression of the subordinate but did change their ratings perhaps because they wanted to be fair to the subordinate by using his/her self-ratings in making their final judgments. The way in which a change in ratings, but not a change in impression impacts a subordinate in the future is another area for additional research.

Based on the results regarding raters' change in impression and change in confidence, it seems that, in general, raters in the low API condition considered their subordinate's self-ratings to be more relevant sources of performance information than did raters in the high API condition. Given that many supervisors and subordinates in actual organizations do work closely with one another, these results call into question whether, in these situations, self-ratings have any value at all other than to provide the ratee with the perception that they have some input into their evaluation. Another question generated by this study is why participants in the low API condition changed their impression when they modified their ratings upwards (overrater condition) but not downwards (underrater condition)? Although reasons were speculated, additional empirical research should investigate this anomalous finding.

One last point that is important to mention is the fact that, although significant differences in amount of modification were detected between the conditions, the overall magnitude of these means was relatively small. For example, the largest mean of

modification scores was .36 for raters in the high API /overrater condition (full sample) which means that only approximately 1 in 3 raters in this condition changed their overall rating at all. These results are similar to those found by Blakely (1993). The mean manipulation of discrepancy was 2 and one-third points versus 2 points in the present study and Blakely also used a 9-point scale to present self-ratings and to collect supervisors' performance ratings. The mean modification score was -.29 with a standard deviation of 1.07. As shown in Table 3, the means in the present study for the over and underrater conditions are very comparable. In general, it seems that people are not very likely to change their original ratings regardless of how much they know about the subordinate. Perhaps future research should explore what individual difference characteristics are related to the likelihood that a rater will change his/her ratings.

The fact that, in general, raters did not adjust their ratings to be more congruent with their subordinates' self-ratings calls into question the usefulness of 360-degree feedback instruments. Several organizations use 360-degree feedback instruments for evaluation purposes (Timmreck & Bracken, 1996). An assumption of 360-degree feedback is that several sources of information will capture different facets of one's performance (Conway & Huffcutt, 1996). Therefore, combining these different perspectives to result in a final evaluation should enhance the accuracy and fairness of the performance evaluation. However, the results of the present study suggest that raters are not significantly influenced by subordinates' self-ratings. Therefore,

companies may want to weigh the costs and benefits of using 360-degree feedback instruments. For example, if raters do not necessarily change their original ratings to reflect the ratings from other sources of feedback, then the time and money invested in implementing 360-degree feedback programs may not be worth the minimal change in performance ratings.

Instead, it may be beneficial for organizations to try to identify certain departments in which morale is low or productivity is especially low so that the effectiveness 360-degree feedback instruments will be maximized. In these situations, it is likely that the feedback from different sources would facilitate discussion about problems within the department and may actually be a first step in resolving these problems.

The results of this study also suggest that 360-degree feedback instruments which are based solely on a standard rating format are not likely to facilitate impression change on the part of an individual who receives ratings from another source which are discrepant from his/her original ratings. In fact, several researchers have investigated under what conditions raters will change their existing impression of a ratee. The tendency to maintain one's original impression of a ratee has been called the "assimilation effect" because subsequent performance information is assimilated into your original impression. On the contrary, the "contrast effect" occurs when incongruent subsequent information causes a change in the rater's original impression. Murphy, Garnett, Herr, and Chen (1986) found that raters only recalled performance

behaviors which were congruent with their original impression (assimilation effect) when memory demands were increased.

Moreover, Foti and Hauenstein (1993) found that raters tended to seek out impression-inconsistent behaviors from the ratee. However, the extent to which they used these impression-inconsistent behaviors depended on how taxed the raters were during observation of subsequent performance behaviors. The contrast effect occurred for those raters who were minimally taxed during observation, whereas the assimilation effect occurred for those raters who had to complete another task while observing the ratee's performance (maximally taxed) (Foti & Hauenstein, 1993).

Given the small means on modification scores in the present study, it seems that the assimilation effect occurred. The results of research on impression change (Foti & Hauenstein, 1993; Murphy, et al., 1986) provide some insight into why raters tended to maintain their original impression of the ratee in the present study. Because the nature of the task in the present study was designed to be realistic, participants were given more problems and issues to resolve than the time allowed. Hence, it is safe to assume that participants in the high API condition were taxed during the time they could have observed their subordinates' performance. Therefore, they may have observed impression-inconsistent behaviors but, because their attention was divided, these behaviors did not change their impression of the subordinate.

In conclusion, it seems that raters need to be presented with impression-inconsistent behaviors that they will be able to remember in order for their original

impression of a ratee to change. Given the fact that rater's attention is often divided during the observation of their subordinate's performance, examples of performance behaviors need to be presented to the supervisor at the time of rating to increase the likelihood that their impression of their subordinate will change. Therefore, if a 360-degree feedback instrument allowed the opportunity for ratees to describe examples of their prior behaviors which were inconsistent with the rater's prior impression, the rater's prior impression would be more likely to change. However, most 360-degree feedback instruments rely on numerical ratings and/or have a standard list of behaviors which raters identify as a "strength" or "needs improvement". Presumably, 360-degree feedback instruments are useful to the extent that individuals use the information provided from other sources. It seems that information from other sources will be more likely to be used if the information has caused an actual change in the rater's impression of that ratee. Therefore, 360-degree instruments may want to incorporate a written component so that actual examples of ratees' prior performance-related behaviors could be presented to the rater.

In general, it seems that more research on the usefulness of 360-degree feedback systems is warranted so that suggestions for improving these instruments to maximize their benefits for the organization could be made.

References

- Ashford, S. J. (1989). Self assessments in organizations: A literature review and integration model. Research in Organizational Behavior, 11, 133-174.
- Atwater, L. E., & Yammarino, F. J. (in press). Antecedents and consequences of self-other rating agreement: A review and model. In Ferris, J. (Ed.), Research in Personnel and Human Resources Management.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51 (6), 1173-1182.
- Bassett, G. A., & Meyer, H. H. (1968). Performance appraisal based on self-review. Personnel Psychology, 21, 421-430.
- Blakely, G. L. (1993). The effects of performance rating discrepancies on supervisors and subordinates. Organizational Behavior and Human Decision Processes, 54, 57-80.
- Borman, W. C. (1974). The rating of individuals in organizations: An alternative approach. Organizational Behavior and Human Performance, 12, 105-124.
- Cantor, N. & Mischel, W. (1977). Traits as Prototypes: Effects on Recognition Memory. Journal of Personality and Social Psychology, 35 (1), 38-38.
- Cawley, B. D., & Levy, P. E. (1995). Performance rating discrepancies: A direct test of multiple explanations. Paper presented at Tenth Annual Conference of the Society for Industrial/Organizational Psychology, Orlando, Florida.

Conway, J. M., & Huffcutt, A. I. (1996). Testing Assumptions of 360-degree feedback: A meta-analysis of supervisor, peer, subordinate, and self-ratings. Paper presented at the Eleventh Annual Conference of the Society for Industrial/Organizational Psychology, San Diego, CA.

Dipboye, R. L., & dePontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. Journal of Applied Psychology, *66* (2), 248-251.

Farh, J., & Dobbins, G. H. (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. L., Dobbins, G. H., & Cheng, B. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U. S. Workers. Personnel Psychology, *44*, 129-147.

Farh, J. L., Werbel, J. D., & Bedeian, A. G. (1988). An empirical investigation of self-appraisal-based performance evaluation. Personnel Psychology, *41*, 141-156.

Favero, J. L., & Ilgen, D. R. (1989). The effects of ratee prototypicality on rater observation and accuracy. Journal of Applied Social Psychology, *19* (11), 932-946.

Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Organizational Behavior and Human Decision Processes, *58*, 101-135.

Folger, R., Konovsky, M. A., & Cropanzano, R. (1992). A due process metaphor for performance appraisal. Research in Organizational Behavior, *14*, 129-177.

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

Fox, S. & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. Personnel Psychology, 41, 581-592.

Fried, Y., Tieggs, R. B., & Bellamy, A. R. (1992). Personal and interpersonal predictions of supervisor's avoidance of evaluating subordinates. Journal of Applied Psychology, 77, (4) 462-468.

Giacalone, R. A. (1985). On slipping when you thought you had put your best foot forward: Self-promotion, self-destruction, and entitlements. Group and Organization Studies, 10(1), 61-80.

Godfrey, D. K., Jones, E. E., & Lord, C. G. (1986). Self-promotion is not ingratiating. Journal of Personality and Social Psychology, 50 (1), 106-115.

Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. Journal of Applied Psychology, 71 (2), 340-342.

Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. Journal of Management, 20 (4), 737-756.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, 41, 43-62.

Heneman, H. G., III. (1974). Comparisons of self-and superior ratings of managerial performance. Journal of Applied Psychology, 59, 638-642.

Heneman, R. L., & Wexley, K. N. (1983). The effects of time delay in rating and amount of information observed on performance rating accuracy. Academy of Management Journal, 26 (4), 677-686.

Holzbach, R. L. (1978). Rater bias in performance ratings: Supervisor, self and peer ratings. Journal of Applied Psychology, 63 (5), 579-588.

Judge, T. A., & Ferris, G. R. (1993). Social context of performance evaluation decisions. Academy of Management Journal, 36 (1), 80-105.

Kacmar, K. M., Carlson, D. S., Wright, P. M., & McMahan, G. C. (1995). Assessing the factors influencing differences between supervisor and subordinate performance ratings: A multiple sample study. Paper presented at the Tenth Annual Conference of the Society for Industrial/Organizational Psychology, Orlando, Florida.

Kacmar, K. M., Delery, J. E., & Ferris, G. R. (1992). Differential effectiveness of applicant impressions management tactics on employment interview decisions. Journal of Applied Social Psychology, 22 (16), 1250-1272.

Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. Journal of Applied Psychology, 59 (4), 445-451.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for Confidence. Journal of Experimental Psychology: Human Learning and Memory, 6 (2), 107-118.

Landy, F.J., Barnes, J.L., & Murphy, K.R. (1978). Correlates of perceived fairness and accuracy of performance evaluation. Journal of Applied Psychology, 63, 751-754.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 67 (3), 280-296.

Meyer, H., Kay, E., & French, J. R. P. (1965). Effects of threat in a performance appraisal interview. Journal of Applied Psychology, 49, 311-317.

Murphy, K. R., & Cleveland, J. V. (1991). Performance appraisal: An organizational perspective. Allyn and Bacon: Boston, MA.

Murphy, K. R., & Cleveland, J. V. (1995). Understanding performance appraisal: Social organizational and goal-based perspectives. Sage Publications: Thousand Oaks, CA.

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

Nilsen, D., & Campbell, D. P. (1993). Self-observer rating discrepancies: Once an overrater, always an overrater? Human Resource Management, 32 (2&3), 265-281.

Oskamp, S. (1965). Overconfidence in case-study judgments. Journal of Consulting Psychology, 29, 261-265.

Paese, R. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. Organizational Behavior and Human Decision Processes, 48, 100-130.

Parker, J. W., Taylor, E. K., Barrett, R. S., & Martens, L. (1959). Rating scale content: III. Relationship between supervisory and self-ratings. Personnel Psychology, 12, 49-63.

Pearce, J. L., & Porter, L. W. (1986). Employee responses to formal performance appraisal feedback. Journal of Applied Psychology, 71 (2), 211-218.

Porter, P. P., & Foti, R. J. (1995). Effect of discrepant information and sex of manager on attributions and ratings of manager's performance: A laboratory and field study. Manuscript submitted for publication.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. Journal of Applied Psychology, 75 (3), 322-327.

Schmitt, N., Noe, R. A., & Gottshalk, R. (1986). Using the lens model to magnify rater's consistency, matching, and shared bias. Academy of Management Journal, 29 (1), 130-139.

Schrader, B. W., & Steiner, D. D. (1995). Effects of different comparison standards on agreement in performance ratings. Poster presented at the Tenth Annual Conference of the Society for Industrial/Organizational Psychology, Orlando, Florida.

Shore, L. M., & Thornton, G. C. (1986). Effects of gender on self- and supervisory ratings. Academy of Management Journal, 29 (1), 115-129.

Steel, R. P., & Ovalle, N. R. (1984). Self-appraisal based upon supervisory feedback. Personnel Psychology, 37, 667-685.

Thornton, G. C. (1968). The relationship between supervisory and self-appraisals of executive performance. Personnel Psychology, 21, 441-455.

Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. Personnel Psychology, 33, 263-271.

Timmreck, C. W., & Bracken, D. W. (1996). Multisource assessment: Reinforcing the preferred "means" to the end. Paper presented in W.W. Tornow (Chair), 360-Feedback Systems: Another Look at Their Uses and Impact. 11th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Tornow, W. W. (1993). Editor's Note: Introduction to special issue on 360-degree feedback. Human Resource Management, 32, (2&3), 211-219.

VanVelsor, E., Taylor, S., & Leslie, J. B. (1993). An examination of the relationships among self-perception accuracy, self-awareness, gender and leader effectiveness. Human Resource Management, 32 (2&3), 249-263.

Wayne, S. J., & Ferris, G. R. (1990). Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. Journal of Applied Psychology, 75 (5), 487-499.

Williams, J. R. & Levy, P. E. (1992). The effects of perceived system knowledge on the agreement between self-ratings and supervisor ratings. Personnel Psychology, 45, 835-847.

Wohlers, A. J., & London, M. (1989). Ratings of managerial characteristics: Evaluation difficulty, co-worker agreement, and self-awareness. Personnel Psychology, 42, 235-261.

Yammarino, F. J., & Atwater, L. E. (1993). Understanding self-perception accuracy: Implications for human resource management. Human Resource Management, 32 (2&3), 231-247.

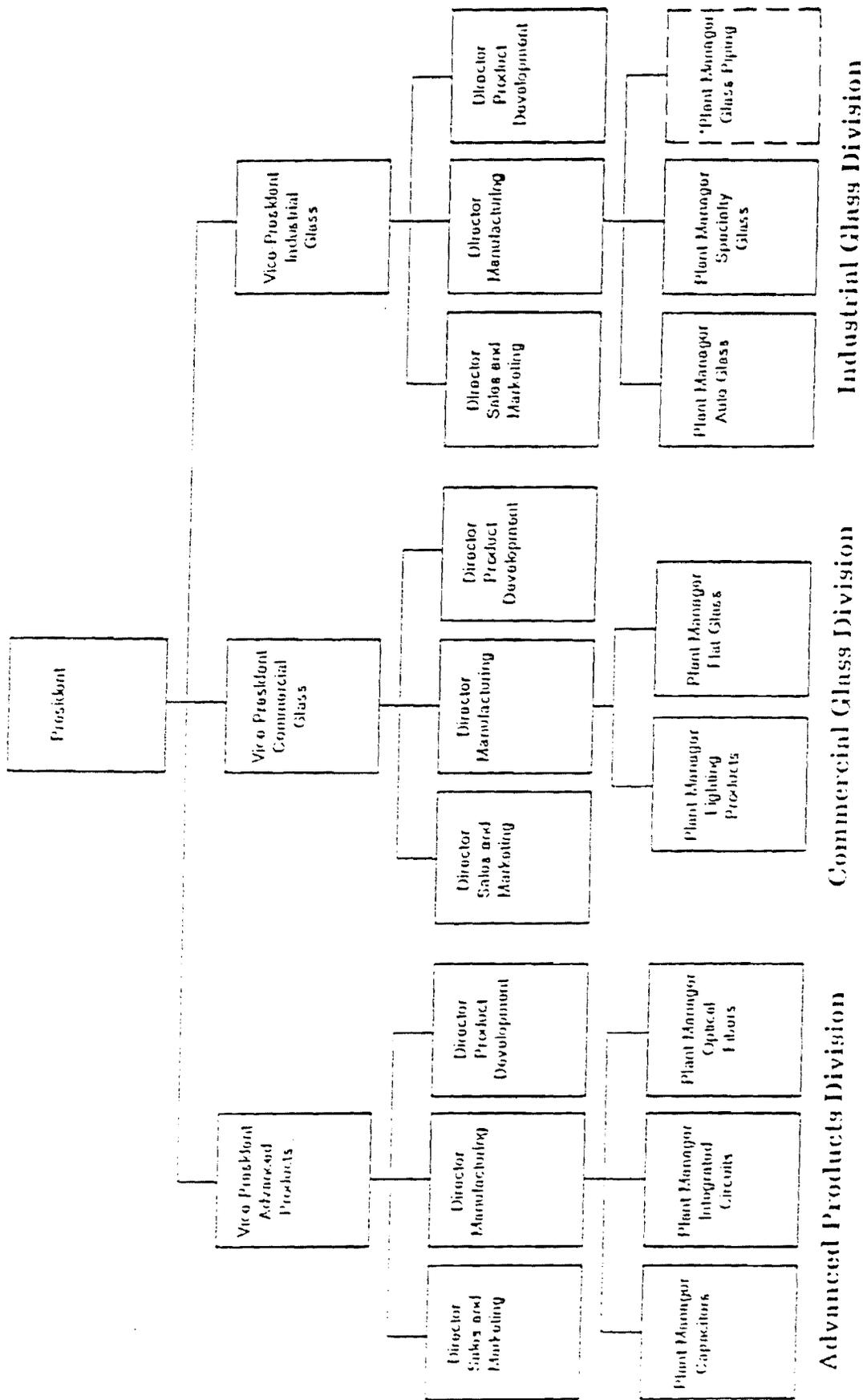
Yu, J., & Murphy, K. R. (1993). Modesty bias in self-ratings of performance: A test of the cultural relativity hypothesis. Personnel Psychology, 46, 357-363.

Zammuto, R. F., London, M., Rowland, K. M. (1982). Organization and rater differences in performance appraisals. Personnel Psychology, 35, 643-658.

Zalesny, M. D. (1990). Rater confidence and social influence in performance appraisals. Journal of Applied Psychology, 75 (3), 274-289.

APPENDIX A
ORGANIZATIONAL CHART

I.O.O.K.I.N.G. G.L.A.S.S., I.N.C.[®]



* This position is intentionally vacant

APPENDIX B
ADAPTED SKILLSCOPE

PERFORMANCE APPRAISAL FORM

The purpose of this evaluation is to gather accurate information about how participants performed during the Looking Glass simulation. The first section of the evaluation asks you to rate the individual you supervised during the Looking Glass simulation on several managerial characteristics. The second section asks you how confident you are that your ratings reflect your subordinate's true performance during the Looking Glass simulation.

SECTION I

INSTRUCTIONS: Circle the number that best represents your subordinate's performance during the Looking Glass simulation on each characteristic. Each managerial performance dimension is defined below. Please read this definition carefully before making your rating.

1. *Problem Identification*: Definition: defining problems effectively; spotting problems, opportunities and threats early; seeking information energetically; creating order out of the large quantity of information.

Of all the possible problems facing your subordinate, rate how many your subordinate identified effectively.

1	2	3	4	5	6	7	8	9
None		Some		Quite a bit		An extreme amount		All

2. *Decision Making*: Definition: Reaching conclusions in a timely manner, acting on judgment, taking actions and committing resources; implementing decisions and following through, weighing consequences of contemplated action carefully.

Of all the possible decisions your subordinate could have made, rate how many your subordinate made effectively.

1	2	3	4	5	6	7	8	9
None		Some		Quite a bit		An extreme amount		All

3. Planning and Organizing: Rate how **effective** your subordinate was at the following:
Establishing a course of action for self/others to accomplish specific goals; allocating resources to execute plans within required timeframes; handling situations in which there is no prescribed method for proceeding; translating strategy into action.

1	2	3	4	5	6	7	8	9
Not at all		Fairly		Effective		Very	Extremely	
Effective		Effective				Effective	Effective	

4. Leadership: Rate how **effective** your subordinate was at the following:
Demonstrating foresight; capitalizing on opportunities to move forward into the future; having a clear idea of where the company should go and conveying that idea to others; delegating work effectively; working well with other people over whom he/she had no direct authority.

1	2	3	4	5	6	7	8	9
Not at all		Fairly		Effective		Very	Extremely	
Effective		Effective				Effective	Effective	

5. Time Management: Rate how **effective** your subordinate was at the following:
setting priorities (e.g., distinguishing between important and unimportant tasks); making the most of the time available (e.g., being productive); dealing with interruptions appropriately.

1	2	3	4	5	6	7	8	9
Not at all		Fairly		Effective		Very	Extremely	
Effective		Effective				Effective	Effective	

6. Overall: Rate the **overall effectiveness** of your subordinate during the Looking Glass simulation.

1	2	3	4	5	6	7	8	9
Not at all		Fairly		Effective		Very	Extremely	
Effective		Effective				Effective	Effective	

**PERFORMANCE APPRAISAL FORM
SECTION II**

INSTRUCTIONS: For each performance dimension described above, please indicate how confident you are that your rating reflects your subordinate's true performance during the Looking Glass simulation. Feel free to reread the definitions of the performance dimensions to refresh your memory.

Please rate how confident you are that your rating on each of the following dimensions reflects your subordinate's **true** performance during the Looking Glass simulation.

1. Problem Identification

1	2	3	4	5
Not at all confident		Moderately confident		Extremely confident

2. Decision Making

1	2	3	4	5
Not at all confident		Moderately confident		Extremely confident

3. Planning and Organizing

1	2	3	4	5
Not at all confident		Moderately confident		Extremely confident

4. Leadership

1	2	3	4	5
Not at all confident		Moderately confident		Extremely confident

5. Time Management

1	2	3	4	5
Not at all confident		Moderately confident		Extremely confident

6. Overall Effectiveness

1	2	3	4	5
Not at all confident		Moderately confident		Extremely confident

SELF-APPRAISAL FORM

The purpose of this evaluation is to gather accurate information about how participants performed during the Looking Glass simulation. You are asked to evaluate your own performance during the Looking Glass simulation.

INSTRUCTIONS: Circle the number that best represents your performance during the Looking Glass simulation on each characteristic. Each managerial performance dimension is defined below. Please read this definition carefully before making your rating.

1. *Problem Identification*: Definition: defining problems effectively; spotting problems, opportunities and threats early; seeking information energetically; creating order out of the large quantity of information.

Of all the possible problems presented to you, rate how many you identified effectively.

1	2	3	4	5	6	7	8	9
None		Some		Quite a bit		An extreme amount		All

2. *Decision Making*: Definition: Reaching conclusions in a timely manner, acting on judgment, taking actions and committing resources; implementing decisions and following through, weighing consequences of contemplated action carefully.

Of all the possible decisions you could have made, rate how many you made effectively.

1	2	3	4	5	6	7	8	9
None		Some		Quite a bit		An extreme amount		All

3. Planning and Organizing: Rate how **effective** you were at the following:

Establishing a course of action for self/others to accomplish specific goals; allocating resources to execute plans within required timeframes; handling situations in which there is no prescribed method for proceeding; translating strategy into action.

1	2	3	4	5	6	7	8	9
Not at all Extremely Effective		Fairly Effective		Effective		Very Effective		

4. Leadership: Rate how **effective** you were at the following:

Demonstrating foresight; capitalizing on opportunities to move forward into the future; having a clear idea of where the company should go and conveying that idea to others; delegating work effectively; working well with other people over whom he/she had no direct authority.

1	2	3	4	5	6	7	8	9
Not at all Effective		Fairly Effective		Effective		Very Effective	Extremely Effective	

5. Time Management: Rate how **effective** you were at the following:

Setting priorities (e.g., distinguishing between important and unimportant tasks); making the most of the time available (e.g., being productive); dealing with interruptions appropriately.

1	2	3	4	5	6	7	8	9
Not at all Effective		Fairly Effective		Effective		Very Effective	Extremely Effective	

6. Overall: Rate your **overall effectiveness** during the Looking Glass simulation.

1	2	3	4	5	6	7	8	9
Not at all Effective		Fairly Effective		Effective		Very Effective	Extremely Effective	

APPENDIX C
INFORMED CONSENT FORM

Informed Consent Form

Looking Glass, Inc. Managerial Simulation

Psyc. HSC 1127-96

Investigators - Paige Porter, M.S. and Roseanne Foti, Ph.D.

IRB Ref. 96-075

I. Purpose

The Industrial/Organizational Psychology Department is conducting research sponsored by the Center for Creative Leadership. The purpose of the study is to collect data on a performance appraisal instrument developed to measure managerial characteristics necessary for successful performance on the University Edition - Looking Glass, Inc. managerial simulation.

II. Procedures

As participants in this research you will take the role of an upper-level manager in a simulated organization called the Looking Glass, Inc. developed by the Center for Creative Leadership, a management consulting firm. You will be given several memos and papers to review before the simulation begins to familiarize you with your position in the organization and the problems you will be faced with. You will interact with other members of the organization to solve the problems embedded in your memos. The preparation time is approximately 1 hour and the entire simulation takes about 3 hours. At the end of the simulation you will be asked to rate your performance during the simulation and the performance of your subordinate (when applicable). Subsequently, you will complete some questionnaires as a part of the research and will be dismissed. Again, the preparation will take approximately one hour and will be completed independently prior to the beginning of the simulation. The session will last approximately four hours and will be conducted in Squires Student Center.

III. Risks

There are no risks associated with participation in this research.

IV. Benefits of this Project

Participating in the Looking Glass, Inc. simulation is a very relevant and useful experience for individuals seeking careers in management or in multi-level organizations. Participants will gain experience in a in-basket exercise which is frequently used by organizations in managerial assessment centers. Moreover, the Looking Glass, Inc. is a shortened version of the management training exercise used by the Center of Creative Leadership and various other certified organizations to assess and develop managers. Participating in the Looking Glass simulation will provide a realistic experience of the day in the life of an upper-level manager in a multi-level organization. The benefits gained from experience with this simulation will depend on the effort and conscientiousness of the individual participant. Essentially, this experience will expose participants to a management assessment tool that they may be subjected to in their future careers.

V. Extent of Anonymity and Confidentiality

Participants will not be identifiable after the study is complete. Participants responses will be identified with code numbers rather than names or social security numbers. At no time will the researchers release the results of the study to anyone other than the individuals working on the project without your written consent.

VI. Compensation

Participants will be paid approximately \$7.00 per hour. At the end of the simulation session, each participant will be paid \$35.00 for his/her participation in this research. This compensation includes a thorough preparation before the simulation begins (approximate preparation time - 1 hour) and approximately four hours of time to complete the simulation, the performance appraisals, and questionnaires.

VII. Freedom to Withdraw

You are free to withdraw from this study at any time and will be compensated for your participation up until the time you decide to withdraw.

VIII. Approval of Research

This research project has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic Institute and State University, by the Department of Psychology.

IX. Subject's Responsibilities

I voluntarily agree to participate in this study. I have the following responsibilities:

1. Prepare for the simulation by reading the in-basket memos.
2. Participate in the Looking Glass, Inc. simulation.
3. Complete performance appraisals on self and others (when applicable) and a few short questionnaires.

X. Subject's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project.

If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

Signature

Date

Should I have any questions about this research or its conduct, I may contact:

Investigator	Paige Porter	961-2748
Faculty Advisor	Roseanne J. Foti, Ph.D.	231-5814
Chair, IRB	E.R. Stout	231-9359

APPENDIX D
PARTICIPANT'S SCHEDULE OF EVENTS

SCHEDULE

- 9:15 - 9:30 Check-in, Coffee, Bagels
- 9:30 - 10:10 Meetings with Supervisors (see attached meeting schedule)
- 10:10 Begin Simulation
- 11:30 Begin Final Meetings with Supervisors (see attached meeting schedule)
- 12:10 End Simulation
- 12:10-12:20 President’s State of the Company Address
- 12:20-12:35 Conduct Self and Subordinate Performance Ratings (where applicable)
- 12:35 The following positions will meet to complete questionnaires outside the simulation room.

Advanced Products Division

Director, Sales and Marketing
 Director, Product Development

Commercial Glass Division

Director, Sales and Marketing
 Director Product

Development

Plant Manager, Capacitors

Plant Manager, Lighting

Products

Plant Manager, Integrated Circuits
 Plant Manager, Optical Fibers

Plant Manager, Flat Glass

Industrial Glass Division

Director, Sales and Marketing
 Director, Product Development
 Plant Manager, Auto Glass
 Plant Manager, Specialty Glass

All other positions will complete questionnaires at their desks.

These are:

President, Vice President, APD; Vice President, CGD; Vice President, IGD; Director of Manufacturing, APD; Director of Manufacturing, CGD; Director of Manufacturing, IGD

Approx. 1:30

Session Ends.

Meeting Schedule

Abbreviation Key

APD - Advanced Products Division	VP - Vice President	S&M - Sales and Marketing	w/ - with
CGD - Commercial Glass Division	Dir. - Director	Manuf. - Manufacturing	
IGD - Industrial Glass Division	PM - Plant Manager	Prod. Dev. - Product Development	

Time of Meeting^^

	<u>Advanced Products Division</u>	<u>Commercial Glass Division</u>	<u>Industrial Glass Division</u>
9:30, 11:30	President meets w/ VP-APD Dir. of Manuf., APD meets w/ PM, Capacitors	VP-CGD meets w/ Dir. S&M, CGD Dir. of Manuf., CGD meets w/ PM, Lighting Products	VP-IGD meets w/ Dir.of S&M, IGD Dir. of Manuf. meets w/ PM, Auto Glass
9:40, 11:40 Glass	VP-APD meets w/ Dir. S&M,APD Dir. Manuf., APD meets w/ PM Integrated Circuits	President meets w/ VP-CGD Dir. Manuf., CGD meets w/ PM Flat Glass	VP-IGD meets w/ Dir. of Prod. Dev. Dir. Manuf. meets w/ PM Specialty
9:50, 11:50	VP-APD meets w/ Dir. Prod. Dev., APD Dir. Manuf., APD meets w/ PM Optical Fibers	VP-CGD meets w/ Dir. Prod. Dev., CGD	President meets w/ VP-IGD
10:00, 12:00	VP-APD meets w/ Dir. Manuf., APD	VP-CGD meets w/ Dir. Manuf., CGD	VP-IGD meets w/ Dir. Manuf., IGD

^^It is important that these meetings take place at the scheduled time and for 10 minutes each. Please make every effort to stay on task and finish the meeting exactly when it is scheduled to end. Thank you.

APPENDIX E
SUMMARY RATING FORM

Supervisor's Position Title _____
 Subordinate's Position Title _____

SUMMARY RATING FORM

Performance Dimension	Supervisor's Rating	Subordinate's Rating
Problem Identification	_____	_____
Decision Making	_____	_____
Planning and Organizing	_____	_____
Leadership	_____	_____
Time Management	_____	_____
Overall	_____	_____
TOTAL*	A. _____	minus B. _____ =
C. _____		

*Sum the supervisor's rating on each dimension and put on line A. Also, sum the subordinate's rating on each dimension and put on line B.

Subtract the value on line B from line A. (i.e., $A - B = ?$) and place on line C.

APPENDIX F
MANIPULATION CHECK QUESTIONNAIRES

Questionnaire Completed by Supervisors for Each One of Their Subordinates

SECTION I

The following are characteristics of successful managers. Indicate whether or not you observed your subordinate perform behaviors that are representative of each of the following characteristics.

	Observed	Not Observed
1. Seeks information energetically.	_____	_____
2. Probes, digs beneath the surface, tests the validity of information.	_____	_____
3. Creates order out of large quantities of information.	_____	_____
4. Keen observer of people, events, things.	_____	_____
5. Defines problems effectively; gets to the heart of a problem.	_____	_____
6. Spots problems, opportunities, threats, trends early.	_____	_____
7. Logical, data-based, rational.	_____	_____
8. Action-oriented; presses for immediate results.	_____	_____
9. Decisive; doesn't procrastinate on decisions.	_____	_____
10. Troubleshooter; enjoys solving problems.	_____	_____
11. Implements decisions, follows through, follows up well; an expediter.	_____	_____
12. Carefully weighs consequences of contemplated action.	_____	_____
13. Establishes and conveys a sense of purpose.	_____	_____
14. A team builder: brings people together successfully around tasks.	_____	_____
15. Structures subordinates' work appropriately.	_____	_____
16. Resourceful; can marshal people, funds, space required for projects	_____	_____

	Observed	Not Observed
17. Manages the process of decision-making effectively; knows who to involve on what issues.	_____	_____
18. Can organize and manage big, long-term projects.	_____	_____
19. Recognizes and rewards people for their work.	_____	_____
20. Can easily handle situations where there is no pat, answer, no prescribed method for proceeding.	_____	_____
21. Can translate strategy into action over the long haul.	_____	_____
22. Inspirational; helps people to see the importance of what they are doing.	_____	_____
23. Good at promoting an idea or vision; persuading.	_____	_____
24. Possesses extensive network of contacts necessary to do the job.	_____	_____
25. Astute sense of "politics."	_____	_____
26. Able to inspire, motivate people; sparks others to take action.	_____	_____
27. Comfortable with the power of the managerial role.	_____	_____
28. Skilled at selling upward, influencing superiors.	_____	_____
29. Delegates effectively.	_____	_____
30. Works effectively with other people over whom he or she has no direct authority.	_____	_____
31. Sets priorities well; distinguishes clearly between important unimportant tasks.	_____	_____
32. Makes the most of the time available; extremely productive.	_____	_____
33. Deals with interruptions appropriately; knows when to admit interruptions and when to screen them out.	_____	_____
34. Avoids spreading self too thin.	_____	_____

SECTION II

1. Rate how different you felt your subordinate's self-ratings of performance were from your original ratings of his or her performance?

1	2	3	4	5
Not at all different		Moderately different		Extremely different

2. On the following scale, rate whether your subordinate's self-ratings of performance were higher, in-agreement, or lower than your original ratings of his or her performance.

1	2	3	4	5
Self-ratings were lower		Self-ratings were the same		Self-ratings were higher

3. Indicate below how well you knew your subordinate BEFORE participating in this simulation.

1	2	3	4	5
Did not know at all		Have had a few interactions in the past		Knew extremely well

4. Did you change your rating of the subordinate after viewing his/her self-ratings?

_____ YES _____ NO

Please list reasons below why you did or did not change your rating. (Your reasons are especially important to this research -- do not leave blank).

5. What was your impression of your subordinate immediately after you completed the first performance appraisal?

1	2	3	4	5
Poor performer		Average performer		Excellent performer

6. What was your impression after you viewed his/her self-rating and rated him/her the second time?

1	2	3	4	5
Poor performer		Average performer		Excellent performer

FINAL QUESTIONNAIRE COMPLETED BY ALL PARTICIPANTS

1. What percentage of your time during the Looking Glass simulation did you spend with the following people:

Supervisor	_____
Peers	_____
Alone	_____
Other	_____

2. In your opinion, how much information about your performance was your supervisor able to obtain during the simulation?

1	2	3	4	5
None		A Moderate Amount		A Large Amount

3. Rate the extent to which you agree with the following statements.

a. My supervisor collected enough information about my performance during the simulation to make an accurate assessment.

1	2	3	4
Strongly Disagree	Disagree	Agree	Strongly Agree

b. I feel that the appraisal process was fair.

1	2	3	4
Strongly Disagree	Disagree	Agree	Strongly Agree

c. My peers in the organization could have made more accurate ratings of my performance than my supervisor.

1	2	3	4
Strongly Disagree	Disagree	Agree	Strongly Agree

d. My self-ratings were probably more accurate than my supervisors'.

1	2	3	4
Strongly Disagree	Disagree	Agree	Strongly Agree

APPENDIX G**TABLES**

Table 1
Descriptive Statistics for Manipulation Check Variables

Variable	<u>M</u>	<u>SD</u>	Min	Max
<u>Manipulation Checks</u>				
1. "Were self-ratings different?"				
High API Condition				
Underrater	3.36	.83	2.00	5.00
In-agreement	1.20	.52	1.00	3.00
Overrater	3.14	1.04	1.00	5.00
Low API Condition				
Underrater	2.95	.97	1.00	5.00
In-agreement	1.47	.92	1.00	4.00
Overrater	2.95	1.07	1.00	5.00
2. "Were self-ratings higher, lower, or in-agreement with your original ratings?"				
High API Condition				
Underrater	1.36	.49	1.00	2.00
In-agreement	2.95	.52	1.00	4.00
Overrater	4.43	.88	1.00	5.00
Low API Condition				
Underrater	1.62	1.02	1.00	5.00
In-agreement	3.00	.65	1.00	4.00
Overrater	4.48	.87	2.00	5.00
3. Mean # of coded interactions across all subordinates for High API condition	10.26	7.07	.00	24.33
4. Mean reported % of time spent with subordinates.				
High API	20.07	4.41	10.00	40.00
Low API	18.41	3.52	5.00	30.00
5. Mean reported % of time spent with supervisors.				
High API	25.78	14.96	.00	80.00
Low API	16.67	11.53	.00	50.00

Table 1 (continued)

Descriptive Statistics for Manipulation Check Variables

Variable	<u>M</u>	<u>SD</u>	Min	Max
6. Mean # of behaviors observed				
High API Condition				
Underrater	21.21	6.66	8.00	31.00
In-agreement	21.45	8.21	2.00	33.00
Ovrrater	22.14	6.40	9.00	33.00
Low API Condition				
Underrater	17.62	6.76	5.00	29.00
In-agreement	17.40	5.23	10.00	26.00
Ovrrater	21.52	7.28	8.00	34.00

Table 2

Intercorrelations Between Amount of Performance Information (API), API Proxy Variables, Mean First Performance Ratings Across Subordinates, and Confidence

Variable	1	2	3	4	5	6	7
1. API Condition ^a --		.69* (49)	.60* (42)	.29 ^b (41)	.25 (49)	.15 (49)	-.06 (49)
2. # of Coded Interactions		--	.53* (42)	.25 (41)	.14 (49)	.02 (49)	-.07 (49)
3. Time Spent w/ Subordinates			--	.36* (36)	.06 (42)	.18 (42)	-.01 (42)
4. Time Spent w/ Supervisor				--	.06 (41)	.11 (41)	.02 (41)
5. # of Behaviors -- Observed					--	.37* (49)	.33* (49)
6. Confidence in Original Ratings						--	.48* (49)
7. First Perform. Ratings							--

* $p < .05$

Note. ^aAPI Condition is the dummy coded variable for the between subjects factor (Low=1, High=2)

^b $p = .06$

Table 3

Descriptive Statistics for Performance Ratings and Modification Scores

Variable	<u>M</u>	<u>SD</u>	Min	Max	Reliability
First Performance Ratings Averaged Across 5 Performance Dimensions					
Underrater	6.18	1.40	2.33	8.83	.91
In-agreement	6.17	1.19	3.67	8.00	.91
Overrater	6.35	1.58	2.67	9.00	.95
First Overall Performance Rating					
Underrater	6.27	1.64	2.00	9.00	
In-agreement	6.29	1.27	3.00	8.00	
Overrater	6.35	1.77	2.00	9.00	
Final Performance Ratings Averaged Across 5 Performance Dimensions					
Underrater	5.96	1.48	2.17	8.83	.93
In-agreement	6.18	1.24	3.67	8.67	.92
Overrater	6.70	1.50	3.00	9.00	.95
Final Overall Performance Rating					
Underrater	6.06	1.60	2.00	9.00	
In-agreement	6.31	1.35	3.00	9.00	
Overrater	6.76	1.61	3.00	9.00	
Modification Scores Across 5 Performance Dimensions					
Underrater	-.23	.52	-1.200	1.200	
In-agreement	.00	.28	-1.200	1.200	
Overrater	.33	.39	-.400	1.400	
Modification scores for Overall Performance Dimension					
Underrater	-.20	.64	-1.00	2.00	
In-agreement	.02	.32	-1.00	1.00	
Overrater	.41	.61	.00	3.00	

Table 4
Descriptive Statistics on Confidence Variables

Variable	<u>M</u>	<u>SD</u>	Min	Max	<u>N</u>
Average confidence ratings on first performance ratings across over and underrater conditions					
High API	4.01	.45	3.17	4.92	28
Low API	3.91	.63	2.25	4.67	21
Average confidence ratings on first performance ratings across all discrepancy conditions					
High API	4.02	.43	3.06	4.92	28
Low API	3.87	.61	2.39	4.67	21
Average confidence ratings on final performance ratings across over and underrater conditions					
High API	3.98	.56	2.50	4.83	27
Low API	4.19	.57	2.92	5.00	21
Average confidence ratings on final performance ratings across all discrepancy conditions					
High API	4.10	.45	3.33	4.83	27
Low API	4.22	.51	2.92	5.00	21

Table 5

Descriptive Statistics for Modification Scores in Each Condition.

Variable	<u>M</u>	<u>SD</u>	Min	Max
Modification Scores				
High Perf. Info. / Full Sample				
Underrater	-.12	.56	-1.167	1.167
In-agreement	.01	.36	-1.167	1.167
Ovrrater	.36	.42	-.333	1.333
Low Perf. Info.				
Underrater	-.37	.40	-1.000	.000
In-agreement	.01	.06	-.167	.167
Ovrrater	.33	.34	.000	1.000
High Perf. Info. / Partial Sample				
Underrater	-.02	.48	-1.000	1.167
In-agreement	.00	.19	-.333	.667
Ovrrater	.32	.40	-.333	1.333

Table 6
Reasons For Why Participants in High Amount of Performance Information Condition Did Or Did Not Change Their Performance Ratings.

Response to whether rater changed rating or not.	Reason	Frequency	%age
YES	Rater thought he/she might have missed something so decided to re-evaluate.	12	15.8
YES	Rater felt he/she had rated too high and had given person benefit of the doubt.	8	10.5
YES	Rater increased rating because ratee underrated his/her performance.	4	5.3
YES	Rater was not confident because he/she had not observed some behaviors.	2	2.6
YES	Rater averaged his/her original ratings with the subordinates'.	2	2.6
YES	Rater decreased rating because ratee rated him/herself too high.	1	1.3
NO	Rater felt they were right in their original rating.	11	14.48
NO	Rater thought ratee was better than the ratee thought.	6	7.9
NO	Ratings were correlated so thought differences were due to cultural differences.	2	2.6
NO	Rater thought subordinate didn't see room for improvement.	4	5.3
NO	Ratings were the same.	9	11.8
NO	Rater was confident in his/her original rating.	4	5.3
	Uninterpretable response or no reasons given.	11	14.48
	TOTAL	76	

Table 7

Reasons For Why Participants in Low Amount of Performance Information Condition Did Or Did Not Change Their Performance Ratings.

Response to whether rater changed rating or not.	Reason	Frequency	%age
YES	Rater was not confident that he/she observed subordinates actions closely enough to pass strict judgment.	7	12.9
YES	The low ratings made by ratee made rater rethink his/her own judgment.	3	5.6
YES	Rater averaged his/her original ratings with the subordinates'.	2	3.7
YES	Rater thought that past interactions with the subordinate affected his ratings unfairly.	1	1.9
NO	Rater was confident/sure of his/her original rating.	13	24.1
NO	Ratings were the same, strengthening the rater's opinion.	8	14.8
NO	Rater did not think anyone could perform so well in such a short exercise.	7	12.9
	Uninterpretable response or no reasons given.	13	24.1
	TOTAL	54	

Table 8

Amount of Impression Change as a Function of Discrepancy and Amount of Performance Information Condition

Amount of Performance Information	Discrepancy		
	Underrater	In-agreement	Ovrrater
Full Sample			
High			
<u>M</u>	-.07	.07	.00
<u>SD</u>	.47	.26	.27
Low			
<u>M</u>	-.05	.24	.24
<u>SD</u>	.67	.77	.44
Partial Sample			
High			
<u>M</u>	.00	.10	-.05
<u>SD</u>	.46	.31	.22

Table 9

Comparison of High and Low Amount of Performance Information on Variables Relating to Perceived Fairness of the Appraisal Process

Variable	Amount of performance information	M	SD	n	df	t-value
My supervisor collected enough information about my performance during the simulation to make an accurate assessment.	High	2.81	.714	80	138	-2.92
	Low	2.45	.743	60		
I feel that the appraisal process was fair.	High	2.89	.638	80	138	-2.59
	Low	2.60	.647	60		
My peers in the organization could have made more accurate ratings of my performance than my supervisor.	High	2.31	.832	80	138	4.24
	Low	2.87	.702	60		
My self-ratings were probably more accurate than my supervisor's ratings.	High	2.62	.586	80	138	3.32
	Low	2.97	.637	60		
^b In your opinion, how much information about your performance was your supervisor able to obtain during the simulation.	High	3.48	.898	80	138	-1.54
	Low	3.23	.947	60		

* $p < .05$.

Note. ^a These variables were responded to on a 4-point scale (1=Strongly Disagree, 4=Strongly Agree)

^b This variable was responded to on a 5-point scale (1=None, 5=A Large Amount)

PAIGE P. PORTER

SCHOOL ADDRESS

917 University City Blvd. Apt. D6
Blacksburg, Virginia 24060
(540) 961-2748

PERMANENT ADDRESS

2503 Kingman Drive
Wilmington, Delaware 19810
(302) 478-3393

EDUCATION

Doctor of Philosophy in Psychology, (Candidate), Expected May 1996

Concentration: Industrial/Organizational Psychology,
Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA

Dissertation: The Influence of Discrepant Perceptions of Performance and Amount of Performance Information on Future Performance Judgments

Master of Science Degree in Psychology, May 1994

Concentration: Industrial/Organizational Psychology
Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA

Thesis: Effect of Discrepant Information and Sex of Manager on Attributions and Ratings of Manager's Performance

Bachelor of Arts Degree in Psychology Education; Minor: History, May 1992.

University of Delaware, Newark, Delaware
Graduated Magna Cum Laude: 3.89 average on a 4.0 scale

HONORS/AFFILIATIONS

Phi Beta Kappa Honor Society (Spring 1992)
Psi Chi National Psychology Honor Society, Treasurer (Fall 1990)
Dean's List (eight semesters)
Student Member of the American Psychological Association
Student Member of the Society for Industrial-Organizational Psychology
Member of Graduate Honor Court, Virginia Tech (Fall 1994 to present)

PROFESSIONAL EXPERIENCE

Intern-Employee Selection, GTE Telephone Operations World Headquarters, Irving, Texas. Wrote Employment Testing Guidelines for company-wide use. Developed structured interviews by performing job analysis and data analysis, conducting phone interviews and observations, organizing and facilitating meetings with subject matter experts to develop questions, and writing technical reports. Conducted structured interview training course and designed supplemental written materials for a video-based structured interview training course. Developed a paper-and-pencil knowledge selection test. (May 1995 - December 1995)

Human Resources Intern, Department of Health and Social Services for State of Delaware, New Castle, Delaware. Designed questionnaire to measure government employee satisfaction with the Human Resources offices throughout the state of Delaware. Collected and analyzed data. Provided a written report of the results. (May 1994 - August 1994)

PROFESSIONAL EXPERIENCE (continued)

Personnel Intern, Alfred I. duPont Institute, Wilmington, Delaware.

Developed job evaluation system for management-level positions, participated in revamping performance appraisal system, revised job descriptions, made reference checks for prospective employees, assisted personnel director with Americans with Disabilities Act cases and compensation concerns. (May 1994 - August 1994)

TEACHING/ADVISING EXPERIENCE

Instructor, Virginia Tech, Blacksburg, Virginia.

Have full responsibility for sophomore-level class in Social Psychology. Prepared and delivered lectures, wrote tests, held office hours to assist students on a one-on-one basis. (January 1996 - Present)

Academic Advisor, Virginia Tech, Blacksburg, Virginia.

Assisted undergraduate students in choosing curricula and courses, gave presentations to students, wrote articles for newsletters, developed surveys and analyzed results, developed advising manual. (August 1993 - May 1995)

Graduate Teaching Assistant, Virginia Tech, Blacksburg, Virginia.

Lab instructor for Introduction to Psychology course, led in-class discussions, graded quizzes and essays, held office hours. (August 1992 - May 1993)

Student Teacher, Brandywine High School, Wilmington, Delaware

Planned and implemented lesson plans for high school seniors in psychology, anthropology, and peer facilitator training courses. (February 1992 - May 1992)

RESEARCH EXPERIENCE

Dissertation Research - Roseanne J. Foti, Ph.D (Chair), Virginia Tech

Designed and implemented laboratory research using MBA students to understand what factors promote a rater's use of self-appraisal information when performance judgments are required. Wrote proposal and received grant to assist in data collection.

Independent Research

Designed and conducted a study comparing theories of impression formation.

Designed and conducted a study exploring the effects of the repressor personality style on reactions to negative feedback.

Thesis Research - Roseanne J. Foti, Ph.D (Chair), Virginia Tech

Designed and implemented a laboratory and field study analyzing the effect of discrepant information and sex of manager on performance ratings and attributions made for performance.

Research Assistant - Robert Eisenberger, Ph.D, University of Delaware

Assisted graduate students and faculty member in entering and analyzing data.

PRESENTATIONS/AWARDS

Porter, P. P. & Foti, R. J. (1995, September) Recipient of the 1995-1996 \$10,000 Looking Glass University Edition Research Grant for "The Influence of Discrepant Perceptions of Performance and Amount of Performance Information on Future Performance Judgments." Awarded by the Center for Creative Leadership, Greensboro, North Carolina.

Porter, P. P. & Foti, R. J. (1995, March) Effect of discrepant information and sex of manager on attributions and ratings of manager's performance. Poster presented at the 1995 Graduate Research Symposium. Third place winner. Virginia Tech, Blacksburg, Virginia.

Porter, P. P. & Foti, R. J. (1995, May) Effect of discrepant information and sex of manager on attributions and ratings of manager's performance. In P. E. Levy (Chair), Feedback from Multiple Sources: Evidence from the Laboratory and Field. Symposium conducted at the 10th Annual Conference of the Society for Industrial/Organizational Psychology, Orlando, Florida.

MANUSCRIPTS

Porter, P. P. & Foti, R. J. (1995). Effect of discrepant information and sex of manager on attributions and ratings of manager's performance. Manuscript submitted for publication.

Porter, P. P. & Foti, R. J. (1995). Individual differences in the processing of negative feedback: Repressors vs. non-repressors. Manuscript submitted for publication.

RELEVANT COURSEWORK

Organizational Staffing
 Industrial Psychology I and II
 Organizational Psychology I and II
 Advanced Psychometric Theory
 Quantitative Topics in Applied Psychology
 Job Analysis and Classification
 Research Methods
 General Statistics
 Psychological Measurement
 Cognitive Psychology

COMPUTER SOFTWARE EXPERIENCE

Experience using MS-DOS, UNIX, SPSS-X (mainframe), SPSS 6.0 for Windows, SAS (Windows, PC, and mainframe versions), Word for Windows, WordPerfect 5.1 (DOS & Windows versions), Freelance Graphics, Powerpoint, Lotus 1-2-3. Excel, PC's, and Macintosh computers.

Paige P. Porter