This is to certify that the
dissertation entitled

LENIENCY BIAS IN THE PERFORMANCE APPRAISAL PROCESS: A PERSON X SITUATION PERSPECTIVE

presented by

BRAD ANTHONY CHAMBERS

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Industrial/Organizational Psychology

Major Professor's Signature

4/17/03

Date

LENIENCY BIAS IN THE PERFORMANCE APPRAISAL PROCESS: A PERSON X SITUATION PERSPECTIVE

By

Brad Anthony Chambers

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Psychology

2003

ABSTRACT

LENIENCY BIAS IN THE PERFORMANCE APPRAISAL PROCESS:
A PERSON X SITUATION PERSPECTIVE

By

Brad Anthony Chambers

Research on inaccuracy in performance ratings has generally focused on rating leniency—i.e., rating someone higher than he/she deserves. As a result of a long history of leniency research, scientists and practitioners have uncovered a variety of influential individual difference characteristics and situational factors that tend to relate to rating leniency. However, leniency research to date focuses on one class of variables to the exclusion of the other—i.e., it focuses on individual differences to the exclusion of situational factors, or it focuses on situational factors to the exclusion of individual differences. Despite the fact that psychologists have long argued the importance of studying human behavior from a person x situation perspective, leniency research does not examine these two classes of variables simultaneously. The current research examined the individual difference characteristics of self-monitoring, agreeableness, conscientiousness, empathy, and ego concern, as well as the situational factors of rater accountability and the presence/absence of a formal grievance policy in a laboratory study. A series of research hypotheses predicting person x situation interactions were advanced and tested using hierarchical linear modeling. Although none of the hypotheses advanced received full support, implications and directions for future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Managers gather employee performance ratings in order to make a variety of important decisions. An analysis of 100 organizations conducted by Cleveland, Murphy, and Williams (1989) revealed that performance evaluations are used to make decisions regarding salary, promotion, termination, and training needs. When these decisions are based on inaccurate or incomplete information, poor decisions will likely be made.

Inaccuracy in performance evaluations has been the topic of research for years. Researchers have documented rating effects such as halo (e.g., Cooper, 1981), primacy and recency (e.g., Farr, 1973), perceived similarity (e.g., Wexley, Alexander, Greenawalt, & Couch, 1980), physical attractiveness (e.g., Stone, Stone, & Dipboye, 1992), personal liking (e.g., Cardy & Dobbins, 1986; Dobbins & Russell, 1986), and leniency (e.g., Jawahar & Williams, 1997; Kane, Bernardin, Villanova, & Peyrefitte, 1995) just to name a few. Appropriately, Dipboye, Smith, and Howell (1994) point out that rating tendencies such as these are not in and of themselves *biases*; rather, they are rating *effects* that may lead to inaccurate ratings if their results are unfounded. Of all the forms of rating inaccuracy, leniency appears to be the most robust and most frequently studied form of inaccuracy in performance ratings.

Traditionally, researchers have examined leniency in performance ratings by looking at either the rater (e.g., Bernardin, Cooke, & Villanova, 2000; Jawahar, 2001; Ralston & Waters, 1996) or the situation (e.g., Judge & Ferris, 1993; Klimoski & Inks, 1990; Shapiro, 1975) in isolation from one another. For example, we know that on average raters' levels of conscientiousness tend to relate negatively to leniency in the ratings that they provide (Bernardin et al., 2000). We also know that requiring raters to

justify to ratees the performance ratings that they assign tends to lead to more lenient ratings on average (Klimoski & Inks, 1990). Thus, we know quite a bit about how individuals can influence ratings, and how situations can influence ratings, but we do not know how these two general classes of factors influence ratings simultaneously. In order to truly understand human behavior, we must look at both the person and the situation simultaneously (Eysenck, 1997). Individuals do not behave in vacuums, and it is unreasonable to assume that different individuals respond to the same situations in the same exact manner.

Recent theory and research has begun to view the act of providing performance ratings as a goal-directed behavior (Murphy & Cleveland, 1995), and they use this reasoning to explain the presence of inaccuracy—and most commonly leniency—in performance ratings. That is, individuals may possess a variety of objectives when they provide performance ratings, and behaving in a manner consistent with some of these different objectives may lead to lenient ratings. To date, little empirical research exists on this subject. Rather, Bjerke, Cleveland, Morrison, and Wilson (1987, as cited in Murphy & Cleveland, 1995) and Longenecker, Sims, and Gioia (1987) provide only anecdotal accounts of the goals and objectives pursued by raters. Despite the lack of empirical research, however, these anecdotal accounts describe what happens when raters are motivated to achieve particular objectives. Finally, these different goals and objectives that raters may seek to achieve when assigning performance ratings can come from within themselves (e.g., a rater wants a particular individual to like him/her) or they can be influenced by the environment (e.g., a rater's own pay is contingent on providing accurate ratings).

One situational or environmental characteristic that has been researched rather extensively in the leniency literature is whether a rater is held accountable for their ratings—typically operationalized by having raters justify ratings in one-on-one feedback discussions with ratees. Intuitively, when raters are accountable to ratees for the ratings that they provide, they tend to rate more leniently compared to situations in which raters are not accountable to ratees (e.g., Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975). Research has not examined the effects of holding raters accountable to others (e.g., his/her peers or superiors)—or, more interestingly, holding raters accountable to *both* the ratee *and* the rater's peers and/or superiors—on the leniency of the ratings that they provide. This latter scenario (i.e., holding raters accountable to ratees *and* the raters' peers/superiors) is especially interesting because it gives raters two competing objectives when assigning performance ratings for poor performers. Trying to satisfy only the ratee leads to leniency (as previous research demonstrates), and trying to satisfy only the rater's peers/superiors would likely lead to accuracy. However, it is somewhat ambiguous how a rater would deal with being in a situation in which he/she needs to satisfy the ratee *and* the rater's peers/superiors simultaneously.

Logically, the rater who is motivated to achieve two incompatible objectives when rating the performance of others has two options. First, he/she can favor one objective to the exclusion of the other and rate in a manner consistent with this favored objective. Alternatively, he/she can make a compromise and try to accomplish *both* objectives to an extent. Drawing from the multiple goals literature, the current research seeks to demonstrate that raters motivated to achieve two incompatible objectives behave

in a manner consistent with the latter strategy (i.e., compromise). When assigning this compromised performance rating, however, the important question becomes "Where will the rating actually fall?"

In addition to the situational factor of whether and to whom a rater is accountable, a second situational variable that is likely to influence the ratings that raters provide is whether individuals (either ratees themselves, or perhaps even peers or senior management) are given the opportunity to challenge the performance ratings provided. Formal grievance policies are often included in performance evaluation systems as a way to increase perceived fairness of these systems (Folger, Konovsky, & Cropanzano, 1992; Greenberg, 1986). Past research has not examined the effects of this situational characteristic on rating leniency to date, but it is likely to have implications for the ratings that raters provide (as detailed in later sections of this paper).

Situational characteristics alone cannot explain entirely the rating behavior of individuals; one must also consider the individuals making these ratings. That is, in order to understand rating behavior more completely, one must examine the interplay between the rater and the environment. Consider the previous example of providing a poorly performing ratee the opportunity to contest or challenge the ratings that a rater provides for him/her. Without considering the rater's personality, one might infer that he/she would provide lenient ratings for a poor performer so as to avoid confrontation. However, one might also infer that the rater would provide accurate (low) ratings for a poor performer to let the individual know how he/she really feels. Considering the rater's level of agreeableness might provide better insight into the action that the rater would actually take (i.e., the rating that he/she would actually provide) in such a situation.

Specifically, a rater with a high level of agreeableness would likely take the former route (i.e., provide an inflated rating so as to avoid conflict), whereas a rater with a low level of agreeableness would likely take the latter route (i.e., provide an accurate, low rating). This idea is discussed further in a later section.

How the situation and individual rater affect the rating process are key issues when it comes to designing effective, non-biased rating systems and selecting managers who will eventually use such rating systems. By examining both individual characteristics *as well as* situational characteristics, this research allowed for the examination of person by situation interactions on rating behavior.

Drawing from theory and research on multiple goals, the current research examines how individual difference variables and situational characteristics influence the performance ratings—specifically, the leniency in these ratings—which raters provide for poorly performing others. This research focuses only on ratings of poorly performing others because that is where the problem of leniency exists—there is no need to rate a good performer leniently (Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975). A conceptual model for this research is presented in Figure 1. As shown in the figure, the specific performance rating provided by a rater is a function of three classes of variables: rater characteristics, ratee characteristics, and the situation. Included in the rater characteristics category are factors such as rater individual differences (e.g., personality), values, beliefs, and personal goals. Included in the ratee characteristics category are factors such as ratee performance as well as ratee personality. Finally, situational characteristics include specific policies and methods relating to the rating system as well as organizational practices regarding performance ratings.

Figure 1. Conceptual Model

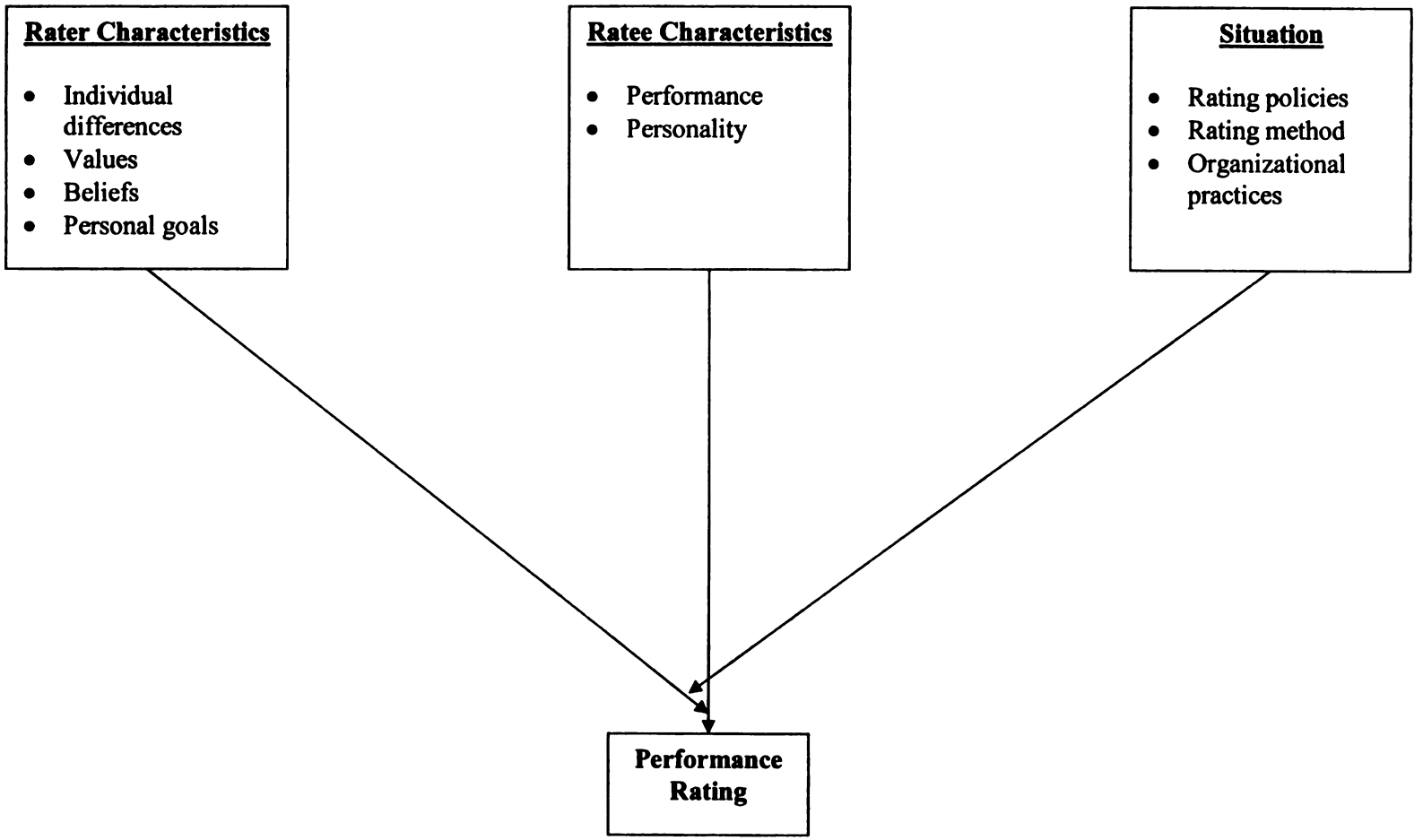


Figure 2 displays the operational model guiding the present research more specifically. The current study examines the interactive effects of a variety of individual difference variables—specifically, agreeableness, conscientiousness, ego concern, self-monitoring, and empathy—and situational characteristics—specifically, the presence/absence of a formal grievance policy, and accountability—on performance ratings of poor performers. Based on this operational model, a series of research hypotheses are advanced and tested.

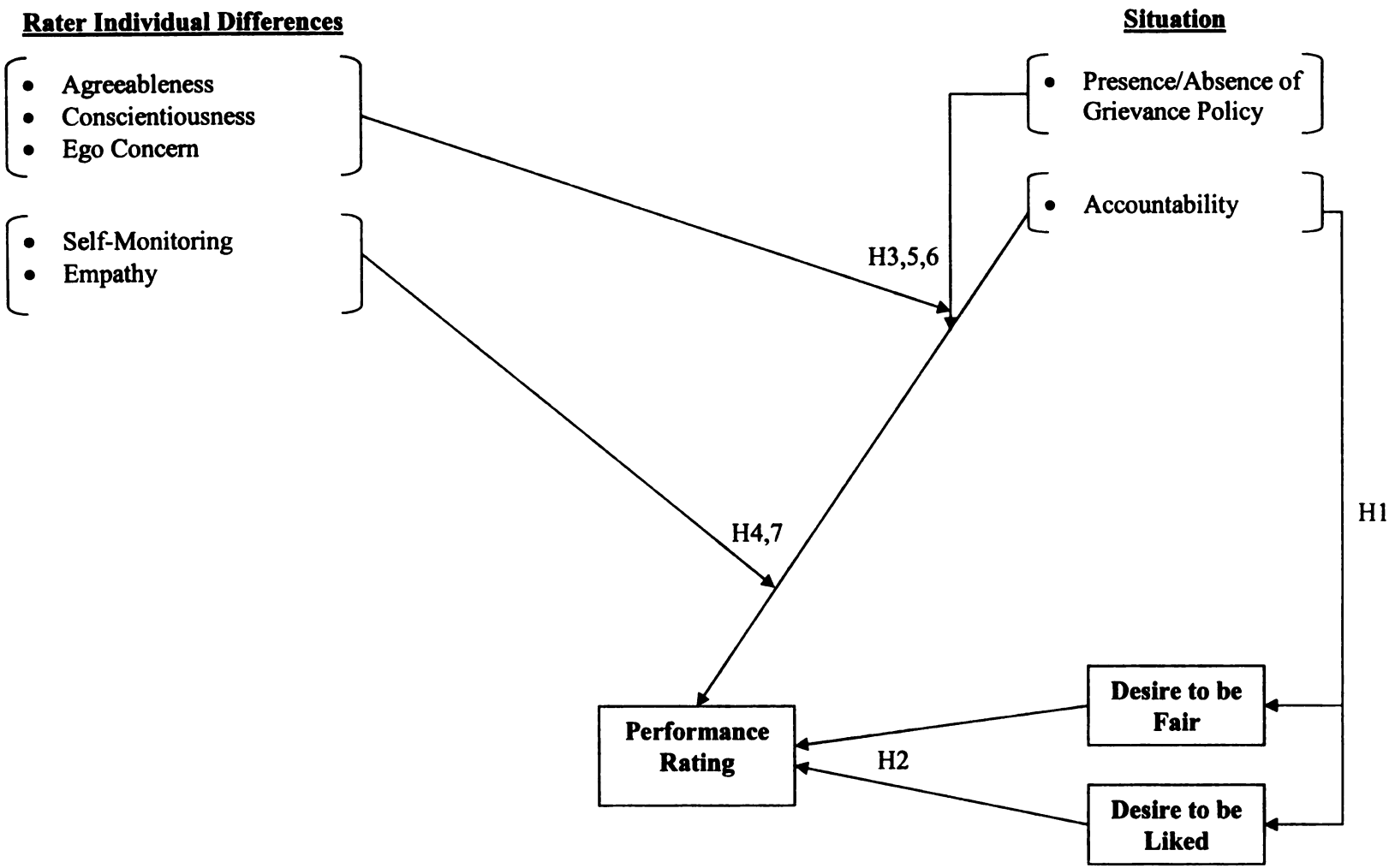
The remainder of this introduction is organized into five sections: 1) a review of leniency research to date, highlighting the contributions from the individual and situational perspectives; 2) a discussion of performance ratings as goal-directed behaviors; 3) a review of the multiple goals literature, and a discussion of how this literature can provide insight into individuals having to deal with competing objectives in the performance evaluation context; 4) a discussion of the interplay between individual difference characteristics and situational characteristics as they relate to performance ratings; and 5) a list of testable hypotheses and an operational model.

Rating Leniency

Leniency In Industry

Performance ratings are used in making decisions regarding salary, promotion, termination, training needs, and other important organizational resource allocation venues (Cleveland et al., 1989). When these decisions are based on inaccurate performance ratings, problems are likely to arise. Systematically lenient—or inflated—performance ratings are problematic because they deny organizational leaders accurate information on which to base important decisions (Kane et al., 1995). The current section substantiates

Figure 2. Operational Model



this claim by examining 1) the pervasiveness of rating leniency in organizations, and 2) the problems associated with lenient performance ratings.

Pervasiveness of rating leniency. Nearly 40 years ago, Barrett (1966) noted the presence of leniency in performance appraisal systems: "When a program is initiated, more than half of the people are given ratings above average and the proportion of high-rated people grows until only the obvious misfits fail to make the top grades" (p. 23, cited in Bernardin et al., 2000). A meta-analysis conducted by Jawahar and Williams (1997) 30 years later demonstrated that the norm in US industry was (still) to rate employees at the top end of the scale. Finally, organizational leaders are not blind to the presence of leniency in their organizations. Multiple business surveys (e.g., Bernardin & Villanova, 1986; Bretz, Milkovich, & Read, 1992; Longenecker, Jaccoud, Sims, & Gioia, 1992) indicate that organizational leaders are indeed cognizant of the pervasiveness of leniency in their performance appraisal systems.

Problems with lenient ratings. The problems associated with lenient performance ratings can best be summarized into two categories. First, lenient ratings may make it difficult for an organization to justify termination decisions (Bernardin et al., 2000). Formal performance evaluations are often the only record of employees' performance (Murphy & Cleveland, 1995). If an individual is performing "well" according to performance evaluation records despite performing quite poorly in actuality, it will be extremely difficult for an organization to justify terminating this individual based on his/her performance. Not only is this a problem in terms of defending the termination to the specific employee, but it is also a potential problem in terms of justifying the

termination in court and to other employees who may learn of the situation (Bernardin et al., 2000).

Second, lenient performance ratings may lead to an inappropriate allocation of funds used to recognize good performance. Consider, for example, that two individuals—a high performer and a low performer—receive their year-end performance evaluations, and these evaluations translate directly into performance bonuses depending on the level of ratings. Suppose that the high performer receives a high rating, and the low performer receives a moderately high rating (i.e., the rating is lenient). Although the high performer may still receive a higher bonus compared to the low performer, the difference between the bonuses granted to the two individuals may not be commensurate with the difference in effort and performance between the two individuals. From an equity perspective (Adams, 1965), the input : outcome ratio for the good performer will be high in comparison to the input : outcome ratio for the poor performer. That is, the high performer will get less for his/her effort in comparison to the low performer. In such a situation, underpayment will likely exist in the eyes of the good performer. As Adams (1965) describes, this underpayment may subsequently lead the good performer to reduce his/her future inputs (i.e., effort and performance) so as to balance the equation.

Despite these and other problems inherent in performance appraisals, some (e.g., Murphy & Cleveland, 1995) suggest that managers should simply expect inaccuracy in performance ratings rather than attempt to combat it. As noted, however, such ratings are often used to make a variety of very important and business-essential decisions in organizations; following Murphy and Cleveland's (1995) advice means that

organizational leaders should simply expect to make poor decisions. This approach to business sounds neither appealing nor economically sound.

Other individuals (e.g., Coens & Jenkins, 2000), recognizing the pitfalls and inaccuracy in performance ratings, have gone so far as to suggest that managers should ban the use of performance ratings altogether and implement some non-rating alternative. The primary alternative to rating-based performance appraisal systems is a ranking-based system—i.e., raters are asked to rank order ratees based on their performance rather than being asked to assign a specific numerical rating for each ratee. Ranking-based systems have been adopted by a number of corporations to date, including IBM, Pratt-Whitney, and Grumman (Kane et al., 1995).

Like rating-based performance appraisal systems, ranking-based systems are not without their disadvantages; Kane et al. (1995) discuss two major criticisms of ranking-based systems. First, performance *rankings* provide less useful information for feedback and employee development purposes. That is, rankings can tell an individual employee how well (or poorly) he/she is performing compared to other people, but such information conveys nothing to the employee regarding how his/her performance compares to company standards and expectations.

A second criticism of ranking-based performance appraisal systems is that although such systems facilitate designation decisions among an intact group of ratees, they confound judgments about specific characteristics of ratees (Kane et al., 1995). That is, ranking-based systems satisfy the needs of most tasks involving designating individuals within a group (e.g., making promotion or retention decisions). However, ranking-based methods imply equal intervals between ranks, thus concealing information

to users regarding how much better (or worse) than others each ratee performed within and across groups. Thus, abandoning the use of rating-based systems—and ultimately disregarding (or limiting information on) employee performance—does not appear to be a viable alternative.

Organizational leaders validate these problems and concerns with leniency discussed in this section. In a survey of Fortune 500 companies conducted by Bretz et al. (1992), 77% of sampled companies indicated that lenient appraisals *jeopardize the validity* of their appraisal systems. That is, companies recognize that leniency is a bad thing and that it taints the information gathered through appraisal systems.

Thus, evidence that organizational leaders recognize the presence of leniency in their performance appraisal systems (Bernardin & Villanova, 1986; Bretz et al., 1992; Longenecker et al., 1992), paired with the fact that they also recognize the problems with leniency, underscores the value and utility of researching the antecedents of rating leniency. Lenient performance evaluations are *not* the ideal scenario, and organizational leaders *are* concerned about imperfect performance evaluation information. The following two sections summarize what research over the past 40 years contributes to our understanding of leniency in performance evaluations, and they highlight the holes that the current research attempts to fill.

Influence of Rater Individual Differences

Guilford (1954) originally asserted that leniency is a stable characteristic of raters; as such, he noted that it could likely be predicted from measures of individual differences. That is, Guilford believed that rating leniency was a systematic characteristic of raters, rather than random noise in the rating process—some raters

simply inflate the ratings that they provide more than others. In almost 50 years since Guilford's original assertion, however, only a handful of studies have directly tested his notions.

Research on rating leniency from the individual differences perspective has examined individuals' (raters') levels of self-monitoring (e.g., Jawahar, 2001; Jawahar & Stone, 1997), agreeableness (e.g., Bernardin et al., 2000; Longenecker et al., 1987), conscientiousness (e.g., Bernardin et al., 2000; Jawahar, 2001), positive human nature (e.g., Villanova, Bernardin, Dahmus, & Sims, 1993), and self-esteem (e.g., Wexley & Yountz, 1985). The current research focuses on the previously examined individual differences of self-monitoring, agreeableness, and conscientiousness, as well as the unexplored individual differences of empathy and ego concern. These five individual difference characteristics are examined in the current research because of their proposed dependence on characteristics of the situation, as discussed in a later section.

Accurately summarizing the literature of rating leniency from the individual differences perspective, Bernardin et al. (2000) note, "too few studies examining the prediction of elevated ratings choose individual difference measures on the basis of something other than convenience" (p. 232). The current research is a step beyond convenience. Similarly, Jawahar (2001) notes, "future research should uncover other personality factors or stable individual differences with the potential to influence appraisal behaviors" (p. 882). Again, the current research, by examining more than just the "big" individual difference predictors of rating leniency (e.g., self-monitoring), is a step in this direction. Not only does the identification of stable individual difference characteristics of lenient raters have practical significance (Jawahar, 2001), it also has

implications for personnel selection in terms of selecting the individuals who will eventually provide performance ratings for an organization (Bernardin et al., 2000). Each of the aforementioned individual difference characteristics of interest in the current study is now discussed below.

Self-monitoring. Self-monitoring refers to the extent to which an individual is sensitive to situational cues of appropriateness and regulates his/her behavior accordingly (Snyder, 1974, 1979). Research generally demonstrates that high self-monitors exhibit greater social conformity than low self-monitors by tailoring their behavior to fit social and interpersonal considerations of situational appropriateness (Fandt & Ferris, 1990; Snyder, 1974, 1979). That is, high self-monitors choose behaviors most likely to maximize approval and minimize disapproval of others (Jawahar, 2001; Jawahar & Williams, 1997; White & Gerstein, 1987).

More generally, high self-monitors act in ways that optimize situational benefits (Snyder & Gangestad, 1982; Snyder, Gangestad, & Simpson, 1983; Snyder & Kendzierski, 1982b), whereas low self-monitors are less concerned with situational appropriateness; instead, low self-monitors are more likely to rely on relevant personal values or attitudes to guide their behavior (Ajzen, Timko, & White, 1982; Gerstein, Ginter, & Graziano, 1985; Snyder & Kendzierski, 1982a; Tunnell, 1980; Zanna, Olson, & Fazio, 1980). A study by Snyder and Monson (1975), for example, demonstrates that high self-monitors report significantly more variability in their behaviors across situations compared to low self-monitors who generally reported more cross-situational stability in their behaviors.

Related to performance ratings more specifically, research supports the notion that high self-monitors modify the ratings that they provide depending on the situation. For example, Jawahar (2001) observed a negative relationship (r = -.72) between individuals' levels of self-monitoring and rating accuracy. Thus, the more individuals self-monitor, the more likely they are to rate in accordance with the situation rather than honestly.

Intuitively, a key variable in the relationship between self-monitoring and performance ratings is the situation. When the situation demands accuracy, then high self-monitors should provide accurate ratings. When the situation demands (or at least promotes) leniency, then high self-monitors should provide lenient ratings. In contrast, low self-monitors should be relatively insensitive to such situational cues; as such, they should provide relatively stable ratings regardless of the situation. This notion is explored further in a later section.

Agreeableness. Agreeableness refers to the extent to which an individual is cooperative, tolerant, and flexible (Barrick & Mount, 1991; McCrae & Costa, 1985). That is, agreeableness is the extent to which individuals go along with others or the situation. Also included in the notion of agreeableness is the notion of conflict avoidance; individuals with high levels of agreeableness tend to dislike and avoid conflict, whereas individuals with low levels of agreeableness tend not to have such an aversion to conflict (McCrae & Costa, 1985).

Empirical and anecdotal evidence generally reveals that individuals with high levels of agreeableness tend to provide more elevated performance ratings than individuals with low levels of agreeableness. For example, Bernardin et al. (2000)

hypothesized and observed that individuals' levels of agreeableness relate positively to the performance ratings that they provide for poor performers. These researchers reason that individuals with high levels of agreeableness are more likely to provide elevated ratings so as to avoid potential conflict with ratees. In contrast, individuals with low levels of agreeableness should provide accurate ratings of poor performers because they are less averse to conflict.

Providing additional support for the notion of conflict aversion leading to rating leniency, Bernardin and Villanova (1986) surveyed the rating habits of 44 supervisors from 27 private and 8 public organizations, and 70 personnel administrators from 31 private and 24 public organizations. These researchers found that supervisors rated the item "Raters rate higher than deserved because they prefer to avoid confrontation" 3.9 on a 5-point scale (1 = not at all, 5 = to a great extent), and administrators rated this item a 3.8. This suggests that some individuals (but not necessarily all individuals) do indeed inflate ratings so as to avoid confrontation. Longenecker et al. (1987) also provide anecdotal accounts of managers providing elevated ratings to avoid confrontation with poorly performing subordinates.

However, with regard to agreeableness, not all situations have the same potential for conflict. To the extent that the situation does not pose the threat of potential conflict, the influence of agreeableness on leniency should decrease. Such situations are discussed in a later section.

Conscientiousness. Conscientiousness refers to the extent to which an individual is thorough, careful, and detail-oriented; individuals with high levels of conscientiousness strive for excellence and set difficult goals for themselves (Barrick & Mount, 1991;

McCrae & Costa, 1985). Related to providing performance ratings, one might expect raters with high levels of conscientiousness to rate others more accurately than raters with lower levels of conscientiousness. The reason for such a relationship could be attributed to the notion that highly conscientious individuals tend to be detail-oriented and follow instructions very carefully—it is highly unlikely that any official performance evaluation instructions read, "Please provide inaccurate ratings."

This notion of rater conscientiousness relating negatively with rating leniency is supported both conceptually (Jawahar, 2001) and empirically (Bernardin et al., 2000). Using a sample of 111 students making performance ratings of their peers after completing a group exercise, Bernardin et al. (2000) found a negative relationship (r = -.37) between raters' levels of conscientiousness and the ratings that they provided. This suggests that highly conscientious raters are more stringent in their ratings and are less likely to rate leniently.

Empathy. Empathy refers to the extent to which individuals are concerned for and understand how others feel (Davis, 1983). Theory and research on empathy generally suggests that empathy relates positively with prosocial (i.e., helping) behaviors toward others (Bentler, 1972; Mehrabian & Epstein, 1972). For example, a meta-analysis of 41 samples by Eisenberg and Miller (1987) found that individuals' self-reported levels of empathy were positively associated (r = .17) with a wide range of prosocial behaviors benefiting other individuals. Similarly, in a sample of 100 secretaries, McNeely and Meglino (1994) found a positive relationship (r = .18) between empathy (measured using the empathetic concern subscale of Davis's (1983) Interpersonal Reactivity Index and

prosocial behavior directed toward individuals (e.g., sending birthday cards to office members, doing a personal favor for someone, etc.).

To date, I am not aware of any research relating individuals' levels of empathy to the performance ratings that they provide for others. Based on the findings noted above, however, one might expect that, all else equal, a rater with a high level of empathy would be less likely to provide accurate ratings of poor performers and more likely to instead provide lenient ratings so as to protect the poor performers' feelings. A rater with a low level of empathy, in contrast, may be more likely to provide accurate ratings of poor performers because he/she is generally insensitive to the feelings of others.

Not all situations necessarily yield the same need to protect the feelings of others; that is, others' feelings are at risk to a greater or lesser extent depending on the situation. Thus, regardless of a rater's level of empathy, there are likely some situations in which this empathy will transpire into lenient ratings more than others. This notion is explored in more detail in a later section.

Ego concern. Another individual difference characteristic that is likely to influence the ratings that raters provide is the extent to which they are concerned with protecting their ego, or what I will refer to as ego concern. Research generally demonstrates that individuals, when their egos are threatened, engage in behaviors that will restore their egos (e.g., Brown & Gallagher, 1992; Crocker, Thompson, McGraw, & Ingerman, 1987; Fein & Spencer, 1997; Gibbons & Gerrard, 1991). Like other individual difference characteristics, it is entirely reasonable to expect individuals to differ in the extent to which they are concerned with protecting their egos.

A logical extension of the noted research findings is that individuals who are concerned with protecting their egos would not wait until their egos have been threatened before engaging in such ego restoration behaviors. Rather, it seems logical to infer that these individuals would consistently behave in ways that will maintain their egos. That is, individuals who are concerned with protecting their egos are likely to take proactive steps to ensure that their egos are not threatened.

Such proactive steps to ensure a protected ego are likely to impact the ratings that individuals provide for others' performance. Specifically, when situational characteristics demand (or promote) accuracy, individuals who are more concerned with protecting their egos (compared to individuals who are less concerned with protecting their egos) will likely rate accurately so as to maintain the way that others think of them. When situational characteristics demand (or tolerate) leniency, in contrast, these individuals will likely rate leniently for the very same reasons. This idea of situational dependency is explored in more detail in a later section.

Summary of rater individual differences. The current section has implicated a number of individual difference characteristics of raters that are likely to influence the performance ratings that they provide for others. Specifically, raters' levels of self-monitoring, agreeableness, conscientiousness, empathy, and ego concern have all been explored as they relate to rating leniency. The next section explores this same problem of rating leniency from an entirely different perspective—situational characteristics that hinder or support leniency.

Influence of Situational Characteristics

The primary shortfall of the research summarized in the previous section is that it generally ignores the influence of situational characteristics on rating leniency. As the previous section demonstrates, rater individual differences are important to take into consideration in any study of rating leniency, but failing to consider the situations in which raters assign their ratings severely limits obtained results. Indeed, previous researchers have warned against being overly reliant on explaining leniency with individual difference variables and cognitive processes; they note that we must also pay attention to the contextual influences within which rating decisions are made (Dipboye, 1985; Ilgen & Favero, 1985; Judge & Ferris, 1993; Nathan, Mohrman, & Milliman, 1991; Wexley & Klimoski, 1984).

A great deal of the research relating situations to rating leniency has focused on the situational characteristics of appraisal purpose and accountability. Results pertaining to each of these avenues of research are summarized in subsequent sections, but more attention is directed toward accountability because it is a key situational variable of interest in the current research. In addition, whether a formal grievance policy is built into the appraisal system is explored as a predictor of rating leniency. To date, this is an unexplored variable in rating leniency research.

Appraisal purpose. More than 50 years ago, Taylor and Wherry (1951) hypothesized that performance ratings made for administrative purposes such as pay raises and promotions are more lenient than ratings obtained for research, feedback, and employee development purposes. Most of the past research testing Taylor and Wherry's (1951) hypothesis has yielded results consistent with the hypothesis (e.g., Harris, Smith,

& Champagne, 1995; Waldman & Thornton, 1988). However, a few studies (e.g., McIntyre, Smith, & Hassett, 1984; Murphy, Kellam, Balzer, & Armstrong, 1984) have observed results inconsistent with this performance appraisal purpose hypothesis; primarily, these studies have found no difference at all between ratings provided for administrative purposes versus those provided for other purposes.

Jawahar and Williams (1997) conducted a meta-analysis of 22 samples to test Taylor and Wherry's (1951) hypothesis and to resolve the inconsistent results just described. The results of this meta-analysis demonstrate that, consistent with the hypothesis, performance ratings are 1/3 of a standard deviation higher when intended for administrative purposes compared to when they are intended for employee development purposes. Results also demonstrated that the relationship between appraisal purpose and performance ratings is stronger when research studies were conducted in field settings, by practicing managers, for real subordinates (vs. paper people). These authors interpret their results as evidence that appraisal purpose does in fact relate to rating leniency.

Accountability. Accountability has been broadly defined in the performance appraisal literature as susceptibility to the expectations or wishes of others (Klimoski & Inks, 1990). Typically, accountability is operationalized in performance appraisal research as the need to justify or rationalize a performance rating to another individual, primarily the ratee. Indeed, Beach and Mitchell (1978) view accountability as strongly influenced by an individual's belief that he/she will have to share a decision's results with others who have a vested interest in the decision (e.g., being required to share a performance rating with a ratee).

Outside of the performance appraisal context, accountability research more generally suggests that the best way to cope with accountability is to exhibit behavior that an individual believes will be acceptable to others (Fandt & Ferris, 1990; Tetlock, 1983). This logic extends to the performance appraisal context as well.

Situations in which a rater anticipates having to share his/her rating of an individual *to* the individual in a face-to-face discussion promotes greater accountability to the individual than situations in which a rater does not anticipate such a face-to-face discussion (Klimoski & Inks, 1990). This increased accountability then creates potential for greater rating distortion—particularly rating leniency. As Klimoski and Inks (1990) note, "the direction of the distortion itself should be toward what the rater perceives the ratee's wishes or expectations to be...in most cases, [this] will be in an upward (more positive) direction" (p. 197). This assertion assumes, however, that the rater feels accountable to the *ratee* (rather than someone else). This notion of being accountable to someone *other than* the ratee is explored later in this section.

Research generally demonstrates that raters provide higher, more lenient evaluations of poor performers when they are required to justify these evaluations to ratees compared to situations in which they are not required to provide such justifications (Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975). In a laboratory study conducted by Fisher (1979), for example, subjects assumed the role of managers who had to rate the performance of a confederate. Prior to providing these performance ratings, subjects were told either that 1) they would never interact with the ratee again, and the ratee would never see the rating, or 2) they would have to convey the rating to the ratee (confederate) in a feedback session. As expected, subjects led to

believe that they would have to convey the rating to the ratee provided significantly higher ratings for a poorly performing confederate compared to subjects led to believe that they would not have to interact with the ratee. There were no differences in ratings provided for high performing confederates.

Ilgen and Knowlton (1980) observed similar results when they had subjects in a laboratory study act as supervisors and rate the performance of either a high or low performer. Subjects in this study were asked to provide two ratings: One rating that would never be seen by the ratee, and a second rating that the rater (subject) would have to convey to the ratee (confederate) in a feedback session. As expected, the ratings provided for feedback purposes were significantly higher than the ratings provided for non-feedback purposes for poor performers. There were no significant differences in ratings provided for high performers.

A third laboratory study conducted by Klimoski and Inks (1990) found essentially the same results as the previous two studies. Like the previous studies, these researchers had subjects assume the role of a supervisor tasked with rating the performance of a "subordinate" (confederate). In this study, however, subjects never met their subordinate; rather, subjects were told that the subordinate was working in another room, and his/her work (i.e., the work to be rated by subjects) was displayed to subjects on a computer. Subjects viewed the work of either a poorly performing subordinate or a high performing subordinate. Subjects were led to believe that they would either 1) have to convey their rating of the subordinate *to* the subordinate, or 2) not have to convey the rating to the subordinate—the rating was completely anonymous. As expected, subjects led to believe that they would have to convey their ratings to their subordinates provided significantly

higher ratings for poorly performing subordinates compared to subjects led to believe that their ratings were anonymous. There were no significant differences in ratings provided for high performers.

All of the studies just summarized, as well as the remaining studies that examine accountability in performance evaluations not discussed above, confound face-to-face discussions/justifications with anonymity of ratings. That is, based on the extant literature we do not know whether the leniency effects of accountability are due to being required to engage in face-to-face discussions with ratees or if these effects are simply due to the fact that ratees will know raters rated them. The current research addresses this confounding of variables by holding lack of anonymity constant; that is, no ratings were completely anonymous in the current study.

The reader will notice that the research summarized in this section looks only at situations in which the rater is/is not accountable to the *ratee*. It is quite possible in an organization, however, for a rater to be accountable to someone *other than* or *in addition to* the ratee. For example, a rater may be accountable to his/her superior regarding the ratings that he/she provides for his/her subordinates, or the situation may place raters accountable to ratees *and* upper management (or peers or some other group of people). Although existing research has not examined this possibility, it suggests that the rater would simply adjust his/her ratings toward whomever he/she is accountable; thus, if a rater is accountable to his/her superior, one would likely expect the ratings that he/she provides to be more accurate than lenient *regardless* of ratee performance. In the event that a rater is accountable to multiple individuals, he/she is likely to have to make some

sacrifices and compromises. How raters are likely to make these compromises is addressed in a later section.

Grievance policy. Some performance evaluation systems provide the opportunity for individuals (e.g., ratees, peers, senior management) to challenge the performance ratings that raters provide. These formal grievance policies are often included in performance evaluation systems as a way to increase perceived fairness of these systems (Folger et al., 1992; Greenberg, 1986). To date, I am aware of no research that examines the impact of having formal grievance policies on ratings provided. However, formal grievance policies are likely to have an impact on the ratings that individuals provide due to the potential for conflict and confrontation inherent to grievance policies. Therefore, knowing whether individuals (e.g., ratees, one's peers, or one's superiors) will have the opportunity to challenge the ratings that one provides is likely to impact these ratings, particularly for individuals who are averse to conflict and confrontation. These ideas are explored in more detail in a later section.

Summary of situational characteristics. The current section has examined the influence of situational characteristics on the ratings that individuals provide for others' performance—specifically, the leniency of these ratings. In particular, previous research suggests that appraisal purpose and rater accountability both relate to rating leniency. In addition, the presence or absence of a formal grievance policy built into the rating system was explored as having a potential influence on rating leniency. The current and previous sections allude to the notion of raters seeking to achieve or make progress toward various goals and objectives when they provide performance ratings for others. The next section examines this notion in more detail.

Performance Ratings As Goal-Directed Behaviors

Recent theory and research has begun to view the act of providing performance ratings as a goal-directed behavior (Murphy & Cleveland, 1995), and researchers use this reasoning to explain the presence of inaccuracy—and most commonly leniency—in performance ratings. That is, individuals may possess a variety of objectives when they provide performance ratings, and behaving in a manner consistent with some of these different objectives may lead to lenient ratings. Indeed, Murphy and Cleveland (1995) note that it is useful to conceptualize the appraisal process as "a goal-directed communication process in which the rater attempts to use the performance appraisal to advance his/her interests" (p. 215).

To date, little empirical research exists on this subject. Rather, Bjerke et al. (1987, as cited in Murphy & Cleveland, 1995) and Longenecker et al. (1987) provide only anecdotal accounts of the goals and objectives pursued by raters. Despite the lack of empirical research, however, these anecdotal accounts describe what happens when raters are motivated to achieve particular objectives. For example, Longenecker et al. (1987) interviewed 60 upper-level executives and identified a variety of reasons why raters may deliberately provide inflated or deflated performance ratings for others. The most common reason supervisors cited for providing lenient ratings for their subordinates was to maximize subordinates' merit increases. Other reasons identified by Longenecker et al. (1987) for providing lenient performance ratings of others include protecting a subordinate whose performance was suffering because of personal problems, promoting a subordinate "up and out" of the work unit, and avoiding a confrontation with a poorly performing subordinate. Similarly, Klimoski and Inks (1990) found that raters tend to

provide lenient ratings of others so as to avoid the negative reactions that often accompany low ratings. Longenecker et al. (1987) also identified a number of reasons for providing *de*flated ratings of subordinates, such as to teach a rebellious subordinate a lesson about who is in charge, and to send a message to a subordinate that he/she should consider leaving the organization. Table 1 summarizes the reasons for providing lenient and deflated ratings identified by Longenecker et al. (1987). This anecdotal evidence (e.g., Longenecker et al., 1987) lends support to the notion of conceptualizing the act of providing performance ratings for others as a goal-directed behavior.

Intuitively, these different goals and objectives that raters may seek to achieve when assigning performance ratings can either originate within themselves (e.g., a rater wants a particular individual to like him/her) or they can be the product of environmental influences (e.g., a rater's own pay is contingent on providing accurate ratings). The current research focuses on the later of these origins (i.e., the environment), but this should *not* imply that an individual's own personal goals irrespective of the environment are unimportant.

Consider the environmentally influenced objectives inherent to justifying ones ratings to others. Research on accountability summarized earlier consistently demonstrates that when raters are required to justify their ratings *to* ratees, they provide more lenient ratings compared to when there is no such requirement (Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975). As noted earlier, existing research looks only at situations in which the rater is/is not accountable to the *ratee*. However, it is quite possible in an organization for a rater to be accountable to someone *other than* or *in addition to* the ratee. For example, a rater may be accountable to his/her

Table 1

Sources of Intentional Rating Bias Identified By Longenecker et al. (1987).

Inflating the Appraisal	Deflating the Appraisal
• To maximize the merit increases a subordinate would be eligible to receive	• To shock a subordinate back onto a higher performance track
• To avoid hanging dirty laundry out in public	• To teach a rebellious subordinate a lesson about who is in charge
• To protect or encourage a subordinate whose performance was suffering because of personal problems	• To send a message to a subordinate that he/she should consider leaving the organization
• To avoid creating a written record of poor performance that would become a permanent part of a subordinate's personnel file	• To build up a strongly documented record of poor performance that could speed up the termination process
• To avoid a confrontation with a subordinate with whom the manager had recently had difficulties	
• To give a break to a subordinate who had improved during the latter part of the performance period	
• To promote a subordinate "up and out" when the subordinate was performing poorly or did not fit the department	

superior regarding the ratings that he/she provides for his/her subordinates, or the situation may make raters accountable to ratees *and* upper management (or peers or some other group of people).

Although existing research has not examined this possibility of being accountable to someone other than or in addition to the ratee, based on the existing research it seems reasonable to infer that a rater would adjust his/her ratings of others depending on *to whom* he/she is accountable. When he/she is accountable to the ratee, there is a need to satisfy the ratee; as such, the rater rates leniently as demonstrated in the literature. When he/she is accountable to others (e.g., his/her peers), there is less of a need to satisfy the ratee and more of a need to satisfy these other individuals. I would argue that being accountable to these other individuals would lead a rater to be more accurate than lenient—it is in the best interest of an organization to collect accurate rather than inaccurate performance ratings (Bretz et al., 1992).

In the event that a rater is accountable to both the ratee *and* other individuals simultaneously, the rater will likely have competing objectives, especially when rating poorly performing individuals. That is, rating in a manner consistent with pleasing the ratee would lead to a lenient rating, whereas rating in a manner consistent with pleasing others (e.g., one's peers, superiors) would lead to a more accurate rating. Since the rater cannot satisfy both constituents completely, he/she is likely to have to make some sacrifices and compromises. This idea of satisfying competing objectives when providing performance ratings is explored further in the next section.

Behavior In the Pursuit of Multiple Goals

The previous section highlighted the notion of an individual being motivated by multiple objectives when assigning performance ratings for others. Indeed, the idea of being motivated by multiple objectives—or multiple goals—exists beyond the realm of performance appraisals; in fact, it exists everywhere. A student in the classroom may have the goals of a) making a good impression on the instructor and b) making new friends. An individual working in a team context may have the goals of a) being a team player by helping others when they are in need and b) ensuring that his/her individual contributions to the team's performance are at a respectable level in case the team is reprimanded for poor performance. Thus, individuals are constantly faced with multiple goals at any given time.

Multiple Goal Relationships

The relationships between multiple goals can take on a variety of different forms, and precise nature of these relationships has implications for how individuals deal with such goals. As outlined in a taxonomy offered by Schmidt (2000), the relationship between multiple goals can take on one of three forms: hierarchically structured goals, complementary goals, and competing or conflicting goals. The current research focuses on the latter relationship (i.e., competing or conflicting goals), but each of the goal relationships is explicated briefly for a more thorough review. Hierarchically structured goals are arranged such that one goal is subordinate to another goal; the attainment of the former, lower-order goal moves an individual toward the attainment of the latter, higher-order goal. An example of two hierarchically structured goals is the goal of completing one's dissertation and the goal of earning one's doctorate. Attainment of the former goal

(i.e., completing one's dissertation) leads an individual toward the attainment of the latter goal (i.e., earning one's doctorate), and attainment of the latter goal *is not possible* without first attaining the former goal. Such a relationship is central to cybernetic control theories of self-regulation (Carver, Lawrence, & Scheier, 1996; Carver & Scheier, 1998; Klein, 1989).

Second, complementary goals are arranged such that the attainment of one goal leads to the attainment of another goal, but the former goal is not subordinate to the latter goal. That is, attainment of the latter goal does not *require* attainment of the former goal, but progress toward this latter goal is furthered by the former goal. An example of two complementary goals is the goal of completing the required courses for one's degree and the goal of learning more about theories central to industrial and organizational psychology. Attainment of the former goal (i.e., completing one's required courses) leads an individual toward the attainment of the latter goal (i.e., learning more about I/O theories). However, unlike hierarchically structured goals, it is possible in the complementary goals scenario to attain the latter goal without first attaining the former goal. For example, an individual could learn about I/O theories by reading relevant books and journal articles rather than by completing one's required courses.

Finally, multiple goals can conflict or compete with one another. One way in which goals may conflict with one another is in terms of resource allocation (Kanfer, 1990), such that an individual may have available enough resources to pursue only one of the goals at a given time. For example, consider an individual who is committed to the goals of getting an A in psychology for the semester and making new friends. The night before the student's psychology final exam, the individual's friends call and ask him/her

to go out. Going out would likely further the individual toward his/her goal of making new friends, but it would likely interfere with his/her goal of getting an A in psychology. Conversely, staying home and studying would likely further the individual toward his/her goal of getting an A in psychology, but it would interfere with his/her goal of making new friends. Since the individual can only be in one place at one time, he/she must allocate his/her resources. He/she may decide to devote all available resources (i.e., time) to one goal or the other, or he/she may decide to split the available resources among the two goals—e.g., study for two more hours and then go out.

A second way in which goals may conflict with one another is in terms of their end states. That is, pursuing Goal 1 would lead to behavior X, whereas pursuing Goal 2 would lead to behavior Y (where X and Y are opposite one another). Consider as an example an individual who has the goals of a) choosing a college where average freshman-level classes have about 10-15 students, and b) choosing a college with at least 15,000 students in the freshman class. Pursuing the first goal would most likely lead the individual to attend a small, private university, whereas pursuing the second goal would likely lead the individual to attend a large, public university. Obviously, the individual cannot attend both universities, so he/she must make a choice among the goals.

Compared to the former kind of goal conflict, this latter kind of goal conflict (i.e., conflicting end states) is likely to be most relevant in the performance appraisal context. Specifically, consider a ratee who performs poorly and a rater who has the goals—or objectives—of satisfying the ratee and satisfying his/her (i.e., the rater's) peers and/or superiors simultaneously. Pursuing the first of these objectives would likely lead the rater to provide a low rating for the individual, whereas pursuing the second objective

would lead the rater to provide an inflated rating. Similar to the previous example involving a choice among colleges, the rater cannot pursue both objectives in this scenario—he/she can only provide one rating. Instead, he/she must choose which of the two goals he/she will pursue in the present situation, or he/she will have to make some sort of compromise. It is this form of goal conflict (i.e., competing end states) on which the current research is focused.

Dealing With Competing Goals: A Literature Review

Surprisingly, research on how individuals deal with and reconcile two or more conflicting goals is lacking in the literature despite its widespread relevance. The idea of multiple (but not necessarily *competing*) goals in general is discussed somewhat, but these discussions offer little guidance in terms of speculating how individuals actually *deal* with these competing goals. For example, existing research often addresses multiple goals from the perspective of person-organization goal congruence—i.e., employees have one set of goals that may or may not be congruent with the organization's set of goals. With respect to person-organization goal congruence, outcome variables most often studied include organizational commitment, job satisfaction, and intentions to leave the organization. Specifically, the aforementioned goal congruence tends to relate positively with organizational commitment and job satisfaction (e.g., Reichers, 1986; Vancouver & Schmitt, 1991) and negatively with voluntary turnover intentions (e.g., Vancouver & Schmitt, 1991). This source of goal conflict is very different from the idea of multiple, conflicting goals that compete for one's attention and resources. Specifically, the focus of the current research is on goals to which the individual is committed. Therefore, the issue of person-organization goal incongruence in and of itself is not of interest in this

research. If, however, an individual is committed to both his/her individual goals as well as the organization's goals (and they are in conflict), this form of person-organization goal incongruence is precisely the goal incongruence of interest in the current research.

Another line of research that deals with competing goals is that of work-family conflict, or, more broadly, inter-role conflict. Inter-role conflict exists when the demands of one role are incompatible with meeting the demands of the other role (Thomas & Ganster, 1995). Suppose an individual is committed to the goal of being a successful businessperson as well the goal of being a good spouse and parent. When the clock strikes 5:00pm on Friday afternoon and the individual has not met his/her deadline for a report, he/she must make a choice. First, he/she can pursue the goal of being a successful businessperson by staying at the office and finishing the report. Alternatively, he/she can pursue the goal of being a good spouse and parent by leaving the office and going on the family camping trip. Given the fact that the individual can only be in one place at one time, these two courses of action conflict with one another. That is, the individuals' *goals* conflict with one another in the current situation. What should the individual do? The inter-role conflict literature approaches this problem from a resource allocation perspective, arguing that the individual should *share* his/her time between the two goals or objectives (e.g., work until 6:00pm and then leave for the camping trip).

Literature more self-regulatory in nature offers similar insight as the work-family conflict literature. Kernan and Lord (1990) note that when individuals are faced with multiple goals, they must establish some sort of priority among the goals and attempt to satisfy both goals according to this priority. They argue that this priority is a function of the goals' valences, perceived discrepancies, and expectancies. Similarly, Naylor,

Pritchard, and Ilgen (1980), in their model of resource allocation, contend that individuals distribute their time and effort in accordance with their expectations for maximizing anticipated positive affect. Their theory suggests that individuals focus more attention and resources on the goal or objective that will lead to the most positive affect. These authors also suggest that individuals might reduce goal conflict by prioritizing goals based on each goal's utility, thus allowing one to work toward more important goals.

Performance Ratings In the Pursuit of Multiple Goals

Thus, existing theory and research offers insight into how raters behave when motivated by one objective vs. another objective. But how likely is it that a rater would be motivated by *only one objective*? In the situation described throughout this proposal where a rater is motivated to satisfy both a poorly performing ratee *and* his/her (i.e., the rater's) peers and/or superiors, the rater is motivated to achieve two conflicting objectives.

Logically, the rater who wants to satisfy both a poorly performing ratee *and* his/her peers has one of two options. First, he/she can favor one objective to the exclusion of the other and rate *either* fairly (i.e., low) *or* leniently. Alternatively, he/she can make a compromise and rate somewhere in between these two endpoints. Proponents of expectancy theories would likely contend that the rater would behave in a manner consistent with the objective with the highest associated valence (Vroom, 1964). When the two goals have equal valence, however, research on dual-tasking offers some insight into how the rater will behave. Work by Pashler (2000) shows that individuals working on two or more tasks (i.e., motivated by two or more objectives) can either perform the tasks simultaneously or switch attention between the two tasks. Thus, in the context of

providing a performance rating for a poorly performing individual, this research would suggest that a rater will attempt to satisfy *both* objectives—i.e., a rater will compromise between rating fairly and rating leniently.

When assigning this (compromised) performance rating, however, where will the rating actually fall? From an organizational leader's perspective, it has already been established that the ideal scenario would be for the rater to align his/her rating more closely with satisfying his/her peers (i.e., rate fairly) (Bernardin et al., 2000; Bretz et al., 1992). Whether this ideal actually transpires, however, is likely dependent on a variety of individual and environmental characteristics.

Person By Situation Interactions In Performance Ratings

The premise of a person by situation interaction is that an individual's behavior at a specific moment in time is a product of his/her trait-like predispositions (i.e., needs, desires, etc.) and situational influences. Specifically, individuals behave similarly in situations that they perceive to be similar; in contrast, situations that are perceived to be different may elicit different personality characteristics (Flavell, 1985; Mischel & Shoda, 1995; Pervin, 1989), leading individuals to behave differently. There are differing perspectives, however, regarding the role of the situation. One perspective is that individuals actively seek out certain situations rather than others (Mischel & Shoda, 1995; Pervin, 1989). An alternative perspective views the individual's behavior as more reactionary compared to the aforementioned, proactive perspective. According to this perspective, individuals passively react to and respond to the situations that they encounter rather than actively seek out certain situations (Mischel & Shoda, 1995; Pervin, 1989). Although both of these perspectives have merit, the current study adopted the

latter. By adopting the latter perspective, this study cannot address questions related to the situations in which different individuals choose to participate. Such questions must be addressed in a study that allows individuals to choose their situations.

Thus, in order to understand an individual's behavior completely, we must look at both the person and the situation. It is unreasonable to assume that all individuals will respond to the same situation in the same manner (i.e., a situation perspective), just as it is unreasonable to assume that a given individual (or an individual with a particular level of a given personality characteristic) will respond to different situations in the same manner (i.e., a person perspective). Instead, different individuals are likely to respond to different situations differently, thus supporting the need to examine person by situation interactions. Indeed, others (e.g., Eysenck, 1997) have noted the importance of examining person by situation interactions, particularly in personality research, arguing that social behavior is influenced by interactions of the individual and the environment.

With regard to rating leniency specifically, these same arguments apply. As summarized previously, one line of research suggests that personality variables, such as self-monitoring, agreeableness, empathy, conscientiousness, and ego concern affect leniency. Similarly, a separate line of research suggests that situational variables, such as rater accountability affect leniency. However, these factors (person and situation) have only been examined in isolation from each other rather than simultaneously. The current study examined both person and situation variables using a between-subjects design as an attempt to integrate what is known about person and situation variables as they influence rating leniency.

Despite the lack of research examining person by situation interactions with regard to rating leniency (or other phenomena of interest to psychologists for that matter), some researchers have begun to note the potential for person by situation interactions as viable predictors of rating behavior. For example, Kane et al. (1995) note, "perhaps some raters are more prone to rate leniently in rating situations that require interpersonal interaction with ratees. They may elevate their ratings because they anticipate that the potential demands placed on them to resolve conflict with ratees given low ratings may exceed their social and self-regulatory competencies" (p. 1048-1049). Similarly, citing Ferris and Mitchell (1987), Fandt and Ferris (1990) note, "just as characteristics of the situation prove more or less conducive to opportunistic behavior, some people are in a better position to capitalize on such opportunities because they are more sensitive and attuned to their task, social, and information environments" (p. 141). Specific person by situation interactions affecting rating leniency addressed in the current study are discussed in the next section.

Research Hypotheses and Operational Model

Figure 2 presents an operational model for the questions of interest in the current study. The following testable hypotheses are based on this operational model.

First, it was hypothesized that accountability would affect the extent to which individuals endorsed a fairness and liking goal. Specifically, when raters were required to justify their rating of a particular ratee to the rater's peers (excluding the ratee), they would report wanting to be fair to a greater extent than wanting to be liked. When raters were required to justify their rating of a particular ratee *to* the ratee (excluding the rater's peers), they would report wanting to be liked to a greater extent than wanting to be fair.

Finally, when raters were required to justify their rating of a particular ratee to both the ratee *and* the rater's peers, they would report wanting to be fair and wanting to be liked to an equal extent.

Hypothesis 1: *Accountability would affect raters' levels of fairness and liking goal valence, such that:*

Hypothesis 1a: *Raters who are accountable to their peers (excluding the ratee) would value a fairness goal to a greater extent than a liking goal.*

Hypothesis 1b: *Raters who are accountable to the ratee (excluding the raters' peers) would value a liking goal to a greater extent than a fairness goal.*

Hypothesis 1c: *Raters who are accountable to both their peers and the ratee would value a liking goal and a fairness goal to an equal extent.*

In addition, these fairness and liking goals were expected to relate to the ratings provided by raters, such that the extent to which an individual endorsed a liking goal would relate positively with ratings provided, and the extent to which an individual endorsed a fairness goal would relate negatively with ratings provided.

Hypothesis 2: *Goal valence would relate to the performance ratings that raters provide, such that:*

Hypothesis 2a: *Valence of a liking goal would relate positively with ratings provided.*

Hypothesis 2b: *Valence of a fairness goal would relate negatively with ratings provided.*

Research examining the effects of rater accountability (i.e., typically operationalized as whether the rater has to justify or explain a rating to the ratee in question) generally finds that raters tend to rate poor performers more leniently when they are accountable to these poor performers, and this effect is attributed to raters attempting to avoid potential confrontation and negative reactions from ratees (e.g., Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975). While this explanation is certainly intuitive, its inherent logic assumes that all individuals have the same aversion to conflict. It is quite possible, however, for some individuals to be more or less conflict averse than other individuals. For example, the individual difference characteristic of agreeableness refers in part to the extent to which one is conflict averse (Barrick & Mount, 1991; McCrae & Costa, 1985).

However, simply being required to justify one's rating of an individual *to* the individual does not necessarily imply that conflict may arise. Consider, for example, the situation in which a rater is required to justify his/her rating of a poorly performing ratee *to* the ratee, but the ratee does not have the opportunity to object to this rating. Contrast this with a situation in which the ratee *will*—by virtue of policy—have the opportunity to object to the rating. There is much greater potential in the latter situation for conflict to arise. As such, it was expected that accountability, raters' levels of agreeableness, and the presence/absence of a grievance policy would interact in their effects on performance ratings in a manner consistent with Figures 3 and 4.

Figure 3. Hypothesis 3a: Accountability X Agreeableness Interaction On Performance Ratings (Grievance Policy Present)

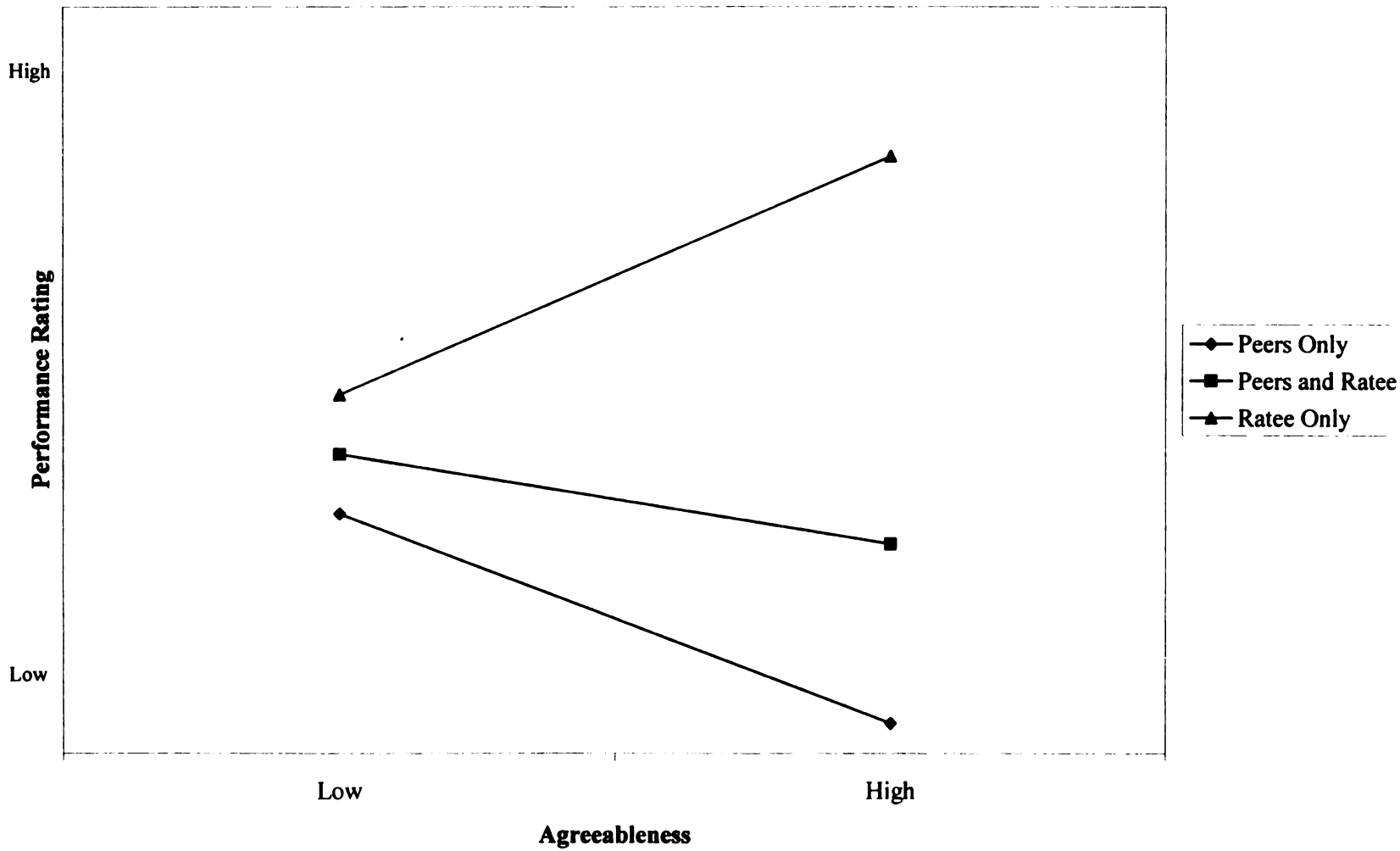
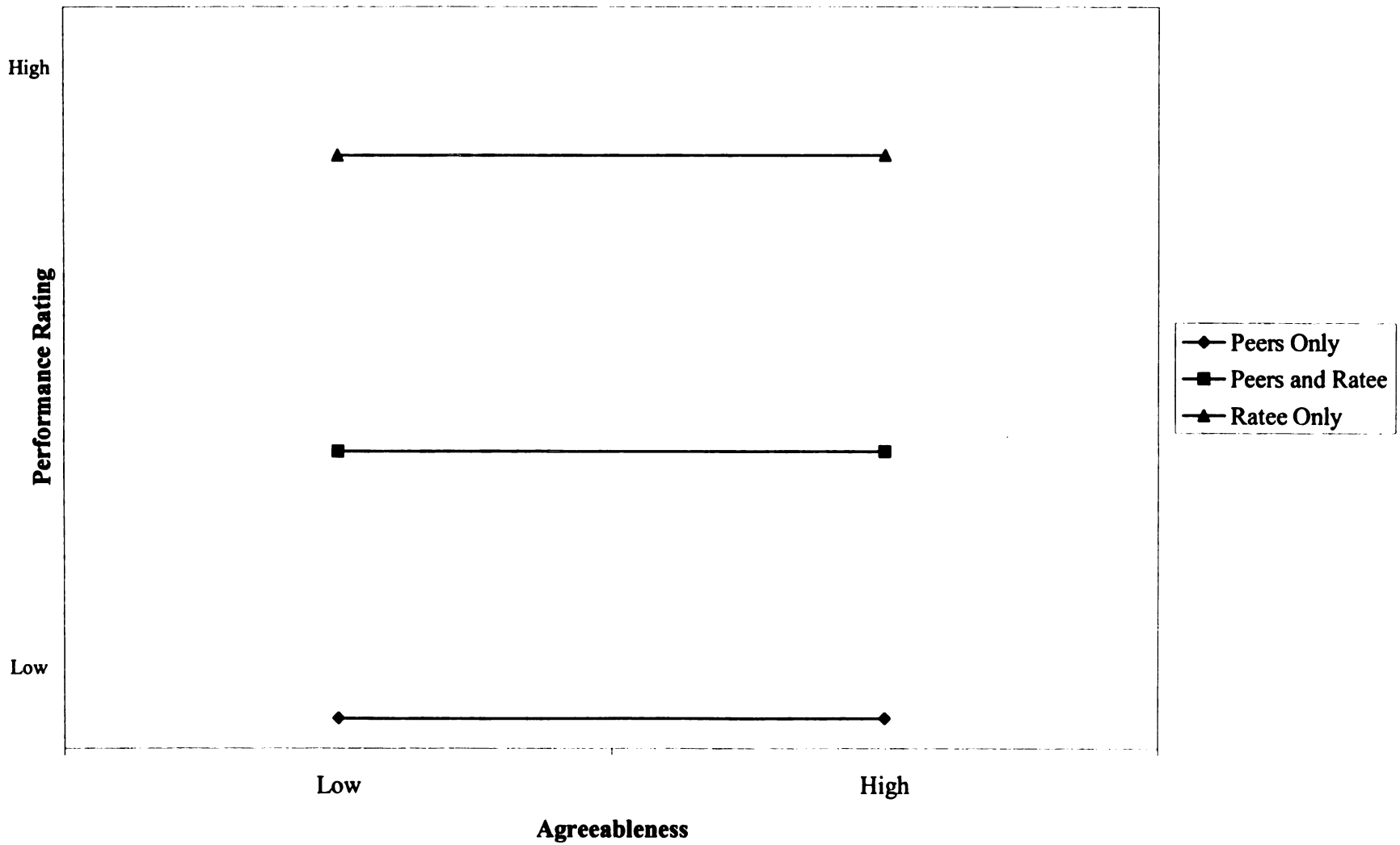


Figure 4. Hypothesis 3b: Accountability X Agreeableness Interaction On Performance Ratings (Grievance Policy Absent)



Hypothesis 3: *Accountability, raters' levels of agreeableness, and the presence/absence of a grievance policy would interact in their effects on performance ratings provided, such that:*

> ***Hypothesis 3a:*** *In the presence of a grievance policy, accountability and raters' agreeableness would interact in their effects on ratings provided, such that 1) mean ratings provided would be lowest when accountable to one's peers (excluding the ratee), highest when accountable to the ratee (excluding the rater's peers), and moderate when accountable to both the one's peers and the ratee simultaneously; and 2) the relationship between raters' agreeableness and ratings provided would be weak and positive when accountable to one's peers (excluding the ratee), zero when accountable to the ratee (excluding the rater's peers), and strong and positive when accountable to both one's peers and the ratee simultaneously.*
>
> ***Hypothesis 3b:*** *In the absence of a grievance policy, accountability would have a main effect on ratings provided, such that accountability to one's peers (excluding the ratee) would result in low ratings, accountability to the ratee (excluding the rater's peers) would result in high ratings, and accountability to both one's peers and the ratee simultaneously would result in moderate ratings.*

As noted, I am aware of no existing research relating empathy to performance ratings. However, given that individuals with higher levels of empathy tend to exhibit more helping behaviors (Berkowitz, 1972; McNeely & Meglino, 1994; Mehrabian &

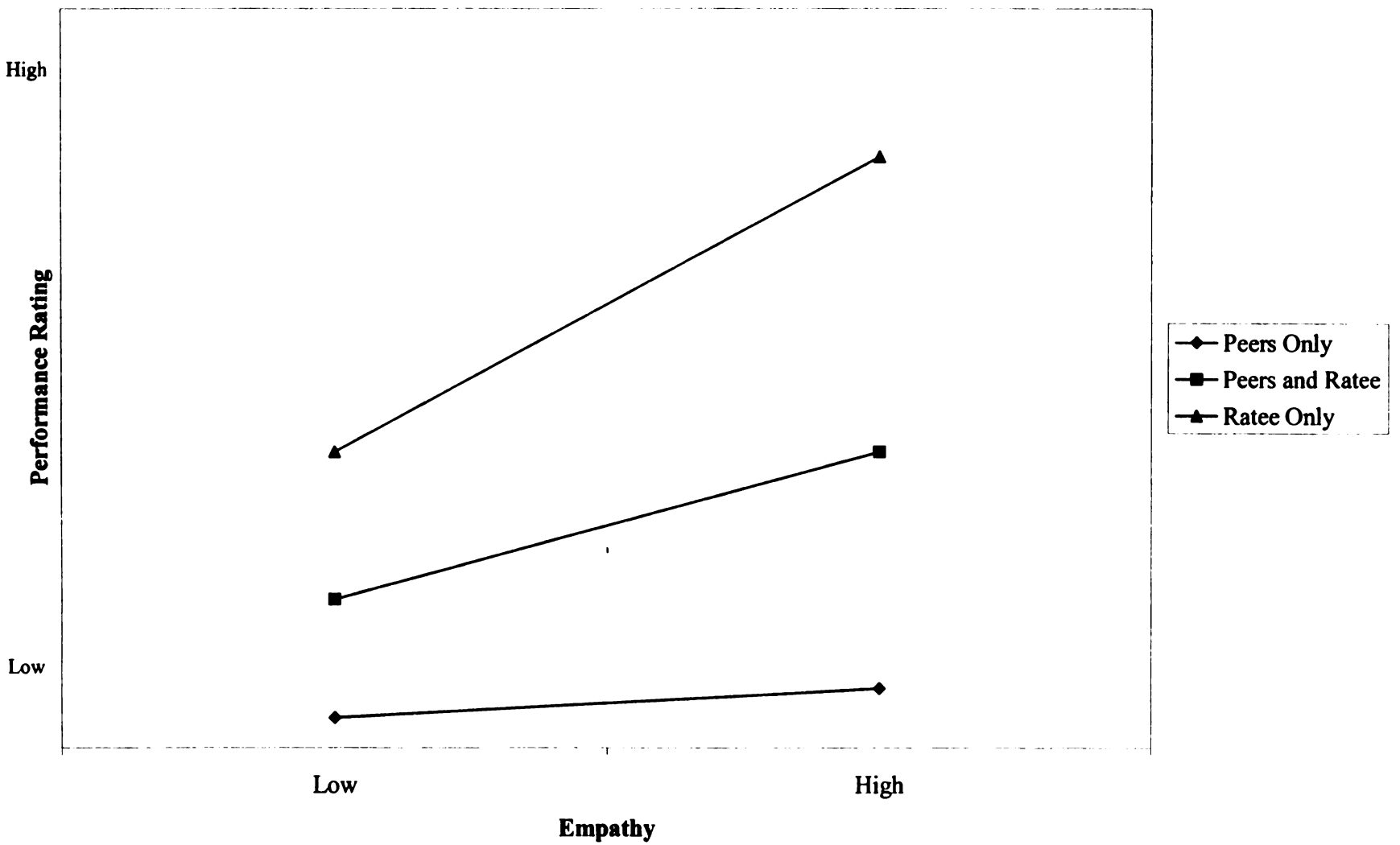
Epstein, 1972) one might expect that, all else equal, a rater with a high level of empathy would be less likely to provide accurate ratings of poor performers so as to protect their feelings. Someone with a low level of empathy, in contrast, may be more likely to provide accurate ratings of poor performers because he/she is generally insensitive to the feelings of others. Further, not all situations necessarily yield the same necessity to protect others' feelings. Consider, for example, a situation in which a rater is accountable to the ratee—receiving negative feedback in such a one-on-one conversation would likely put at risk the ratee's feelings to a greater extent than a situation in which such a conversation would not take place. Thus, raters' empathy—or their concern for others—was expected to interact with accountability in their effects on performance ratings provided by raters in a manner consistent with Figure 5.

Hypothesis 4: *Accountability and raters' levels of empathy would interact in their effects on performance ratings provided, such that:*

Hypothesis 4a: *The relationship between raters' empathy and ratings provided would be weak and positive when accountable to one's peers (excluding the ratee), strong and positive when accountable to the ratee (excluding the rater's peers), and moderately strong and positive when accountable to both one's peers and the ratee simultaneously.*

Hypothesis 4b: *Mean ratings provided would be lowest when accountable to one's peers (excluding the ratee), highest when accountable to the ratee (excluding the rater's peers), and moderate when accountable to both one's peers and the ratee simultaneously.*

Figure 5. Hypothesis 4: Accountability X Empathy Interaction On Performance Ratings



Research relating raters' levels of conscientiousness to the performance ratings that they provide for others generally demonstrates that conscientiousness relates negatively with performance ratings provided (Bernardin et al., 2000). That is, individuals with higher levels of conscientiousness tend to provide lower, presumably more accurate ratings of others' performance. Thus, with regard to the situational variables examined in the current study, these variables were not expected to have any effects on ratings provided by highly conscientious individuals. However, as raters' levels of conscientiousness decrease, it was expected that accountability would affect performance ratings in a manner consistent with Figure 6.

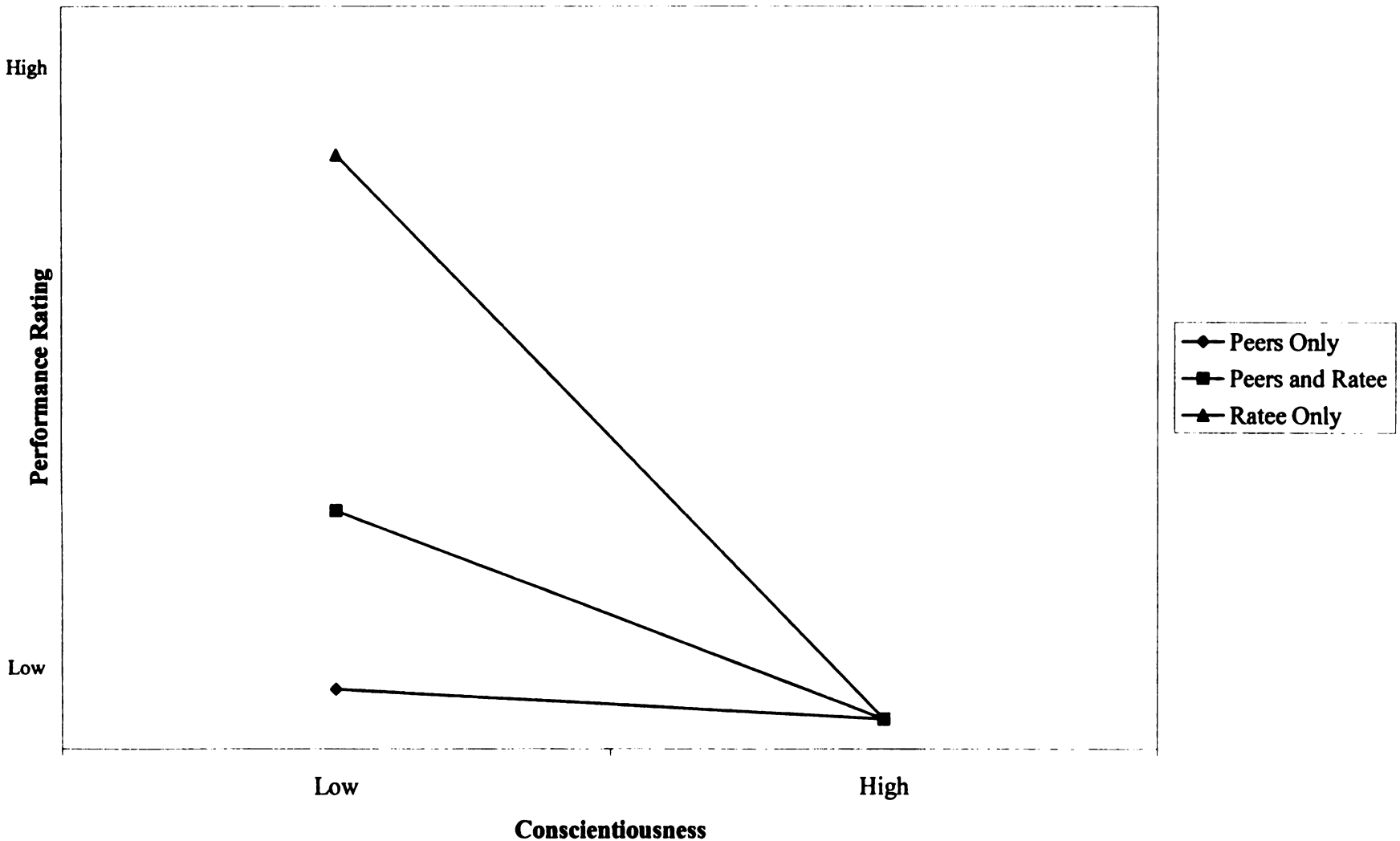
Hypothesis 5: *Accountability and raters' levels of conscientiousness would interact in their effects on performance ratings provided, such that:*

Hypothesis 5a: *The relationship between raters' conscientiousness and ratings provided would be null when accountable to one's peers (excluding the ratee), strong and negative when accountable to the ratee (excluding the rater's peers), and moderately strong and negative when accountable to both one's peers and the ratee simultaneously.*

Hypothesis 5b: *Mean ratings provided when conscientiousness is high would be equally low regardless of accountability.*

Work by Snyder (1974, 1979) shows that, compared to individuals with lower levels of self-monitoring, individuals with higher levels of self-monitoring exhibit greater social conformity by tailoring their behavior to fit social and interpersonal considerations of situational appropriateness—i.e., they modify their behavior to fit the situation. For example, Jawahar (2001) observed a negative correlation (-.72) between self-monitoring

Figure 6. Hypothesis 5: Accountability X Conscientiousness Interaction On Performance Ratings



and rating accuracy—thus, the more individuals self-monitored, the more likely they were to rate in accordance with the situation rather than honestly. Therefore, because of this heightened sensitivity to situational cues and demands, one might expect self-monitoring to interact with accountability. When a rater is accountable only to the ratee, he/she is likely to rate more leniently than when he/she is accountable to his/her peers (in which case he/she is likely to rate more fairly/honestly). Thus, consistent with Figure 7, it was expected that accountability and raters' levels of self-monitoring would interact in their effects on performance ratings provided.

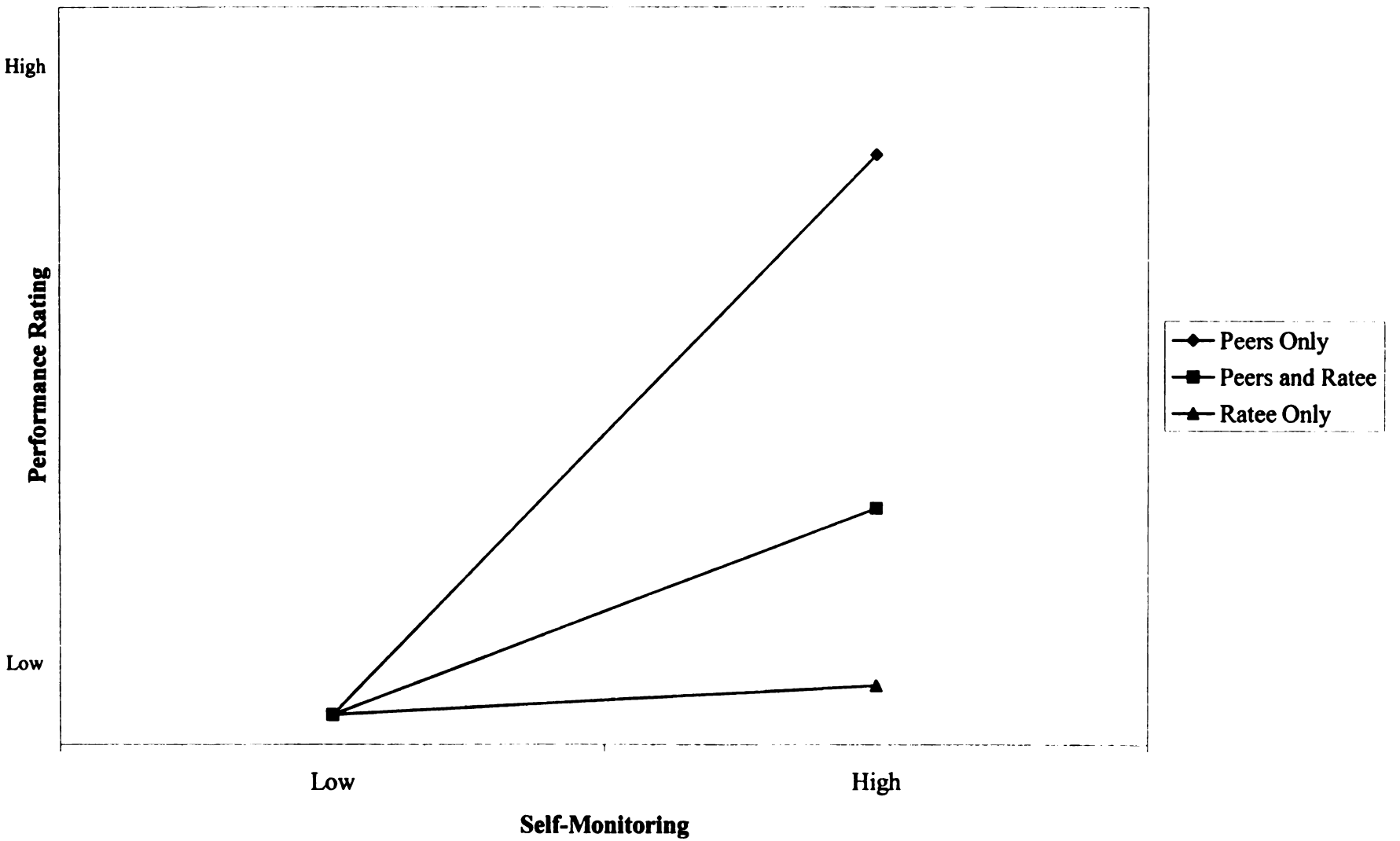
Hypothesis 6: *Accountability and raters' levels of self-monitoring would interact in their effects on performance ratings provided, such that:*

Hypothesis 6: *The relationship between raters' self-monitoring and ratings provided would be null when accountable to one's peers (excluding the ratee), strong and negative when accountable to the ratee (excluding the rater's peers), and moderately strong and negative when accountable to both one's peers and the ratee simultaneously.*

Hypothesis 6: *Mean ratings provided when self-monitoring is low would be equally low regardless of accountability.*

Finally, it was demonstrated earlier that individuals, when their egos are threatened, engage in behaviors that will restore their egos (e.g., Brown & Gallagher, 1992; Crocker et al., 1987; Fein & Spencer, 1997; Gibbons & Gerrard, 1991). As noted, a logical extension of this idea is that individuals who are typically more concerned with protecting their egos would not wait until it has been threatened—they would consistently behave in ways that will protect their egos. Thus, when situational characteristics

Figure 7. Hypothesis 6: Accountability X Self-Monitoring Interaction On Performance Ratings



demand accuracy, individuals who are highly concerned with protecting their egos would likely rate accurately so as to maintain their egos. When situational characteristics demand leniency, raters would likely rate leniently for the same reasons. Finally, consider the effects of a formal grievance policy on these relationships. If a rater knows that a ratee with have the opportunity to question and challenge the rating provided by the rater, this is likely to threaten a highly ego concerned rater to a greater extent than knowing that a rater will not have such an opportunity. Thus, accountability, raters' levels of ego concern, and the presence/absence of a grievance policy were expected to interact in their effects on performance ratings provided in a manner consistent with Figures 8 and 9.

***Hypothesis** 7: Accountability, raters' levels of ego concern, and the presence/absence of a grievance policy would interact in their effects on performance ratings provided, such that:*

***Hypothesis** 7a: In the presence of a grievance policy, the relationship between ego concern and ratings provided would be strong and negative when accountable to one's peers (excluding the ratee), strong and positive when accountable to the ratee (excluding the rater's peers), and null when accountable to both one's peers and the ratee simultaneously.*

***Hypothesis** 7b: In the absence of a grievance policy, accountability and raters' ego concern would interact in their effects on ratings provided in a manner similar to Hypothesis 7a, but these relationships would be attenuated.*

Figure 8. Hypothesis 7a: Accountability X Ego Concern Interaction On Performance Ratings (Grievance Policy Present)

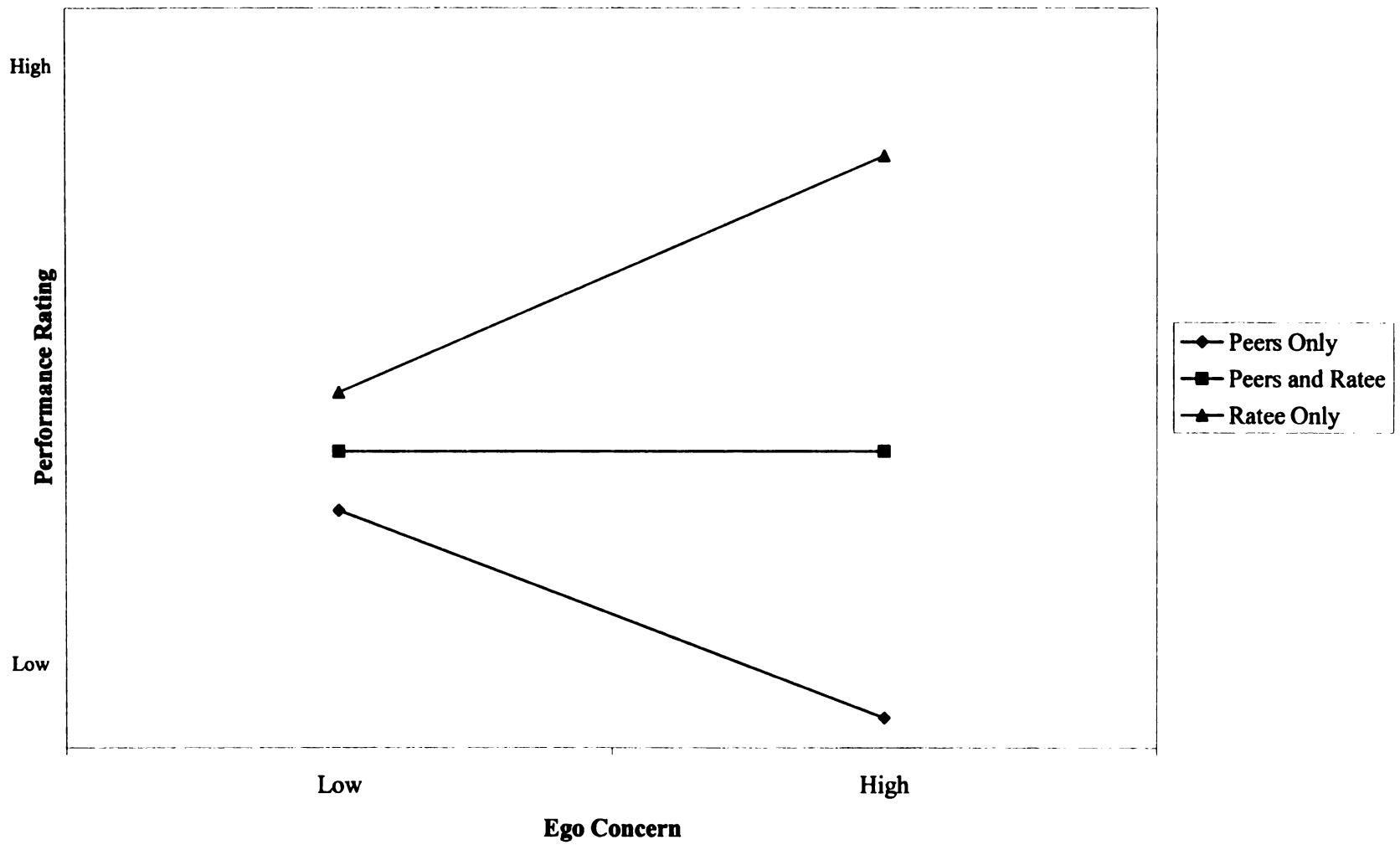
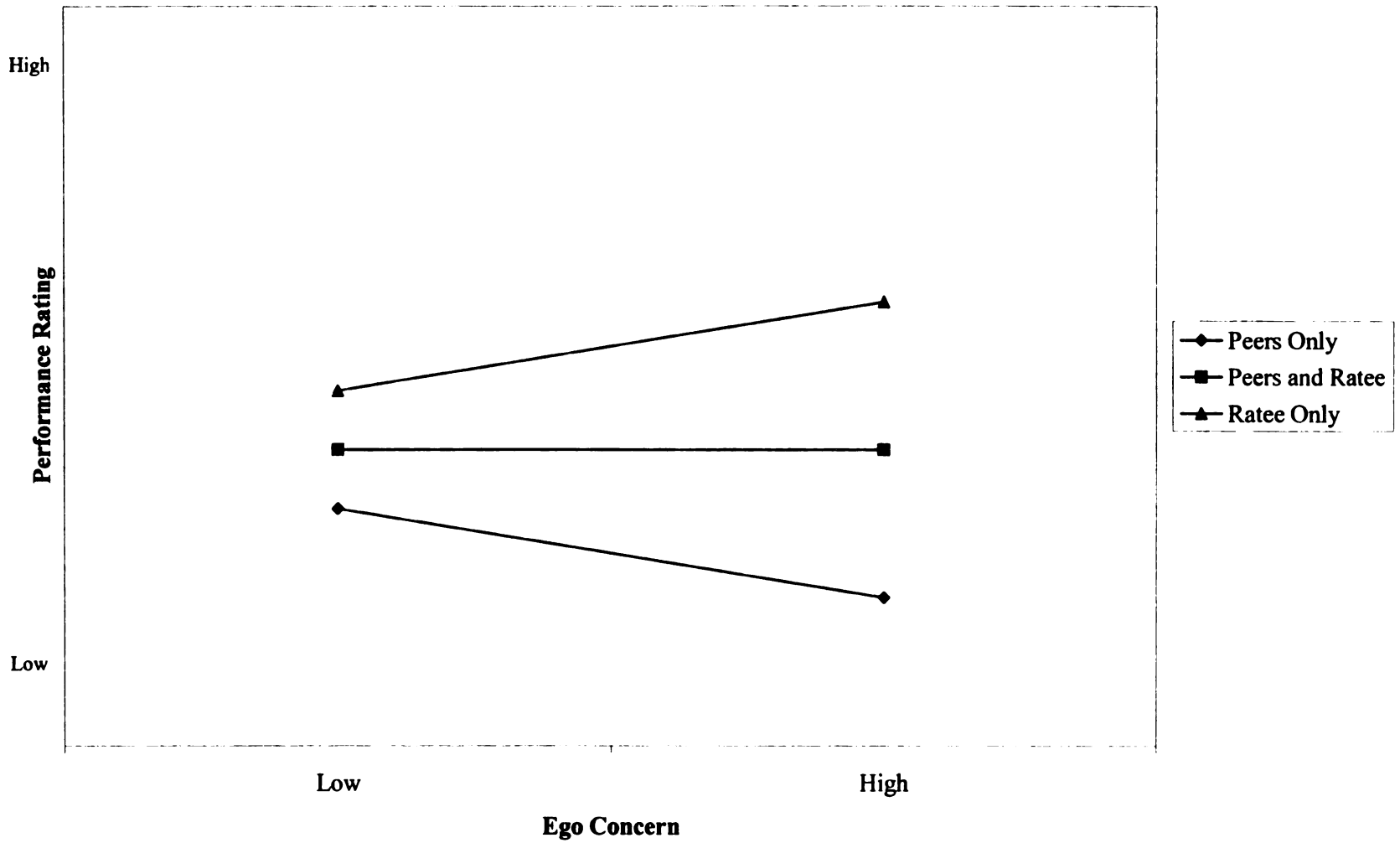


Figure 9. Hypothesis 7b: Accountability X Ego Concern Interaction On Performance Ratings (Grievance Policy Absent)



METHOD

Subjects and Study Design

Three hundred forty undergraduate psychology students participated as subjects in the current study in exchange for research participation credits. Eight subjects were removed from analyses because they had incomplete data prohibiting the matching of pre-experiment questionnaire data and session data. Of the 332 subjects in the final sample, approximately 65% (217) were female, 38% (125) were freshmen, and their mean age was 19.31 years (*SD* = 1.30 years). On average, subjects reported that they had little-some (*M* = 2.57, *SD* = 1.14; 1 = none, 2 = a little, 3 = some, 4 = a lot, 5 = very much) experience rating the performance of others. Finally, a sample size of 332 yields respectable levels of power (see Appendix A for power analysis results).

Subjects participated in groups of 3-7 (*M* = 5.99, *SD* = 1.03) people in this 4 (rater accountability) X 2 (presence/absence of a grievance policy) fully crossed design. Each group of subjects received one of the four levels of the accountability manipulation (1: accountable to peers only; 2: accountable to ratee only; 3: accountable to peers and ratee; 4: not accountable to anyone – control condition) and one of the two levels of the grievance policy manipulation (1: formal grievance policy was present; 2: formal grievance policy was absent). All subjects rated the performance of one of four poorly performing confederates, and subjects in the same group rated the same confederate.

Task Overview and Criterion Measure

Subjects formed groups upon arrival to the experiment and completed an adaptation of the Winter Survival Task (BusinessBalls.com, 2002). This task asked subjects to pretend that they were on an airplane that crashed in the wilderness in the

middle of winter. Group members were instructed to examine a list of 35 items (e.g., box of signal flares, shovel, electronic calculator) on board the airplane and choose the 10 most important items for survival before the plane burst into flames. Individuals were given 3 minutes to examine the list and choose the top 10 items independently from the rest of the group, and then the entire group had 10 minutes to discuss and agree upon a final list of 10 items on behalf of the entire group. Variations of this task have been used to study a variety of individual- and group-level phenomena, including goal setting (e.g., Durham, Locke, Poon, & McLeod, 2000) and group member ability and experience (e.g., Littlepage, Robison, & Reddington, 1997). The specific task used in the current study is displayed in Appendix B.

Confederate behavior. One of four poorly performing confederates participated in each group. Each confederate attended a brief confederate training session; the materials from this training session are contained in Appendix C. As shown in the appendix, confederates were trained to 1) choose "bad" items for their individual lists of the 10 items that the group should take, and 2) be unwilling to provide input after the initial list was read aloud. In an attempt to make the experiment the same across confederates, each confederate chose the exact same list of 10 items in every session. Some of the items on this list were jumper cables, Blockbuster rental card, and an electronic calculator. The remaining items on this list are displayed in Appendix C. The reason for using confederates was to guarantee that a consistent poor performer existed in each group. Finally, all confederates were males (between the ages of 20 and 22 years) so as to eliminate any potential effects that could result from having male vs. female poor performers.

Performance ratings. After subjects completed the Winter Survival Task, they rated each group member on three different dimensions: cooperativeness, contributions to the group, and overall performance. Subjects made each of these ratings on a scale from 1-100 (1 = extremely poor, 100 = excellent). These rating instructions are shown in Appendix D. Ultimately, the only ratings of interest were the ratings that each subject provided for the poor performer (i.e., the confederate).

Experimental Manipulations

Rater accountability. Groups received one of four levels of the rater accountability manipulation: 1) accountable only to one's peers, 2) accountable only to the ratee, 3) accountable to one's peers *and* the ratee simultaneously, and 4) not accountable to anyone—control. Eighty-nine subjects in 18 groups received the first level of this manipulation, 86 subjects in 20 groups received the second level, 82 subjects in 18 groups received the third level, and 75 subjects in 16 groups received the third level. This manipulation was operationalized by informing subjects that they would have to justify their ratings to group members excluding the ratee, only the ratee, the entire group including the ratee, or no one, respectively. Specifically, subjects in the first condition (i.e., accountable only to one's peers) were told that they would have to reveal to everyone their ratings (i.e., simply read off the numbers) for each group member, but that that they would then have to *justify* their ratings to the rest of the group (the ratee in question would leave the group while he/she was being discussed). For example, Group Member A would know how Group Member B rated him/her, but Group Member B would not have to justify this rating to Group Member A. Instead, Group Member B would have to justify this rating to Group Members C, D, E, F, and G. Thus, subjects

were led to believe that a ratee would know how a given rater rated him/her, but the rater would not have to justify these ratings to the ratee. Specific wording for this and other levels of this manipulation (crossed with the grievance policy manipulation discussed in the next section) are displayed in Appendix E.

Subjects in the second condition (i.e., accountable only to the ratee) were told that they would have to reveal to everyone their ratings for each group member, but that they would then have to *justify* their ratings to each ratee (rater and ratee would have a one on one discussion). For example, Group Members C, D, E, F, and G would know how Group Member B rated Group Member A, but Group Member B would not have to justify this rating to Group Members C, D, E, F, and G. Instead, Group Member B would only have to justify this rating to Group Member A. Thus, subjects were led to believe that the group would know how a given rater rated all group members, but the rater would only have to justify the ratings to respective group members.

Subjects in the third condition (i.e., accountable to one's peers *and* the ratee simultaneously) were told that they would have to justify their ratings to both the group as a whole as well as each ratee simultaneously. Subjects were told that they would go around in a circle revealing their ratings for each member in the group and providing their justifications. For example, Group Member B would have to justify his/her rating of Group Member A to Group Members A, C, D, E, F, and G simultaneously.

Finally, subjects in the fourth condition (i.e., not accountable to anyone—control) were told that they would have to reveal to everyone their ratings for each group member, but that they would not have to justify these ratings to anyone. Thus, ratings provided by subjects receiving this and every other level of the rater accountability manipulation were

not entirely anonymous. This was done to address the confounding of accountability and anonymity often found in performance appraisal research (e.g., Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975).

Presence of grievance policy. Groups received one of two levels of the presence of grievance policy manipulation (see Appendix E). One hundred sixty-seven subjects in 36 groups were told that group members *receiving* the justification would have a chance to challenge the rater on the rating in question. For example, if Group Member B was required to justify his/her rating of Group Member A *to* Group Member A, then Group Member A would have a chance to challenge Group Member B on this rating. If Group Member B was required to justify his/her rating of Group Member A to Group Members C, D, E, F, and G, then Group Members C, D, E, F, and G would have a chance to challenge Group Member B on this rating. Finally, if Group Member B was required to justify his/her rating of Group Member A to Group Members A, C, D, E, F, and G simultaneously, then Group Members A, C, D, E, F, and G would have a chance to challenge Group Member B on this rating. The remaining subjects (i.e., 165 subjects in 36 groups) received the second level of this manipulation and were told that no one would be allowed to question or challenge any of the ratings that they provide.

Measures

Rater accountability manipulation check. Three items were used to assess if subjects were aware of whether and to whom they would have to justify their performance ratings of others. An example item is "I am going to have to ***justify*** my rating of each specific group member ***to*** each specific group member (for example, I will have to justify my rating of Member A ***to*** Member A)." Subjects responded "yes" or

"no" to each of these questions. These three items and their instructions are displayed in Appendix F.

Grievance policy manipulation check. A single item was used to assess whether subjects were aware of a formal grievance policy regarding the performance ratings that they provided. This item asked subjects "Will group members have the opportunity to object to the ratings that you provide?" Subjects responded "yes" or "no" to this question. This item and its instructions are displayed in Appendix G.

Confederate behavior. Three items were used to assess subjects' perceptions of the confederate's behavior throughout the experiment. Specifically, subjects were asked to rate whether the confederate a) "resisted making contributions to the group," b) "took the experiment as a joke," and c) "chose ridiculous items as 'important items'." Subjects were asked to rate only the confederate on these behaviors. However, to reduce the chances that subjects would suspect the identity of the confederate and inform future subjects, subjects were led to believe that they were each rating a different group member when answering these questions. This was accomplished by printing these questions on paper of different colors. Each subject then received a different color, but the instructions on all colors indicated that the subject should rate "Member A" (the confederate). By receiving papers of different colors, it was expected that subjects would think they were doing something different than each other. These items and their instructions are displayed in Appendix H.

Goal valence. An 8-item scale developed specifically for this study was used to assess the extent to which individuals were concerned with being fair and honest (four items, $\alpha = .63$) vs. being liked by others (four items, $\alpha = .83$). An example item used to

measure the extent to which individuals were concerned with being fair and honest is "It is important to me that I provide fair and honest ratings of my group members." An example item used to measure the extent to which individuals were concerned with being liked by others is "It is important to me that my group members like me." Subjects responded to these items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). These items and their instructions are displayed in Appendix I.

Agreeableness. A 10-item (α = .78) scale from the International Personality Item Pool (IPIP, 2001) was used to assess individuals' levels of agreeableness. An example item is "I have a soft heart." Subjects responded to these items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). These items and their instructions are displayed in Appendix J.

Conscientiousness. A 10-item (α = .83) scale from the International Personality Item Pool (IPIP, 2001) was used to assess individuals' levels of conscientiousness. An example item is "I pay attention to details." Subjects responded to these items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). These items and their instructions are displayed in Appendix K.

Empathy. A 7-item (α = .71) subscale of Davis's (1983) Interpersonal Reactivity Index was used to assess individuals' levels of empathy. An example item is "I often have tender, concerned feelings for people less fortunate than me." Subjects responded to these items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). These items and their instructions are displayed in Appendix L.

Self-monitoring. A 23-item (α = .68) scale adapted from Snyder (1974) was used to assess individuals' levels of self-monitoring. An example item is "My behavior is

usually an expression of my true inner feelings, attitudes, and beliefs." Subjects responded to these items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). These items and their instructions are displayed in Appendix M.

Ego concern. A 6-item (α = .71) scale developed specifically for this study was used to assess individuals' levels of ego concern. An example item is "Before I decide to do something, I consider what others will think of me." Subjects responded to these items on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). These items and their instructions are displayed in Appendix N.

Demographics. Demographic information regarding subjects' age, sex, year in school, major, grade point average, ACT or SAT scores, and plans after graduation was collected in order to explore post hoc explanations if needed. Also included in the set of demographics questions were two items assessing subjects' work experience and familiarity with rating the performance of others. These items and their instructions are shown in Appendix O.

Procedure

Prior to the experiment (i.e., when subjects signed up for the experiment on the internet), subjects filled out the agreeableness (see Appendix J), conscientiousness (see Appendix K), empathy (see Appendix L), self-monitoring (see Appendix M), ego concern (see Appendix N), and demographics (see Appendix O) measures. Upon arrival to the experiment, subjects provided their written consent to participate in the experiment (see Appendix P), and they were formed into groups of no more than seven people. Each group member was given a name card with a letter from A to G on the front; subjects were asked to assume this letter as their identity throughout the experiment and were

instructed to use these letters to refer to each other. A confederate was always included in each group. Subjects were assigned to receive (verbally) one level of each of the two manipulations (see Appendix E). Following these manipulations, subjects completed the Winter Survival Task (see Appendix B). Subjects then filled out the accountability (see Appendix F) and grievance policy (see Appendix G) manipulation checks, followed by the goal valence (see Appendix I) measure. Following these questionnaires, subjects rated the performance of their group members (see Appendix D) and justified these ratings as specified by the manipulations. After subjects had a chance to justify their performance ratings to the group members, subjects completed the confederate behavior measure (see Appendix H). To reduce the possibility of subjects learning that there was a confederate in the group, the confederate behavior measure was printed on a different colored piece of paper for everyone. Again, by receiving papers of different colors, it was expected that subjects would think they were doing something different than each other, thus reducing the chance of subjects identifying the confederate as such and informing future subjects. After completing this final measure, subjects were debriefed (see Appendix Q) fully and thanked for their time.

RESULTS

Means, standard deviations, scale internal consistency reliabilities, and intercorrelations among the variables of interest are displayed in Table 2. Table 3 displays raw means and standard deviations for performance ratings by condition.

Manipulation Checks

Rater accountability. Subsequent to receiving the rater accountability manipulation, subjects responded to three manipulation check questions: 1) Will the rater (subject) have to justify his/her ratings of ratees to his/her peers?, 2) Will the rater have to justify his/her rating of ratees *to* ratees?, and 3) Will the rater have to justify his/her ratings of ratees to both his/her peers and ratees? (see Appendix F for specific wording of these questions). Table 4 summarizes subjects' responses to these questions. With regard to subjects who were accountable to their peers (i.e., Level 1 of the rater accountability manipulation), a chi-square testing the null hypothesis that the percentages of "No" and "Yes" responses to the first manipulation check question (i.e., Justify ratings to peers?) are equal indicates that more subjects answered "Yes" (95.5%) to this question ($\chi^2 = 73.719, p < .01$) as expected. Also as expected, a chi-square of "No" and "Yes" responses to the second manipulation check question (i.e., Justify ratings to ratees?) indicates that more subjects answered "No" (94.4%) to this question ($\chi^2 = 70.124, p < .01$). Finally, a third chi-square of "No" and "Yes" responses to the third manipulation check question (i.e., Justify ratings to peers *and* ratees?) indicates that more subjects answered "No" (94.4%) to this question ($\chi^2 = 70.124, p < .01$) as expected.

Second, with regard to subjects who were accountable to ratees (i.e., Level 2 of the rater accountability manipulation), a chi-square of "No" and "Yes" responses to the

Table 2

Means, Standard Deviations, and Intercorrelations Among Variables.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1. Cooperation Rating	50.61	40.31	-									
2. Contributions Rating	24.63	32.81	.62**	-								
3. Overall Performance Rating	33.58	33.89	.74**	.89**	-							
4. Fairness Goal Valence	3.61	.64	-.02	-.06	-.05	(.63)						
5. Liking Goal Valence	2.72	.71	.06	.12*	.14**	.36**	(.82)					
6. Agreeableness	3.97	.46	.03	-.02	.01	.16**	.24**	(.78)				
7. Conscientiousness	3.48	.59	-.01	.00	.01	.11*	.23**	.23**	(.83)			
8. Self-Monitoring	3.04	.34	.08	.05	.08	-.03	.09	.01	-.21**	(.68)		
9. Empathy	3.75	.48	.05	-.01	.02	.20**	.17**	.68**	.12*	.01	(.71)	
10. Ego Concern	3.36	.57	.01	-.07	-.04	-.03	.02	.13*	-.02	.11*	.06	(.71)
11. Group Size	5.99	1.03	-.14**	-.24**	-.22**	.05	-.02	.03	.03	-.04	.06	-.05

Note: Scale reliabilities are reported along the diagonal. * $p < .05$; ** $p < .01$

Table 3

Summary of Raw Performance Ratings By Condition.

Condition	Cooperation		Contributions		Overall Performance	
	M	*SD*	*M*	*SD*	*M*	*SD*
All	50.61	40.31	24.63	32.81	33.58	33.89
Grievance Policy Present						
Accountable to Peers	46.55	45.68	20.66	31.55	29.68	35.75
Accountable to Ratee	49.28	36.96	23.00	28.87	31.93	32.25
Accountable to Peers and Ratee	47.81	39.71	27.23	34.32	35.88	35.13
Not Accountable to Anyone	56.65	40.90	30.84	36.80	40.59	35.87
Grievance Policy Absent						
Accountable to Peers	43.91	37.37	18.42	27.41	30.69	27.84
Accountable to Ratee	50.70	39.56	22.77	34.01	30.35	33.96
Accountable to Peers and Ratee	51.69	40.32	24.36	33.37	37.03	33.56
Not Accountable to Anyone	60.87	42.44	31.79	36.74	34.05	38.15

Table 4

Rater Accountability Manipulation Check: Frequencies of Responses to Justification Questions.

Condition/Question	n	No	Yes	χ^{2a}
Manipulation: Accountable to Peers				
Question: Justify to Peers?	89	4 (4.5%)	85 (95.5%)	**73.719****
Question: Justify to Ratee?	89	84 (94.4%)	5 (5.6%)	**70.124****
Question: Justify to Peers and Ratee?	89	84 (94.4%)	5 (5.6%)	**70.124****
Manipulation: Accountable to Ratee				
Question: Justify to Peers?	86	82 (95.3%)	4 (4.7%)	**70.744****
Question: Justify to Ratee?	86	5 (5.8%)	81 (94.2%)	**67.163****
Question: Justify to Peers and Ratee?	86	68 (79.1%)	18 (20.9%)	**14.535****
Manipulation: Accountable to Peers and Ratee				
Question: Justify to Peers?	82	56 (68.3%)	26 (31.7%)	10.976**
Question: Justify to Ratee?	82	7 (8.5%)	75 (91.5%)	**56.390****
Question: Justify to Peers and Ratee?	82	1 (1.2%)	81 (98.8%)	**78.049****
Manipulation: Not Accountable to Anyone				
Question: Justify to Peers?	75	71 (94.7%)	4 (5.3%)	**59.853****
Question: Justify to Ratee?	75	65 (86.7%)	10 (13.3%)	**40.333****
Question: Justify to Peers and Ratee?	75	57 (76.0%)	18 (24.0%)	**10.140****

[a] Chi-Square tests the null hypothesis that percentages of "No" and "Yes" repsonses are equal (50%). Bold values indicate that associated percentages are in the expected direction.
** $p < .01$

first manipulation check question (i.e., Justify ratings to peers?) indicates that more subjects answered "No" (95.3%) to this question ($\chi^2 = 70.744, p < .01$) as expected. A second chi-square of "No" and "Yes" responses to the second manipulation check question (i.e., Justify ratings to ratees?), also as expected, indicates that more subjects answered "Yes" (94.2%) to this question ($\chi^2 = 67.163, p < .01$). Finally, as expected, a third chi-square of "No" and "Yes" responses to the third manipulation check question (i.e., Justify ratings to peers *and* ratees?) indicates that more subjects answered "No" (79.1%) to this question ($\chi^2 = 14.535, p < .01$).

Third, with regard to subjects who were accountable to their peers and ratees (i.e., Level 3 of the rater accountability manipulation), a chi-square of "No" and "Yes" responses to the first manipulation check question (i.e., Justify ratings to peers?) indicates that more subjects answered "No" (68.3%) to this question ($\chi^2 = 10.976, p < .01$); this finding is counter to expectations. Second, as expected, a chi-square of "No" and "Yes" responses to the second manipulation check question (i.e., Justify ratings to ratees?) indicates that more subjects answered "Yes" (91.5%) to this question ($\chi^2 = 56.390, p < .01$). Finally, a third chi-square of "No" and "Yes" responses to the third manipulation check question (i.e., Justify ratings to peers *and* ratees?) indicates that more subjects answered "Yes" (98.8%) to this question ($\chi^2 = 789.049, p < .01$) as expected. The unexpected pattern of results pertaining to the first question (i.e., Justify ratings to peers?) asked of subjects in this condition is not particularly problematic given that they were able to answer the third question (i.e., Justify ratings to peers *and* ratees?) correctly.

Finally, with regard to subjects who were not accountable to anyone (i.e., Level 4 of the rater accountability manipulation), a chi-square of "No" and "Yes" responses to the

first manipulation check question (i.e., Justify ratings to peers?) indicates that more subjects answered "No" (94.7%) to this question ($\chi^2 = 59.853, p < .01$) as expected. A second chi-square of "No" and "Yes" responses to the second manipulation check question (i.e., Justify ratings to ratees?), also as expected, indicates that more subjects answered "No" (86.7%) to this question ($\chi^2 = 40.333, p < .01$). Finally, as expected, a third chi-square of "No" and "Yes" responses to the third manipulation check question (i.e., Justify ratings to peers *and* ratees?) indicates that more subjects answered "No" (76.0%) to this question ($\chi^2 = 10.140, p < .01$). Thus, with the exception of responses to the first manipulation check question (i.e., Justify ratings to peers?) from subjects who were accountable to their peers and ratees (i.e., Level 1 of the rater accountability manipulation), subjects' responses to the three manipulation check questions indicate that the rater accountability manipulation was successful.

Grievance policy. Subsequent to receiving the grievance policy manipulation, subjects were asked whether individuals (to whom ratings would be justified) would have the opportunity to challenge the ratings provided (see Appendix G for specific wording of this questions). Table 5 summarizes subjects' responses to this question. With regard to subjects in the Policy Absent condition, a chi-square testing the null hypothesis that the percentages of "No" and "Yes" responses to this question are equal indicates that more subjects answered "No" (90.3%) to this question ($\chi^2 = 107.206, p < .01$) as expected. A second chi-square of "No" and "Yes" responses to this question from subjects in the Policy Present condition indicates that more subjects answered "Yes" (93.4%) to this question ($\chi^2 = 125.898, p < .01$). Thus, subjects' responses to this manipulation check question indicate that the grievance policy manipulation was successful.

Table 5

Grievance Policy Manipulation Check: Frequencies of Responses to Presence of Policy Question[a]

Condition	n	No	Yes	χ^{2b}
Grievance Policy Absent	165	149 (90.3%)	16 (9.7%)	**107.206****
Grievance Policy Present	167	11 (6.6%)	156 (93.4%)	**125.898****

[a] Question: Is there a grievance policy?

[b] Chi-Square tests the null hypothesis that percentages of "No" and "Yes" repsonses are equal (50%). Bold values indicate that associated percentages are in the expected direction.

** $p < .01$

Confederate behavior. At the end of the experiment subjects rated three aspects of the confederate's behavior: 1) whether the confederate resisted making contributions, 2) whether the confederate took the experiment as a joke, and 3) whether the confederate chose "ridiculous" items as "important" (see Appendix H for specific wording of these questions). Table 6 summarizes subjects' responses to these questions. With regard to the first of these questions (i.e., Resisted making contributions?), a chi-square testing the null hypothesis that the percentages of "No" and "Yes" responses to this first question are equal indicates that more subjects answered "Yes" (80.5%) to this question (χ^2 = 122.799, $p < .01$) as expected. Also as expected, a chi-square of "No" and "Yes" responses to the second question (i.e., Took experiment as a joke?) indicates that more subjects answered "Yes" (96.0%) to this question ($\chi^2 = 278.061$, $p < .01$). Finally, a third chi-square of "No" and "Yes" responses to the third question (i.e., Chose ridiculous items as important?) indicates that more subjects answered "Yes" (97.9%) to this question (χ^2 = 302.594, $p < .01$) as expected. Table 7 summarizes subjects' responses to these same questions, reported separately for each confederate. As seen in the table, the pattern of results reported above is consistent across confederates.

Finally, because four different confederates were utilized in this study, it was necessary to test whether different confederates received different performance ratings from subjects. Interactions between dummy-coded confederate and condition (i.e., accountability crossed with grievance policy) variables were not significant predictors of any of the outcome variables (Cooperation: F (18, 302) = 1.40, $p > .05$; Contributions: F (18, 302) = 0.78, $p > .05$; Overall Performance: F (18, 302) = 1.20, $p > .05$). Therefore, these results, paired with the results of subjects' ratings of various aspects of

Table 6

Confederate Behavior Check: Frequencies of Responses to Confederate Behavior Questions.

Question	n	No	Yes	χ^{2a}
Resisted Making Contributions?	329	64 (19.5%)	265 (80.5%)	**122.799****
Took Experiment As A Joke?	328	13 (4.0%)	315 (96.0%)	**278.061****
Chose Ridiculous Items as "Important"	330	7 (2.1%)	323 (97.9%)	**302.594****

[a] Chi-Square tests the null hypothesis that percentages of "No" and "Yes" repsonses are equal (50%). Bold values indicate that associated percentages are in the expected direction.
** $p < .01$

Table 7

Confederate Behavior Check: Frequencies of Responses to Confederate Behavior Questions, Separated By Confederate.

Confederate/Question	n	No	Yes	χ^{2a}
Confederate 1				
Resisted Making Contributions	59	8 (13.6%)	51 (86.4%)	**31.339****
Took Experiment As A Joke?	59	1 (1.7%)	58 (98.3%)	**55.068****
Chose Ridiculous Items as "Important"	59	0 (0.0%)	59 (100.0%)	**59.000****
Confederate 2				
Resisted Making Contributions	104	7 (6.7%)	97 (92.3%)	**77.885****
Took Experiment As A Joke?	104	3 (2.9%)	101 (97.1%)	**92.346****
Chose Ridiculous Items as "Important"	104	3 (2.9%)	101 (97.1%)	**92.346****
Confederate 3				
Resisted Making Contributions	118	43 (36.4%)	75 (63.6%)	**8.678****
Took Experiment As A Joke?	117	8 (6.8%)	109 (93.2%)	**87.188****
Chose Ridiculous Items as "Important"	119	3 (2.5%)	116 (97.5%)	**107.303****
Confederate 4				
Resisted Making Contributions	48	6 (12.5%)	42 (87.5%)	**27.000****
Took Experiment As A Joke?	48	1 (2.1%)	47 (97.9%)	**44.083****
Chose Ridiculous Items as "Important"	48	1 (2.1%)	47 (97.9%)	**44.083****

[a] Chi-Square tests the null hypothesis that percentages of "No" and "Yes" repsonses are equal (50%). Bold values indicate that associated percentages are in the expected direction.

** $p < .01$

confederates' behavior suggest that subjects did not perceive or rate the four confederates differently. Regardless, a dummy-coded confederate variable was included in the hierarchical linear modeling (HLM) analyses discussed below.

Overview of Analyses

HLM (Bryk & Raudenbush, 1987; Bryk & Raudenbush, 1992) is a set of statistical procedures for analyzing datasets that are *nested* in structure—i.e., information regarding the group (or individual) to which an individual (or observation) belongs tells us something about that individual (or observation). Examples of nested data structures include individuals nested within teams, students nested within schools, individuals nested within workgroups nested within organizations, and observations nested within individuals. In the current dataset, individuals are nested within groups, which are nested within confederates. Knowing that a particular individual is a member of a particular group and experienced a particular confederate tells us something about that individual—one would expect individuals from the same groups and confederates to be more similar to each other with regard to some outcome compared to individuals from different groups and confederates.

Computing an ordinary least squares (OLS) regression on nested datasets forces one to ignore that observations are in fact nested; dependent variables (e.g., performance ratings) would be regressed onto independent variables (e.g., rater accountability and grievance policy manipulations), ignoring group membership. Because of the nested data structure, resulting errors of prediction (residuals) would be more similar for observations from the same group and confederate, thus resulting in correlated residuals—a problem

for OLS regression (Bliese, 2002). By analyzing such datasets using HLM procedures, the nesting and correlated residuals are taken into consideration.

Therefore, each of the experimental hypotheses advanced previously were analyzed using SAS PROC MIXED, taking into consideration the nesting of individuals within groups and groups within confederates. When graphing interactions including continuous independent variables (i.e., individual differences), these graphs were created at plus and minus one standard deviation on the continuous variables. Additionally, analyses utilizing any of the three performance ratings (cooperation, contributions, or overall performance) as dependent variables included group size as a covariate due to the significant ($p < .01$) negative correlations between group size and performance ratings (see Table 2 for exact relationships). Because none of the demographics variables related significantly to the outcome variables, these variables did not serve as covariates in any of the analyses.

Hypothesis 1: Accountability and Goal Valence

Hypothesis 1 predicted that the rater accountability manipulation would affect individuals' levels of fairness and liking goal valence, such that raters accountable only to their peers would value a fairness goal to a greater extent than a liking goal (Hypothesis 1a), raters accountable only to ratees would value a fairness goal to a lesser extent than a liking goal (Hypothesis 1b), and raters accountable to both their peers and ratees would value a fairness goal and a liking goal to an equal extent (Hypothesis 1c). These hypotheses were analyzed through Panel Analysis, an extension of HLM that allows for the comparison of multiple dependent variables (i.e., repeated measurements) in the same model.

Table 8 summarizes the results obtained for Hypothesis 1. Consistent with Hypothesis 1a, raters who were accountable only to their peers did indeed value being fair to a significantly greater extent than being liked (F (1,17) = 74.87, $p < .01$). However, counter to Hypotheses 1b and 1c, respectively, this same pattern also emerged when raters were accountable only to ratees (F (1,19) = 91.06, $p < .01$) as well as when they were accountable to their peers and ratees (F (1,17) = 76.34, $p < .01$). Thus, Hypothesis 1 was not supported.

Hypothesis 2: Goal Valence and Performance Ratings

Hypothesis 2 predicted that individuals' levels of fairness and liking goal valence would relate to the performance ratings that they provide, such that liking goal valence would relate positively with performance ratings (Hypothesis 2a) and fairness goal valence would relate negatively with performance ratings (Hypothesis 2b). As seen in Table 9, individuals' levels of liking goal valence related positively with all three performance ratings (Cooperation: t (259) = 0.84, $p > .05$; Contributions: t (259) = 2.06, $p < .01$; Overall Performance: t (259) = 2.30, $p < .01$), but this relationship was significant only with respect to Contributions and Overall Performance; these findings provide partial support for Hypothesis 2a. Consistent with Hypothesis 2b, individuals' levels of fairness goal valence related negatively with Contributions (t (259) = -0.42, $p > .05$) and Overall Performance (t (259) = -0.59, $p > .05$), but neither of these relationships reached statistical significance. Counter to expectations, individuals' levels of fairness goal valence related positively (t (259) = 0.29, $p > .05$) with Cooperation, but this relationship did not reach statistical significance. Thus, these results provide only partial support for Hypothesis 2.

Table 8

Hypothesis 1: Estimated Least Squared Means of Goal Valence Levels For Each Accountability Condition.

Condition	n	Being Liked	Being Fair	F[a]	*df*
Accountable to Peers	89	2.78	3.67	**74.87****	1,17
Accountable to Ratee	86	2.76	3.67	91.06**	1,19
Accountable to Peers and Ratee	82	2.67	3.55	76.34**	1,17
Not Accountable to Anyone	75	2.67	3.56	63.05**	1,15

[a] Bold values indicate that the associated effect is in the hypothesized direction.
** $p < .01$

Table 9

Hypothesis 2: Summary of Relationships (t-Values)[a] Between Goal Valence and Performance Ratings, Controlling for Group Size.

	Cooperation		Contributions		Overall Performance	
Goal	*t*	R^2	*t*	R^2	*t*	R^2
Being Liked	**0.84**	.00	**2.06***	.02	**2.30***	.02
Being Fair	0.29	.00	**-0.42**	.00	**-0.59**	.00

[a] $df = 259$. Bold values indicate that the associated effect is in the hypothesized direction.

* $p < .05$

Hypothesis 3: Accountability, Grievance Policy, Agreeableness, and Performance Ratings

Hypothesis 3 predicted that accountability, presence/absence of a grievance policy, and raters' levels of agreeableness would interact in their effects on performance ratings provided in a manner consistent with Figures 3 (Hypothesis 3a) and 4 (Hypothesis 3b). As indicated in Tables 10-12, this three-way interaction was not observed for any of the outcome variables (Cooperation: F (2,195) = 0.32, p > .05; Contributions: F (2,195) = 1.22, p > .05; Overall Performance: F (2,195) = 0.75, p > .05). Regardless, Hypotheses 3a (grievance policy present) and 3b (grievance policy absent) were examined for exploratory purposes.

Grievance policy present (Hypothesis 3a). With regard to Cooperation, neither accountability (F (2,24) = 0.00, p > .05) nor raters' levels of agreeableness (F (1,101) = 0.47, p > .05) had significant main effects on performance ratings provided. Further, counter to Hypothesis 3a, these two variables did not interact significantly (F (2,99) = 0.84, p > .05) in their effects on performance ratings provided. These results are displayed in Table 10 and Figure 10.

Similarly, neither accountability (F (2,24) = 0.17, p > .05) nor raters' levels of agreeableness (F (1,101) = 0.92, p > .05) had significant main effects on Contributions ratings provided, nor did these two variables interact significantly (F (2,99) = 0.96, p > .05) in their effects on performance ratings provided. These results are displayed in Table 11 and Figure 11.

Finally, with regard to ratings of Overall Performance, neither accountability (F (2,24) = 0.11, p > .05) nor raters' levels of agreeableness (F (1,101) = 0.43, p > .05) had

Table 10

Hypothesis 3: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Agreeableness Effects On Ratings of Cooperation.

	Accountability			Agreeableness				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability X Policy X Agreeableness						0.32	2,195	.00
Accountability								
Grievance Policy Present	46.70	43.53	50.04			0.00	2,24	.00
Grievance Policy Absent	72.34	73.18	77.85			0.22	2,24	.01
Agreeableness								
Grievance Policy Present				44.56	40.18	0.47	1,101	.00
Grievance Policy Absent				69.23	67.67	0.26	1,98	.00
Accountability X Agreeableness								
Grievance Policy Present						0.84	2,99	.01
Low Agreeableness	61.92	66.89	70.71					
High Agreeableness	75.34	70.68	65.72					
Grievance Policy Absent						1.28	2,96	.01
Low Agreeableness	67.32	86.44	85.52					
High Agreeableness	82.40	74.35	81.18					

Table 11

Hypothesis 3: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Agreeableness Effects On Ratings of Contributions.

	Accountability			Agreeableness				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability X Policy X Agreeableness						1.22	2,195	.01
Accountability								
Grievance Policy Present	72.23	72.04	71.69			0.17	2,24	.01
Grievance Policy Absent	78.65	83.15	85.99			0.14	2,24	.01
Agreeableness								
Grievance Policy Present				68.48	72.59	0.92	1,101	.01
Grievance Policy Absent				72.16	75.80	0.10	1,98	.00
Accountability X Agreeableness								
Grievance Policy Present						0.96	2,99	.01
Low Agreeableness	46.64	46.29	57.37					
High Agreeableness	48.80	43.94	44.88					
Grievance Policy Absent						4.28*	2,96	.04
Low Agreeableness	68.33	61.71	89.26					
High Agreeableness	72.27	78.72	67.77					

* $p < .05$

Table 12

Hypothesis 3: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Agreeableness Effects On Ratings of Overall Performance.

	Accountability			Agreeableness				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability X Policy X Agreeableness						0.75	2,195	.00
Accountability								
Grievance Policy Present	62.37	59.21	65.30			0.11	2,24	.00
Grievance Policy Absent	80.76	77.31	86.40			0.32	2,24	.01
Agreeableness								
Grievance Policy Present				59.29	56.27	0.43	1,101	.00
Grievance Policy Absent				70.06	71.66	0.11	1,98	.00
Accountability X Agreeableness								
Grievance Policy Present						0.60	2,99	.01
Low Agreeableness	61.06	62.20	69.91					
High Agreeableness	64.64	57.17	61.89					
Grievance Policy Absent						1.60	2,96	.02
Low Agreeableness	76.03	67.08	90.63					
High Agreeableness	80.38	81.19	80.27					

Figure 10. Hypothesis 3a (Results): Accountability X Agreeableness Interaction On Cooperation Ratings (Grievance Policy Present)

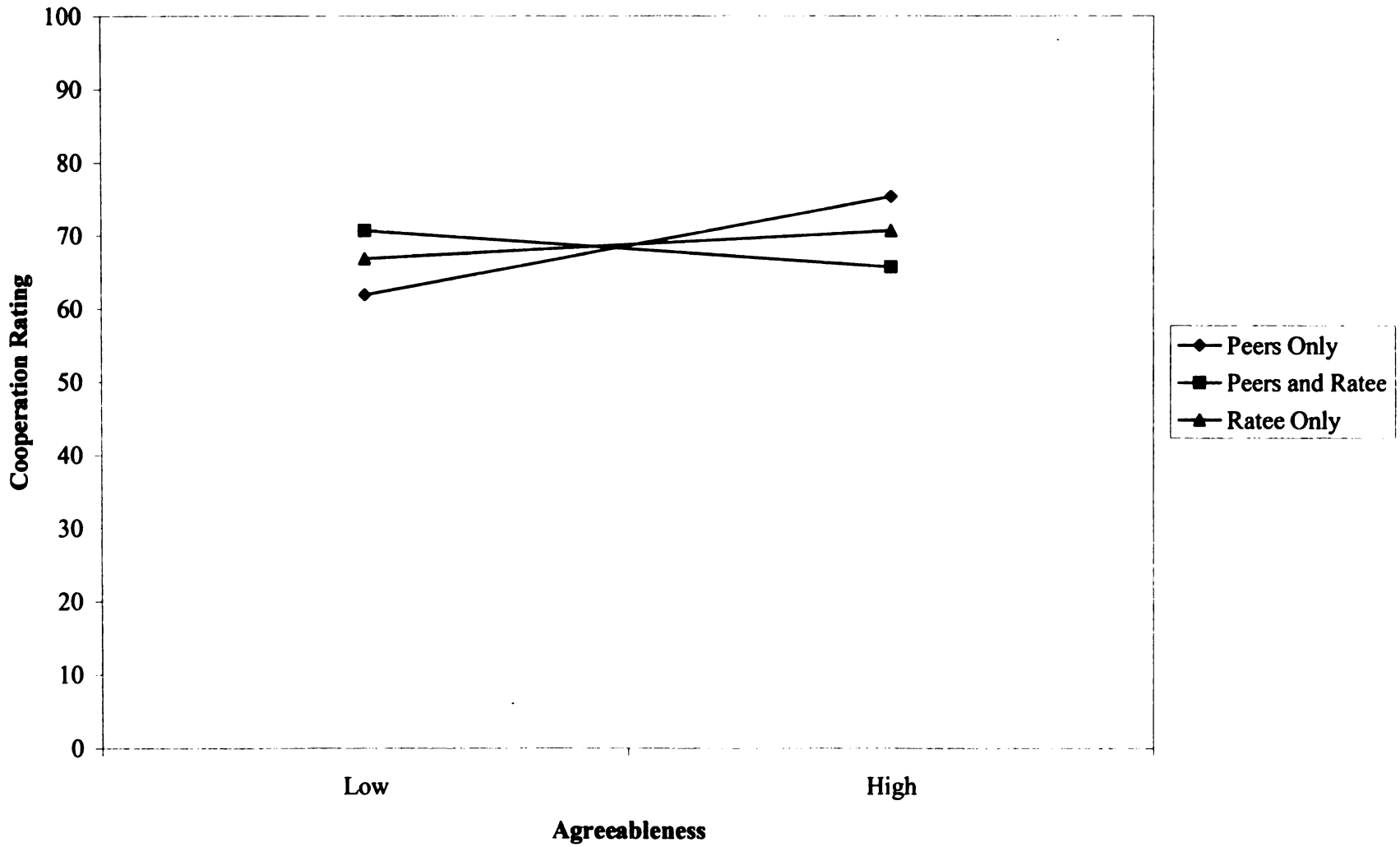
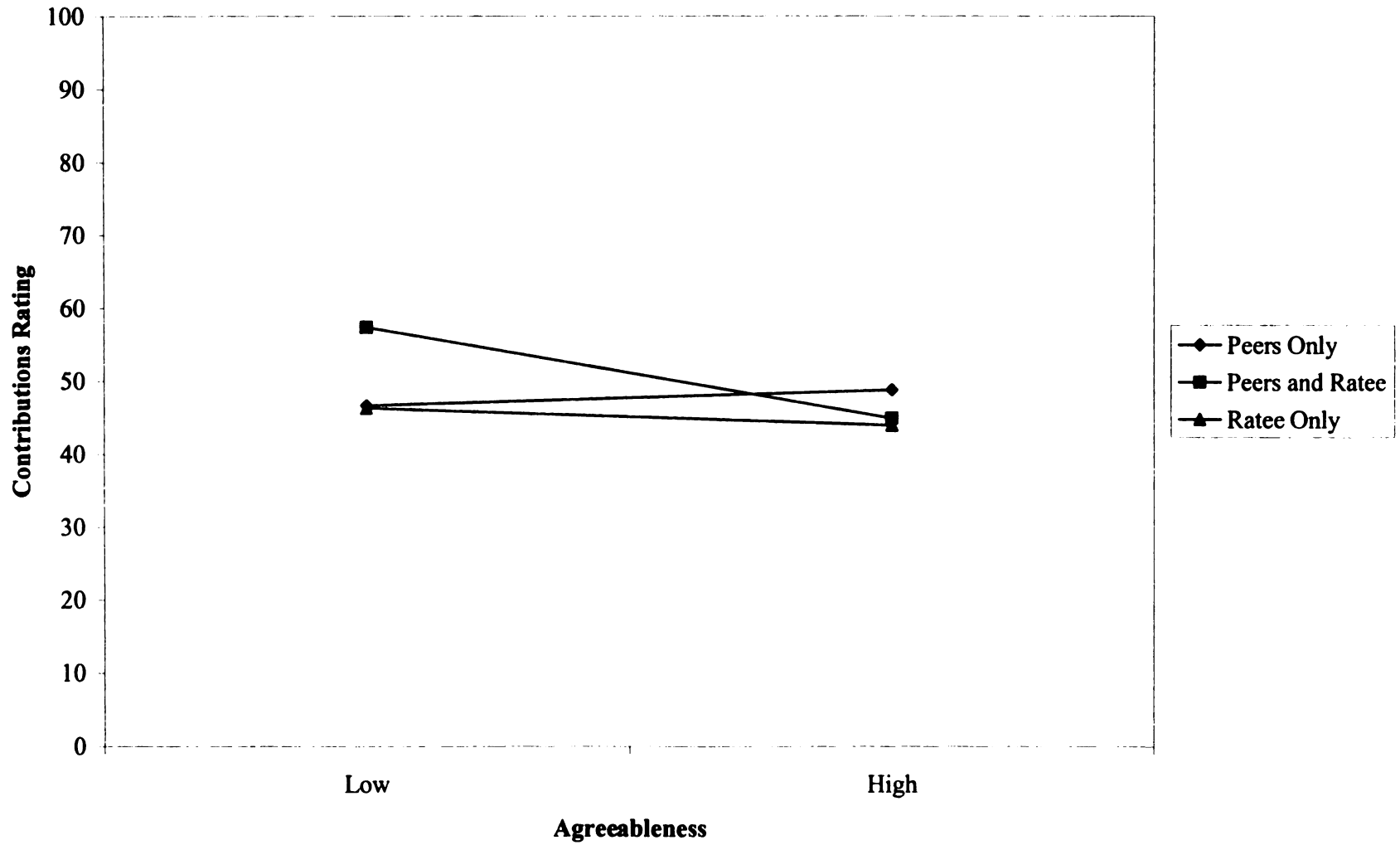


Figure 11. Hypothesis 3a (Results): Accountability X Agreeableness Interaction On Contributions Ratings (Grievance Policy Present)



significant main effects on performance ratings provided. Similar to Cooperation and Contributions and counter to expectations, accountability and raters' levels of agreeableness did not interact significantly (F (2,99) = 0.60, p > .05) in their effects on ratings of Overall Performance. These results are displayed in Table 12 and Figure 12.

Grievance policy absent (Hypothesis 3b). Counter to expectations, accountability did not have a significant main effect (F (2,24) = 0.22, p > .05) on ratings of Cooperation, nor was there a significant main effect (F (1,98) = 0.26, p > .05) of raters' levels of agreeableness or interaction (F (2,96) = 1.28, p > .05) between these two predictors on ratings of Cooperation. These results are displayed in Table 10 and Figure 13.

With regard to ratings of Contributions, accountability did not have a significant main effect (F (2,24) = 0.14, p > .05) on performance ratings; this finding is counter to expectations. Also with regard to Cooperation, ratee's levels of agreeableness did not have a significant main effect (F (1,98) = 0.10, p > .05) on performance ratings, but these two predictors did interact significantly (F (2,96) = 4.28, p < .05) in their effects. These results are displayed in Table 11 and Figure 14.

Finally, counter to expectations, accountability did not have a significant main effect (F (2,24) = 0.32, p > .05) on ratings of Overall Performance, nor was there a significant main effect (F (1,98) = 0.11, p > .05) of raters' levels of agreeableness or interaction (F (2,96) = 1.60, p > .05) between these two predictors on ratings of Overall Performance. These results are displayed in Table 12 and Figure 15. Thus, Hypothesis 3 was not supported.

Figure 12. Hypothesis 3a (Results): Accountability X Agreeableness Interaction On Overall Performance Ratings (Grievance Policy Present)

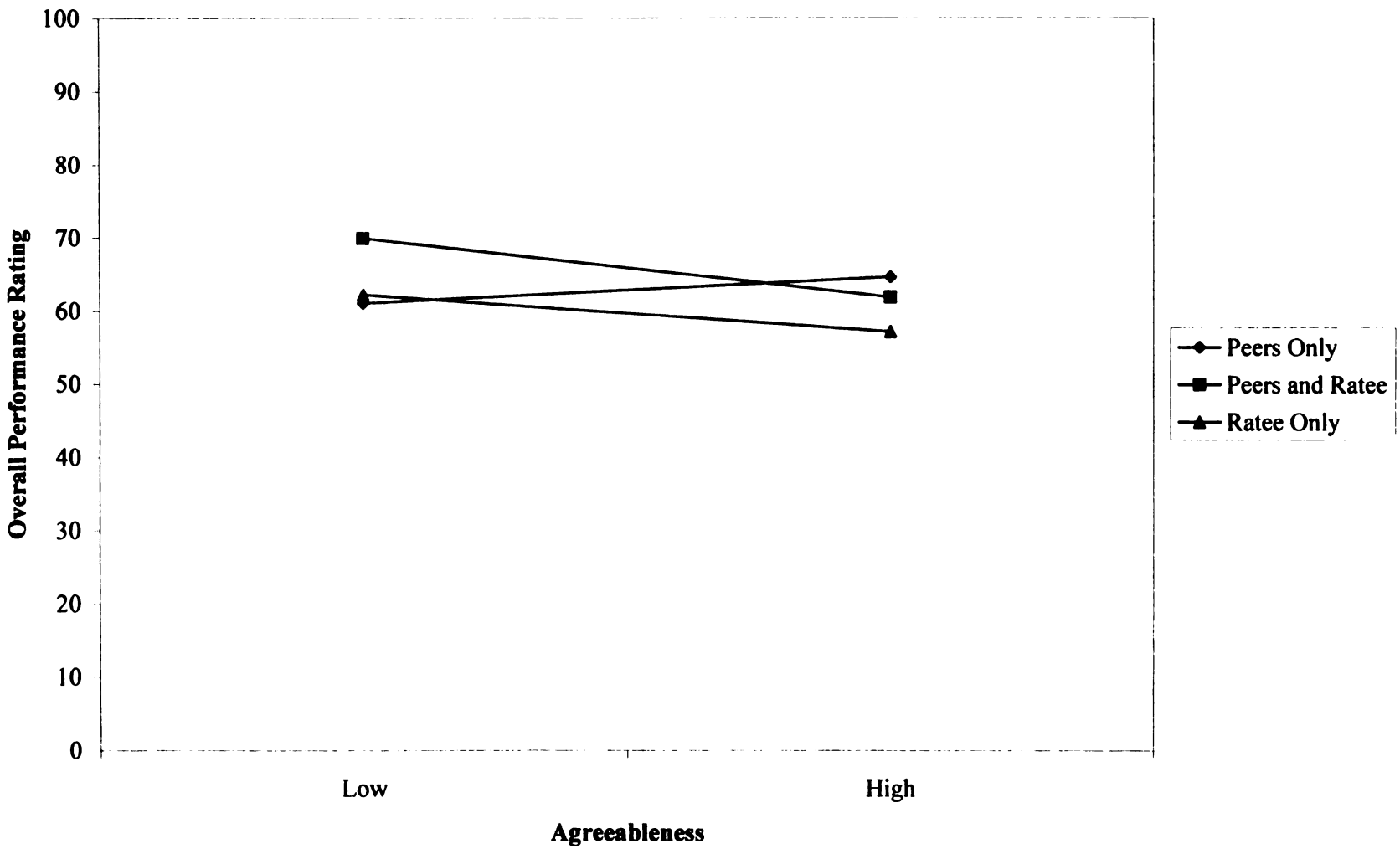


Figure 13. Hypothesis 3b (Results): Accountability X Agreeableness Interaction On Cooperation Ratings (Grievance Policy Absent)

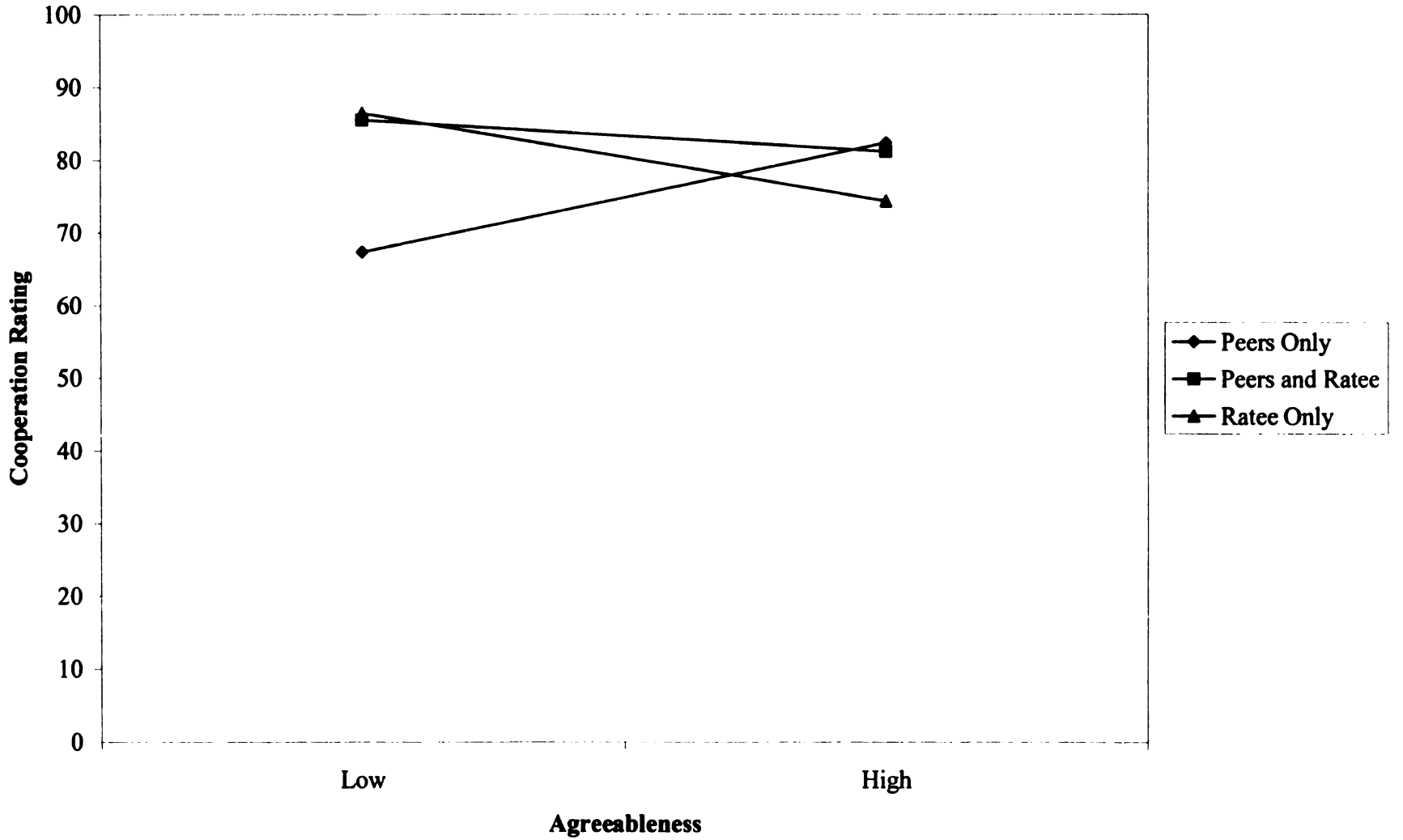


Figure 14. Hypothesis 3b (Results): Accountability X Agreeableness Interaction On Contributions Ratings (Grievance Policy Absent)

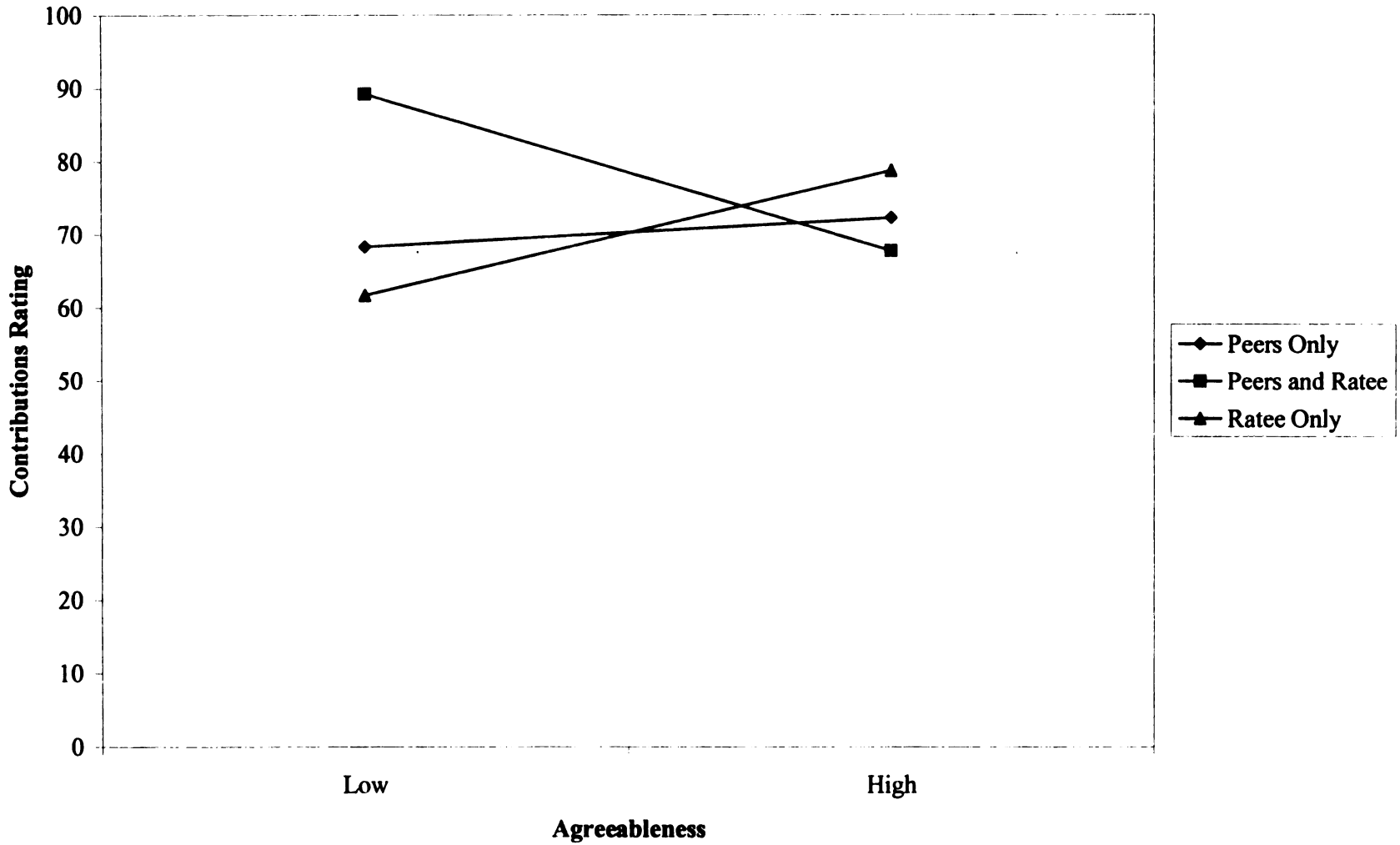
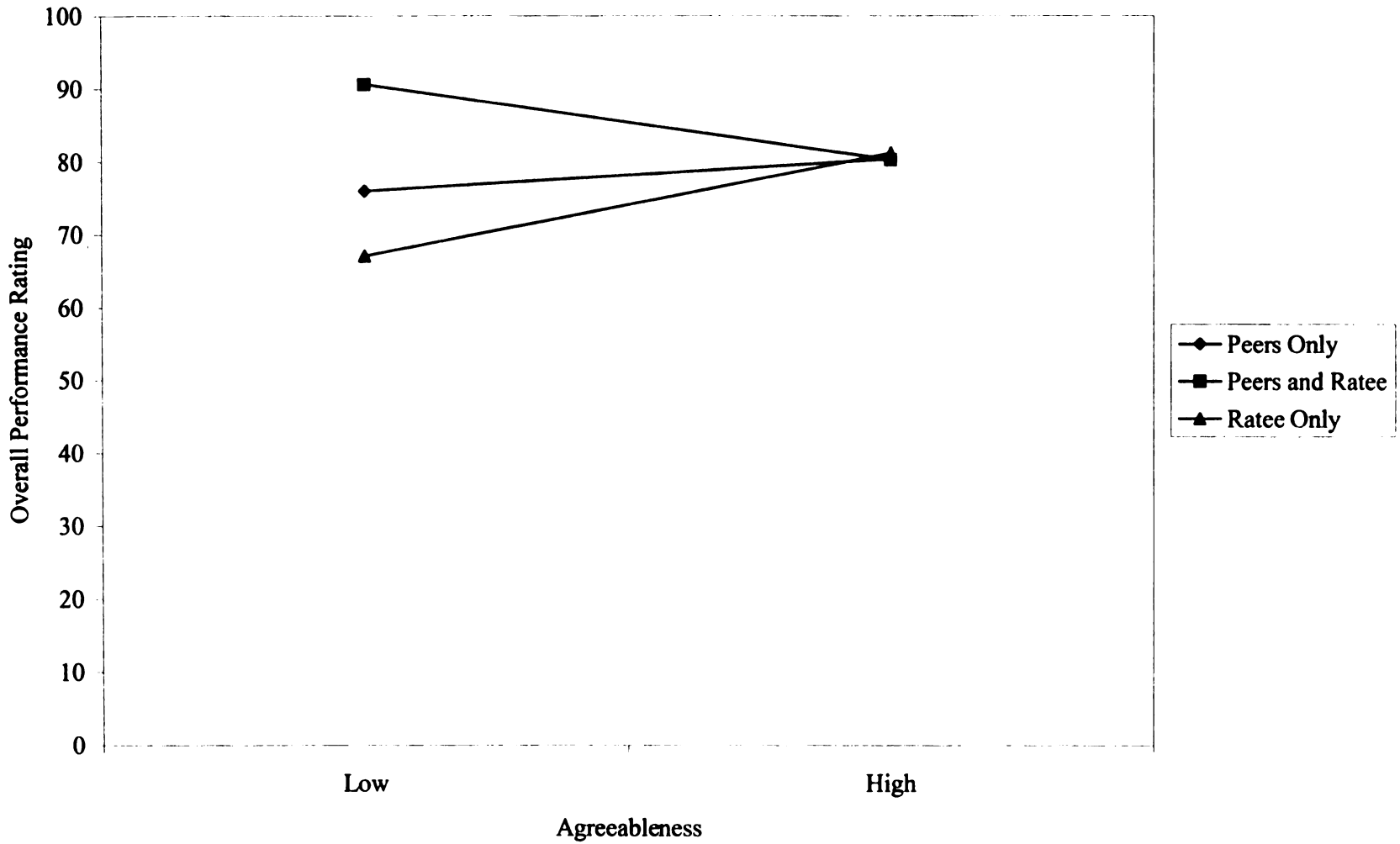


Figure 15. Hypothesis 3b (Results): Accountability X Agreeableness Interaction On Overall Performance Ratings (Grievance Policy Absent)



Hypothesis 4: Accountability, Empathy, and Performance Ratings

Hypothesis 4 predicted that accountability and raters' levels of empathy would interact in their effects on performance ratings provided in a manner consistent with Figure 5. Counter to Hypothesis 4b, accountability did not have a significant main effect (F (2,52) = 0.08, p > .05) on ratings of Cooperation, nor was there a significant main effect (F (1, 200) = 1.64, p > .05) of raters' levels of empathy. Counter to Hypothesis 4a, these two predictors also did not interact significantly (F (2,198) = 0.17, p > .05) in their effects on ratings of Cooperation. These results are displayed in Table 13 and Figure 16.

Also counter to Hypothesis 4b, accountability did not have a significant main effect (F (2,52) = 0.40, p > .05) on ratings of Contributions, nor was there a significant main effect (F (1,200) = 1.62, p > .05) of raters' levels of empathy. Similar to ratings of Cooperation and counter to Hypothesis 4a, accountability and raters' levels of empathy did not interact significantly (F (2,198) = 1.43, p > .05) in their effects on ratings of Contributions. These results are displayed in Table 14 and Figure 17.

Finally, similar to ratings of Cooperation and Contributions and counter to Hypothesis 4a, accountability did not have a significant main effect (F (2,52) = 0.48, p > .05) on ratings of Overall Performance, nor was there a significant main effect (F (1,200) = 0.09, p > .05) of raters' levels of empathy. Counter to Hypothesis 4a, accountability and raters' levels of empathy did not interact significantly (F (2,198) = 1.14, p > .05) in their effects on ratings of Overall Performance. These results are displayed in Table 15 and Figure 18. Hypothesis 4 was not supported.

Table 13

Hypothesis 4: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Empathy Effects On Ratings of Cooperation.

	Accountability			Empathy				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	76.78	78.10	80.20			0.08	2,52	.00
Empathy				68.94	74.68	1.64	1,200	.01
Accountability X Empathy						0.17	2,198	.00
Low Empathy	69.66	74.02	74.95					
High Empathy	78.01	76.03	79.34					

Table 14

Hypothesis 4: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Empathy Effects On Ratings of Contributions.

	Accountability			Empathy				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	55.04	54.50	60.59			0.40	2,52	.01
Empathy				53.12	48.95	1.62	1,200	.01
Accountability X Empathy						1.43	2,198	.01
Low Empathy	55.10	58.04	69.03					
High Empathy	61.34	56.08	58.33					

Table 15

Hypothesis 4: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Empathy Effects On Ratings of Overall Performance.

Effect	Accountability			Empathy		F	df	R^2
	Peers	Ratee	Both	Low	High			
Accountability	69.68	66.68	74.66			0.48	2,52	.01
Empathy				62.70	61.70	0.09	1,200	.00
Accountability X Empathy						1.14	2,198	.01
Low Empathy	68.32	71.24	76.51					
High Empathy	72.04	62.66	73.38					

Figure 16. Hypothesis 4 (Results): Accountability X Empathy Interaction On Cooperation Ratings

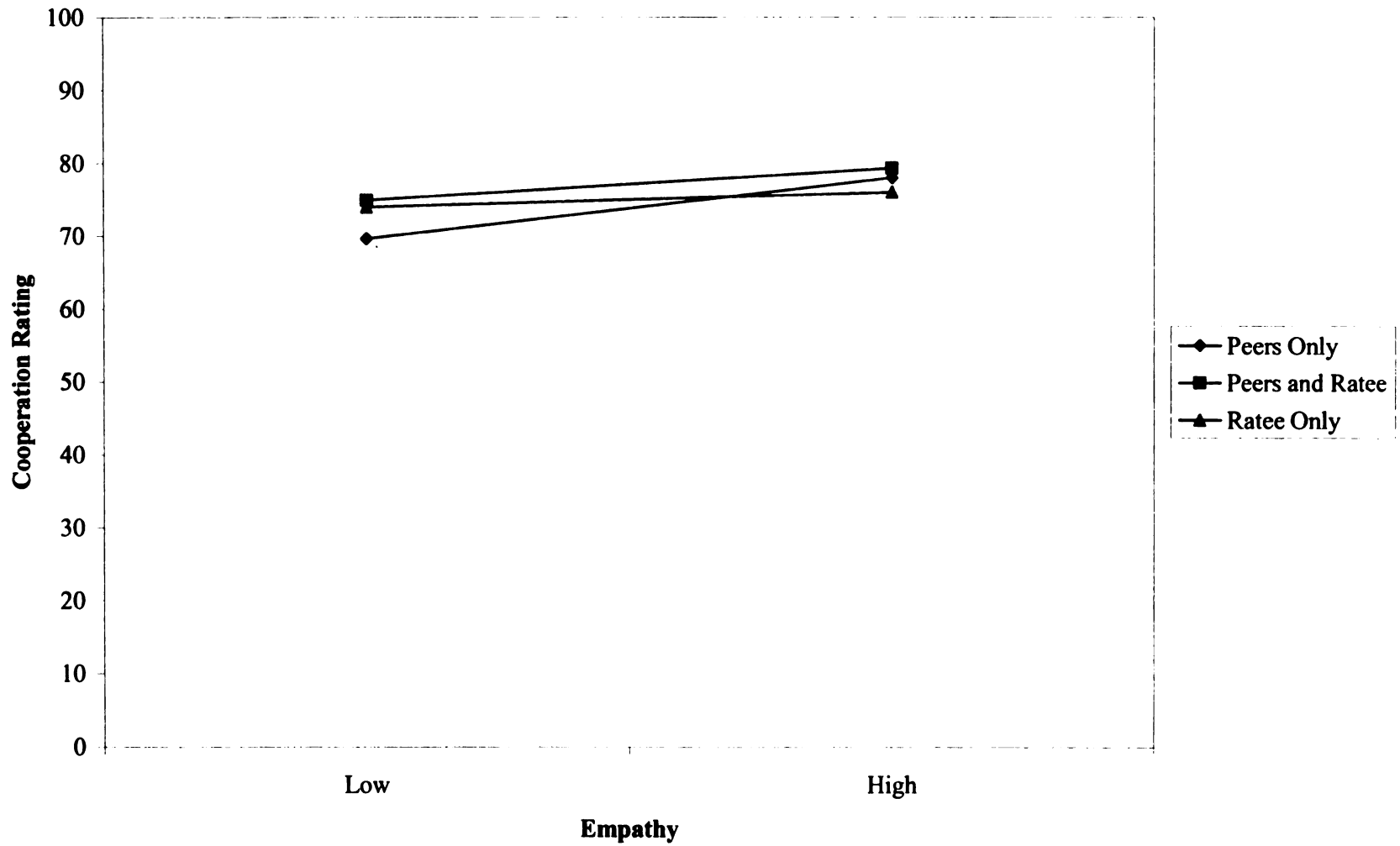


Figure 17. Hypothesis 4 (Results): Accountability X Empathy Interaction On Contributions Ratings

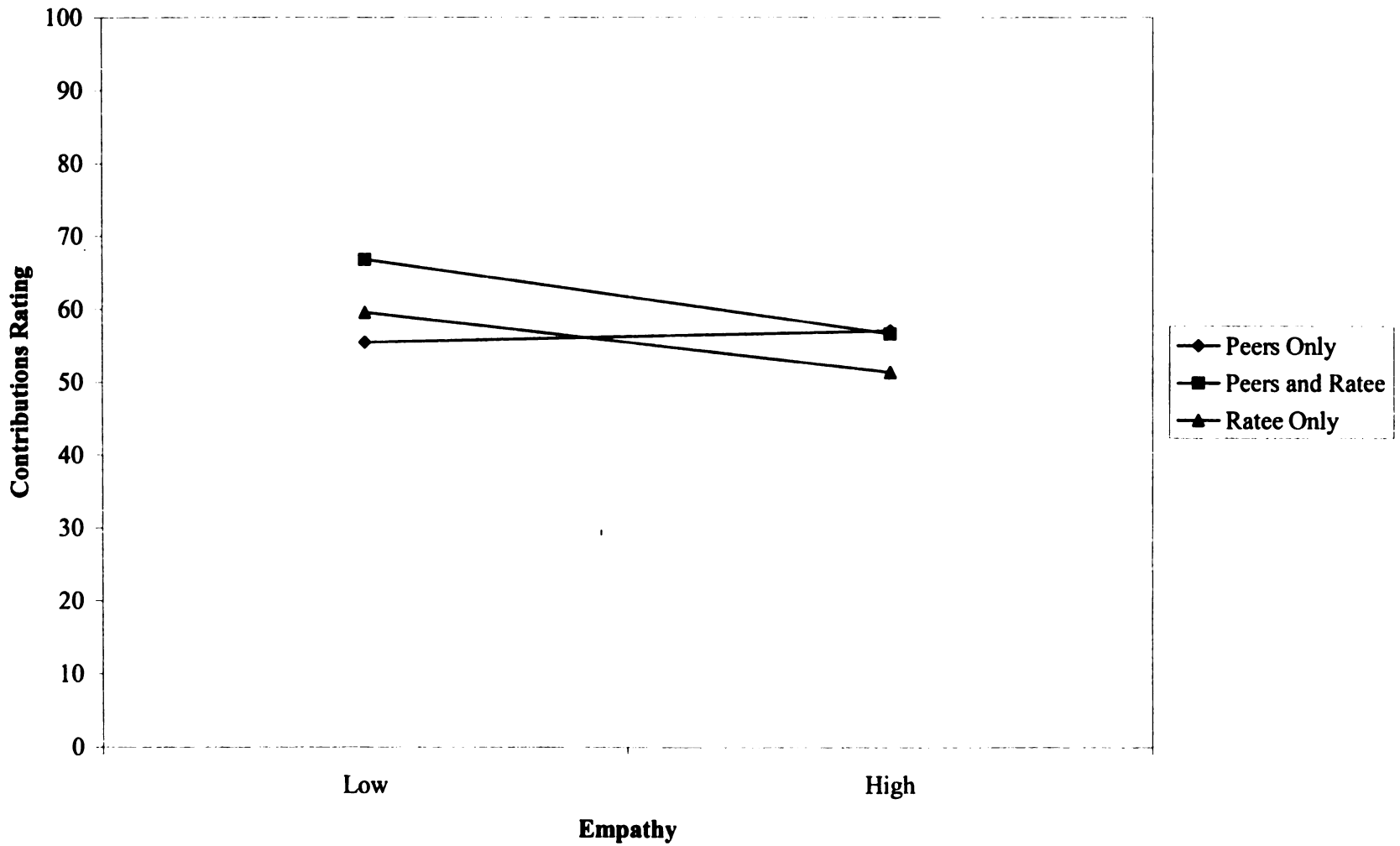
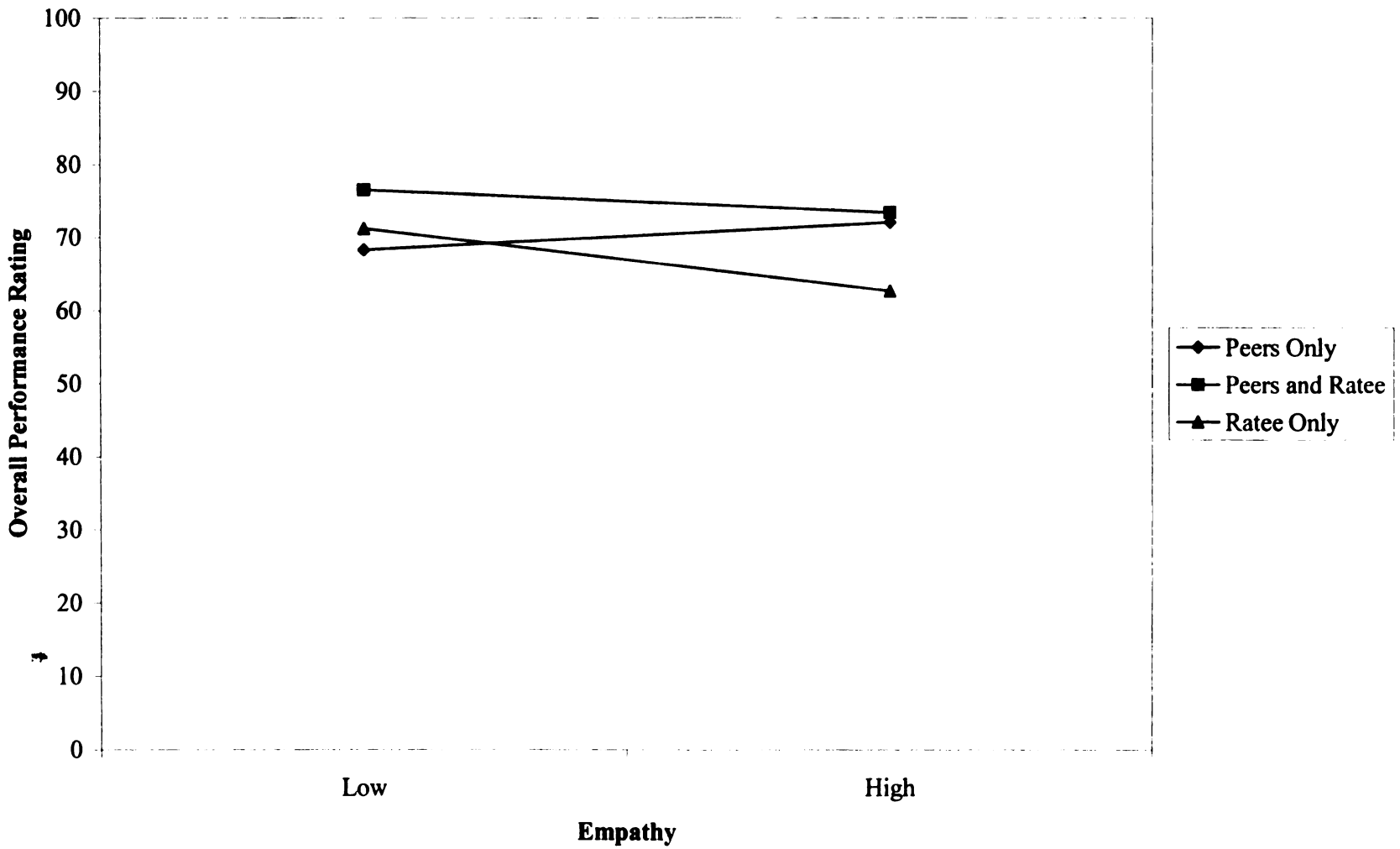


Figure 18. Hypothesis 4 (Results): Accountability X Empathy Interaction On Overall Performance Ratings



Hypothesis 5: Accountability, Conscientiousness, and Performance Ratings

Hypothesis 5 predicted that accountability and raters' levels of conscientiousness would interact in their effects on performance ratings provided in a manner consistent with Figure 6. With regard to ratings of Cooperation, neither accountability (F (2,52) = 0.13, $p > .05$) nor raters' levels of conscientiousness (F (1, 200) = 0.69, $p > .05$) had a significant main effect on performance ratings. Counter to Hypotheses 5a and 5b, these two predictors also did not interact significantly (F (2,198) = 0.30, $p > .05$) in their effects on ratings of Cooperation. These results are displayed in Table 16 and Figure 19.

Similarly, with regard to ratings of Contributions, neither accountability (F (2,52) = 0.39, $p > .05$) nor raters' levels of conscientiousness (F (1,200) = 0.72, $p > .05$) had a significant main effect on performance ratings. Counter to Hypotheses 5a and 5b, these two predictors also did not interact significantly (F (2,198) = 2.13, $p > .05$) in their effects on ratings of Cooperation. These results are displayed in Table 17 and Figure 20.

Finally, consistent with ratings of Cooperation and Contributions, neither accountability (F (2,52) = 0.49, $p > .05$) nor raters' levels of conscientiousness (F (1,200) = 0.42, $p > .05$) had a significant main effect on ratings of Overall Performance. Counter to Hypotheses 5a and 5b, these two predictors also did not interact significantly (F (2,198) = 2.27, $p > .05$) in their effects on ratings of Overall Performance. These results are displayed in Table 18 and Figure 21. Hypothesis 5 was not supported.

Hypothesis 6: Accountability, Self-Monitoring, and Performance Ratings

Hypothesis 6 predicted that accountability and raters' levels of self-monitoring would interact in their effects on performance ratings provided in a manner consistent with Figure 7. Neither accountability (F (2,52) = 0.14, $p > .05$) nor raters' levels of self-

Table 16

Hypothesis 5: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Conscientiousness Effects On Ratings of Cooperation.

	Accountability			Conscientiousness				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	70.67	72.63	75.13			0.13	2,52	.00
Conscientiousness				72.07	68.17	0.69	1,200	.00
Accountability X Conscientiousness						0.30	2,198	.00
Low Conscientiousness	72.66	78.28	81.09					
High Conscientiousness	74.33	72.09	74.67					

Table 17

Hypothesis 5: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Conscientiousness Effects On Ratings of Contributions.

	Accountability			Conscientiousness				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	56.07	55.29	61.42			0.39	2,52	.01
Conscientiousness				52.85	49.93	0.72	1,200	.00
Accountability X Conscientiousness						2.13	2,198	.01
Low Conscientiousness	55.10	58.04	69.03					
High Conscientiousness	61.34	56.08	58.33					

Table 18

Hypothesis 5: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Conscientiousness Effects On Ratings of Overall Performance.

	Accountability			Conscientiousness				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	69.04	66.05	74.12			0.49	2,52	.01
Conscientiousness				63.20	60.96	0.42	1,200	.00
Accountability X Conscientiousness						2.27	2,198	.01
Low Conscientiousness	67.27	70.28	80.85					
High Conscientiousness	75.69	65.72	72.13					

Figure 19. Hypothesis 5 (Results): Accountability X Conscientiousness Interaction On Cooperation Ratings

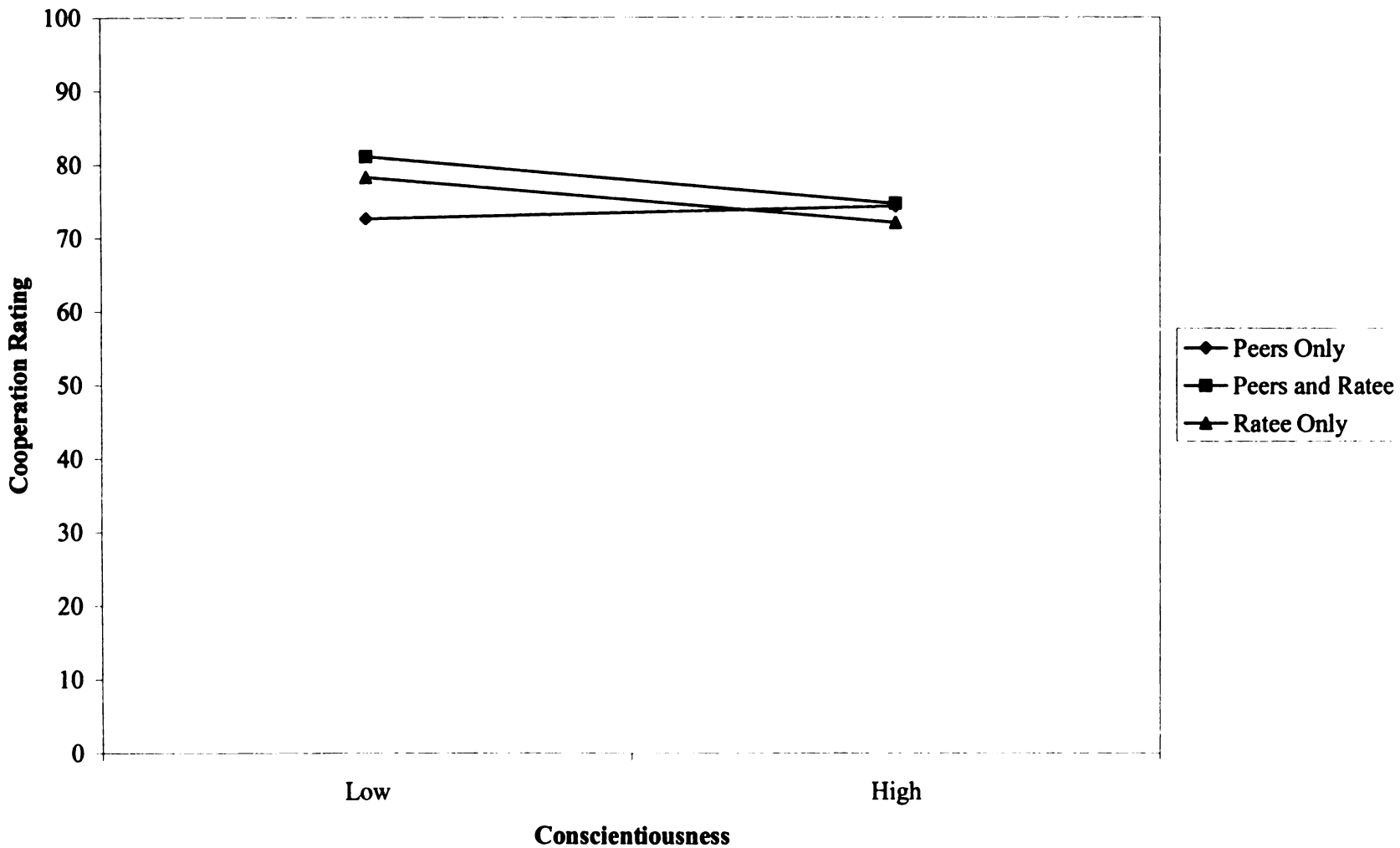


Figure 20. Hypothesis 5 (Results): Accountability X Conscientiousness Interaction On Contributions Ratings

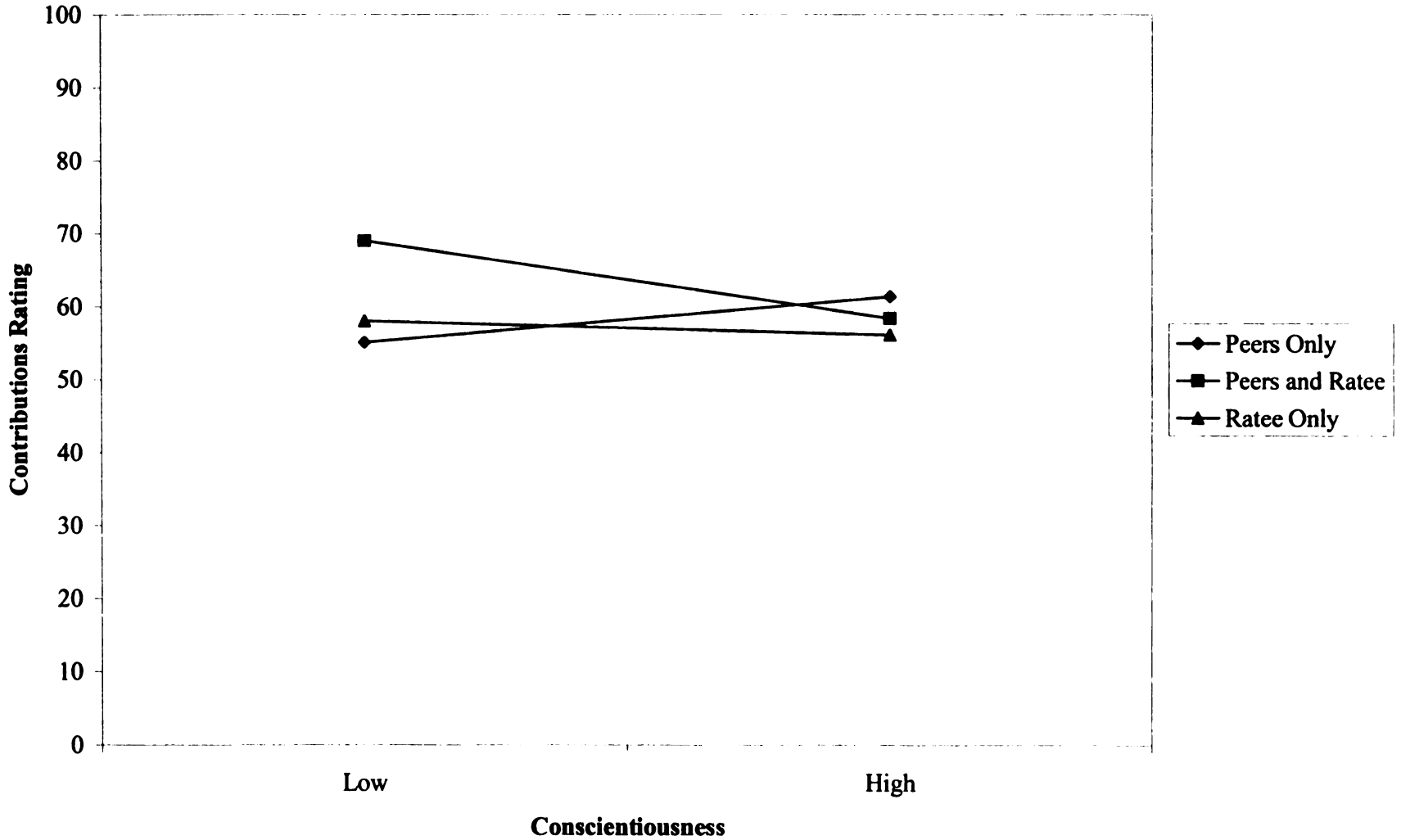
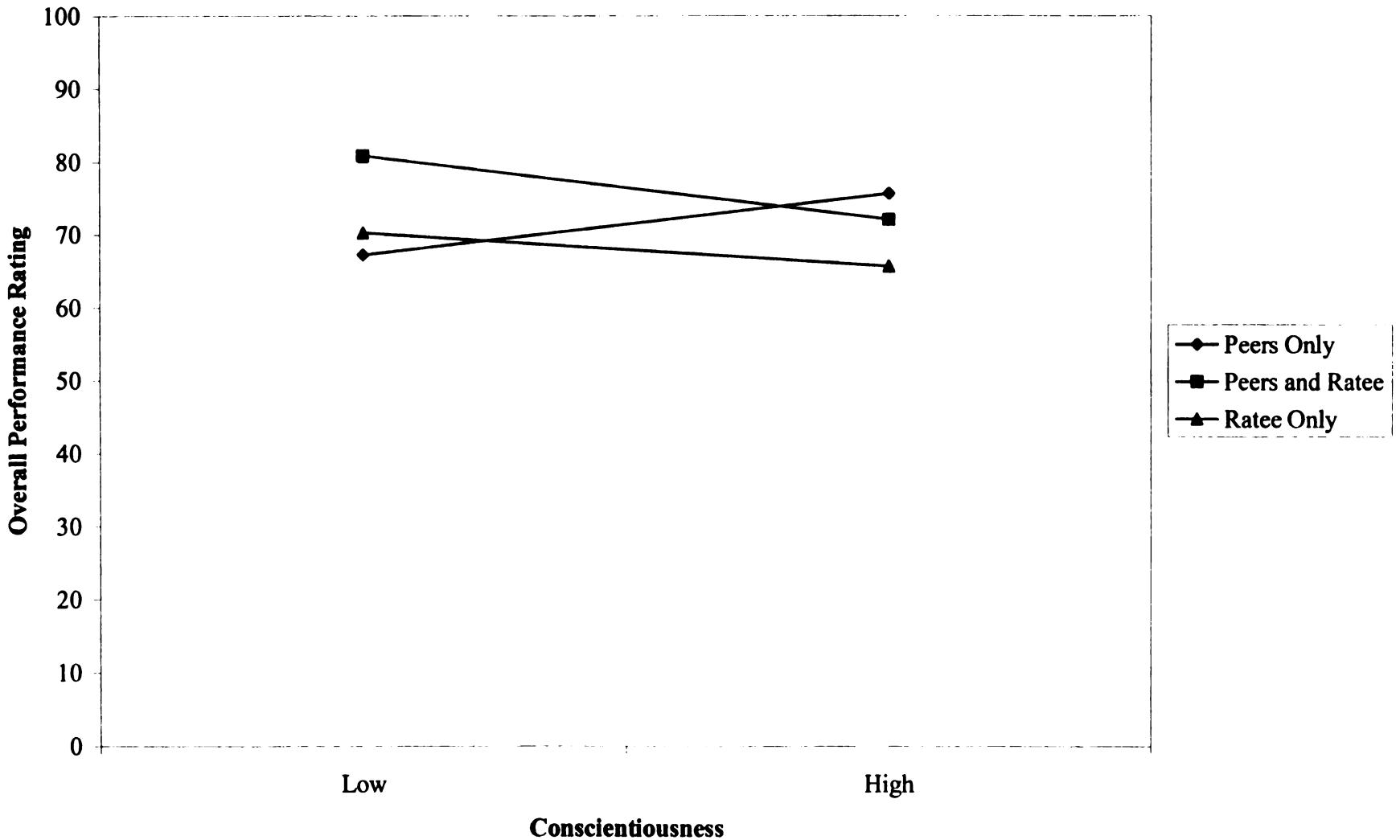


Figure 21. Hypothesis 5 (Results): Accountability X Conscientiousness Interaction On Overall Performance Ratings



monitoring (F (1, 200) = 1.86, p > .05) had a significant main effect on ratings of Cooperation. Counter to Hypotheses 6a and 6b, these two predictors also did not interact significantly (F (2,198) = 1.58, p > .05) in their effects on ratings of Cooperation. These results are displayed in Table 19 and Figure 22.

Similarly, with regard to ratings of Contributions, neither accountability (F (2,52) = 0.37, p > .05) nor raters' levels of self-monitoring (F (1,200) = 1.22, p > .05) had a significant main effect on performance ratings. Counter to Hypotheses 6a and 6b, these two predictors also did not interact significantly (F (2,198) = 0.70, p > .05) in their effects on ratings of Cooperation. These results are displayed in Table 20 and Figure 23.

Finally, consistent with ratings of Cooperation and Contributions, accountability (F (2,52) = 0.46, p > .05) did not have a significant main effect on ratings of Overall Performance, but the effect (F (1,200) = 3.56, p = .06) for raters' levels of self-monitoring was marginal, such that higher levels of self-monitoring are related to higher performance ratings. Counter to Hypotheses 6a and 6b, accountability and raters' levels of self-monitoring did not interact significantly (F (2,198) = 1.68, p > .05) in their effects on ratings of Overall Performance. These results are displayed in Table 21 and Figure 24. Hypothesis 6 was not supported.

Hypothesis 7: Accountability, Grievance Policy, Ego Concern, and Performance Ratings

Hypothesis 7 predicted that accountability, presence/absence of a grievance policy, and raters' levels of ego concern would interact in their effects on performance ratings provided in a manner consistent with Figures 8 (Hypothesis 7a) and 9 (Hypothesis 7b). As indicated in Tables 22-24, this three-way interaction was not observed for any of the outcome variables (Cooperation: F (2,195) = 0.39, p > .05; Contributions: F (2,195) =

Table 19

Hypothesis 6: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Self-Monitoring Effects On Ratings of Cooperation.

	Accountability			Self-Monitoring				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	75.31	77.80	79.85			0.14	2,52	.00
Self-Monitoring				66.68	73.27	1.86	1,200	.01
Accountability X Self-Monitoring						1.58	2,198	.01
Low Self-Monitoring	66.88	76.94	68.98					
High Self-Monitoring	75.91	70.70	83.17					

Table 20

Hypothesis 6: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Self-Monitoring Effects On Ratings of Contributions.

	Accountability			Self-Monitoring				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	59.14	58.66	64.48			0.37	2,52	.01
Self-Monitoring				49.39	53.32	1.22	1,200	.01
Accountability X Self-Monitoring						0.70	2,198	.00
Low Self-Monitoring	55.04	57.20	57.98					
High Self-Monitoring	58.77	55.43	66.43					

Table 21

Hypothesis 6: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Self-Monitoring Effects On Ratings of Overall Performance.

	Accountability			Self-Monitoring				
Effect	Peers	Ratee	Both	Low	High	*F*	*df*	R^2
Accountability	72.67	70.29	78.00			0.46	2,52	.01
Self-Monitoring				58.27	64.97	3.56[a]	1,200	.02
Accountability X Self-Monitoring						1.68	2,198	.01
Low Self-Monitoring	64.71	67.65	67.68					
High Self-Monitoring	72.75	64.92	80.56					

[a] $p = .06$

Table 22

Hypothesis 7: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Ego Concern Effects On Ratings of Cooperation.

Effect	Accountability			Ego Concern		F	df	R^2
	Peers	Ratee	Both	Low	High			
Accountability X Policy X Ego Concern						0.39	2,195	.00
Accountability								
Grievance Policy Present	70.62	69.71	69.93			0.00	2,24	.00
Grievance Policy Absent	77.17	81.74	84.26			0.20	2,24	.01
Ego Concern								
Grievance Policy Present				70.60	70.40	0.00	1,101	.00
Grievance Policy Absent				69.87	74.65	0.54	1,98	.01
Accountability X Ego Concern								
Grievance Policy Present						0.78	2,99	.01
Low Ego Concern	71.11	78.02	69.52					
High Ego Concern	75.44	64.52	75.21					
Grievance Policy Absent						3.22*	2,96	.03
Low Ego Concern	76.71	95.95	74.43					
High Ego Concern	83.59	73.69	94.33					

* $p < .05$

Table 23

Hypothesis 7: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Ego Concern Effects On Ratings of Contributions.

	Accountability			Ego Concern				
Effect	Peers	Ratee	Both	Low	High	F	df	R^2
Accountability X Policy X Ego Concern						1.59	2,195	.01
Accountability								
Grievance Policy Present	46.10	43.62	49.82			0.16	2,24	.01
Grievance Policy Absent	73.50	74.18	78.58			0.12	2,24	.00
Ego Concern								
Grievance Policy Present				45.53	39.89	1.28	1,101	.01
Grievance Policy Absent				67.64	69.10	0.10	1,98	.00
Accountability X Ego Concern								
Grievance Policy Present						0.78	2,99	.01
Low Ego Concern	50.11	43.98	57.79					
High Ego Concern	44.63	46.83	44.27					
Grievance Policy Absent						1.11	2,96	.01
Low Ego Concern	74.58	75.70	71.34					
High Ego Concern	71.19	72.03	81.94					

Table 24

Hypothesis 7: Estimated Least Squared Means of Rater Accountability, Grievance Policy, and Ego Concern Effects On Ratings of Overall Performance.

Effect	Accountability			Ego Concern		F	df	R^2
	Peers	Ratee	Both	Low	High			
Accountability X Policy X Ego Concern						1.44	2,195	.01
Accountability								
Grievance Policy Present	63.33	60.70	66.41			0.10	2,24	.00
Grievance Policy Absent	79.92	76.54	85.59			0.32	2,24	.01
Ego Concern								
Grievance Policy Present				57.99	57.63	0.01	1,101	.00
Grievance Policy Absent				69.73	70.87	0.06	1,98	.00
Accountability X Ego Concern								
Grievance Policy Present						0.08	2,99	.00
Low Ego Concern	64.49	60.00	65.64					
High Ego Concern	61.70	61.40	66.99					
Grievance Policy Absent						3.66*	2,96	.04
Low Ego Concern	82.02	86.64	77.49					
High Ego Concern	81.35	70.61	92.43					

* $p < .05$

Figure 22. Hypothesis 6 (Results): Accountability X Self-Monitoring Interaction On Cooperation Ratings

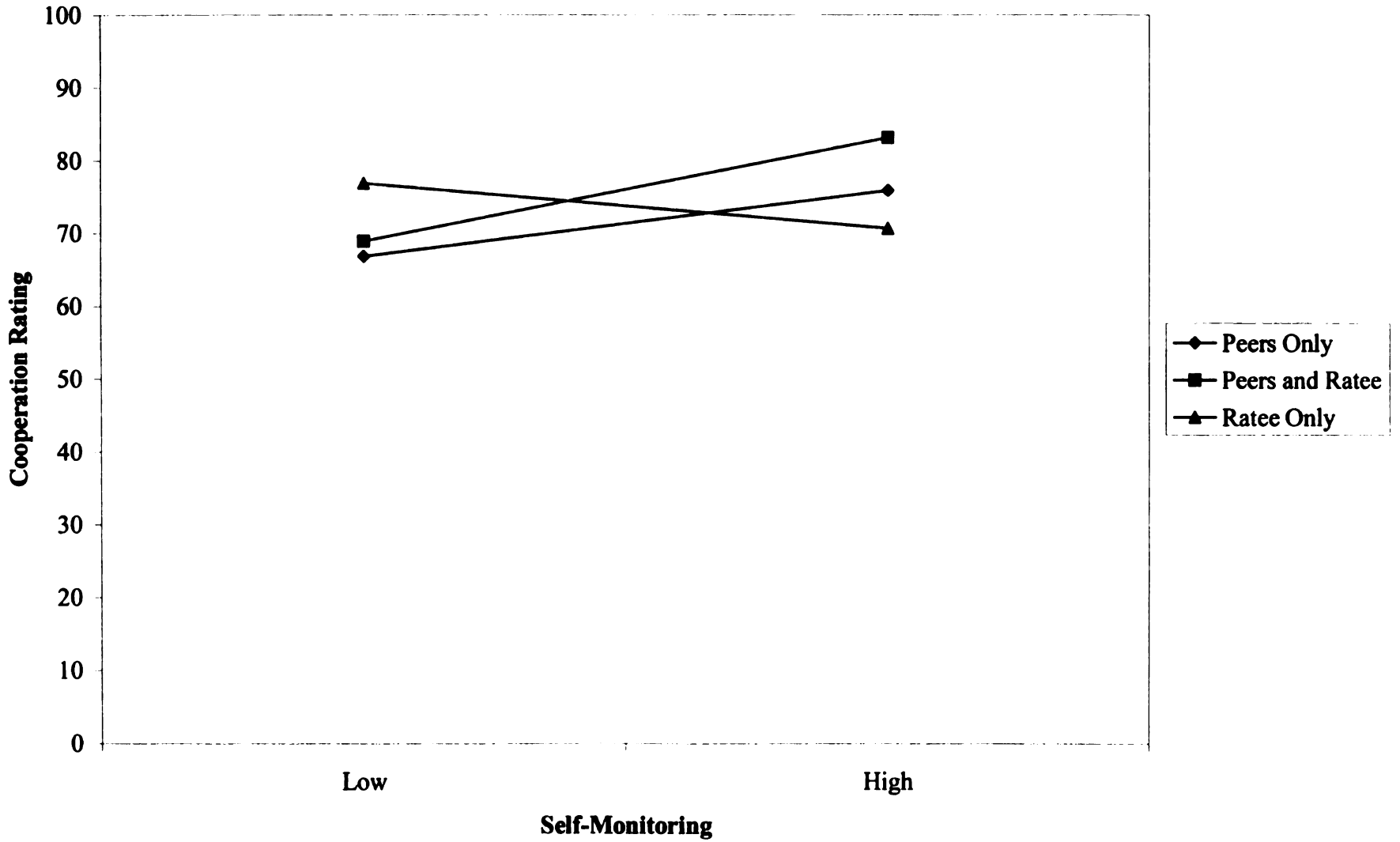


Figure 23. Hypothesis 6 (Results): Accountability X Self-Monitoring Interaction On Contributions Ratings

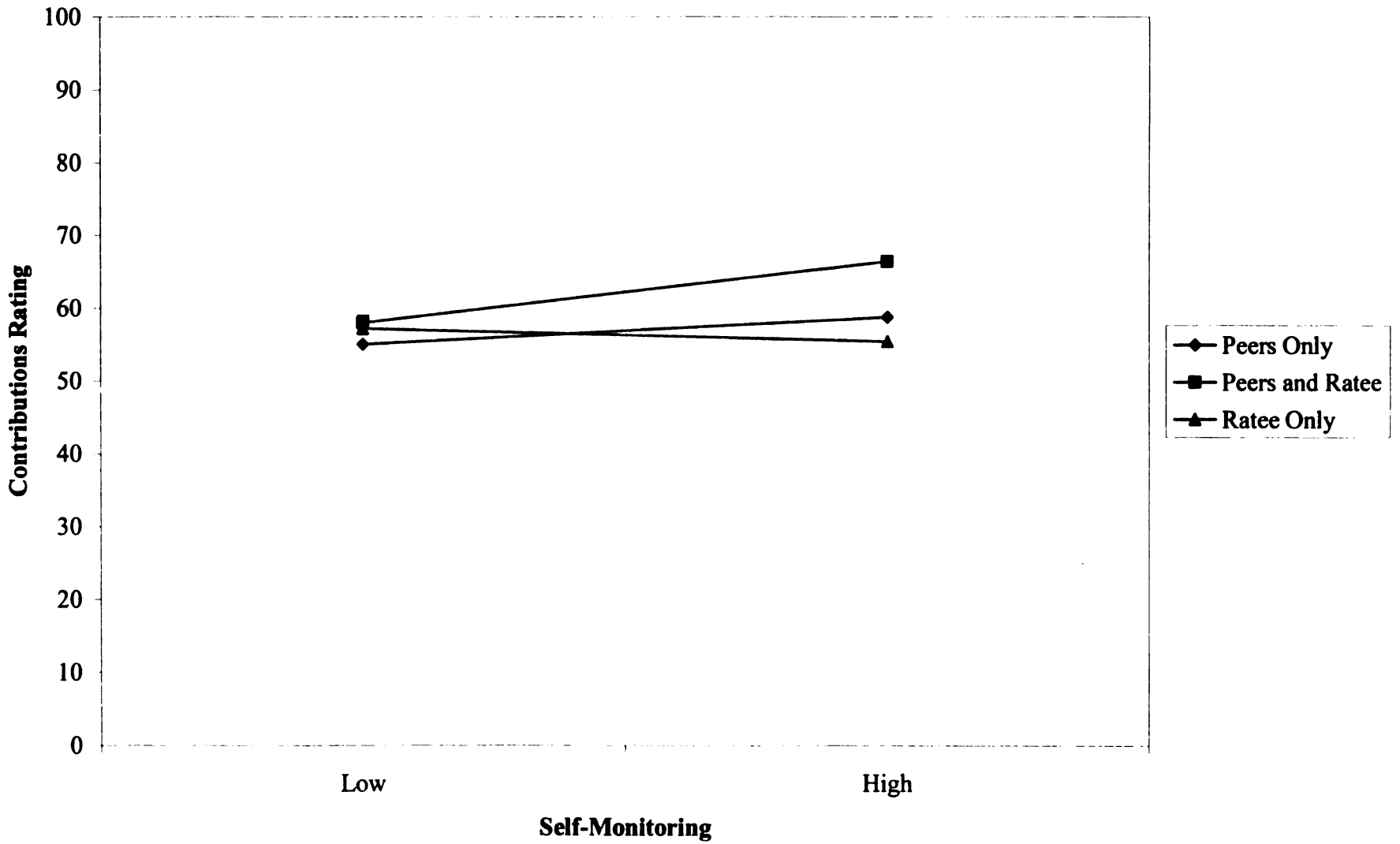
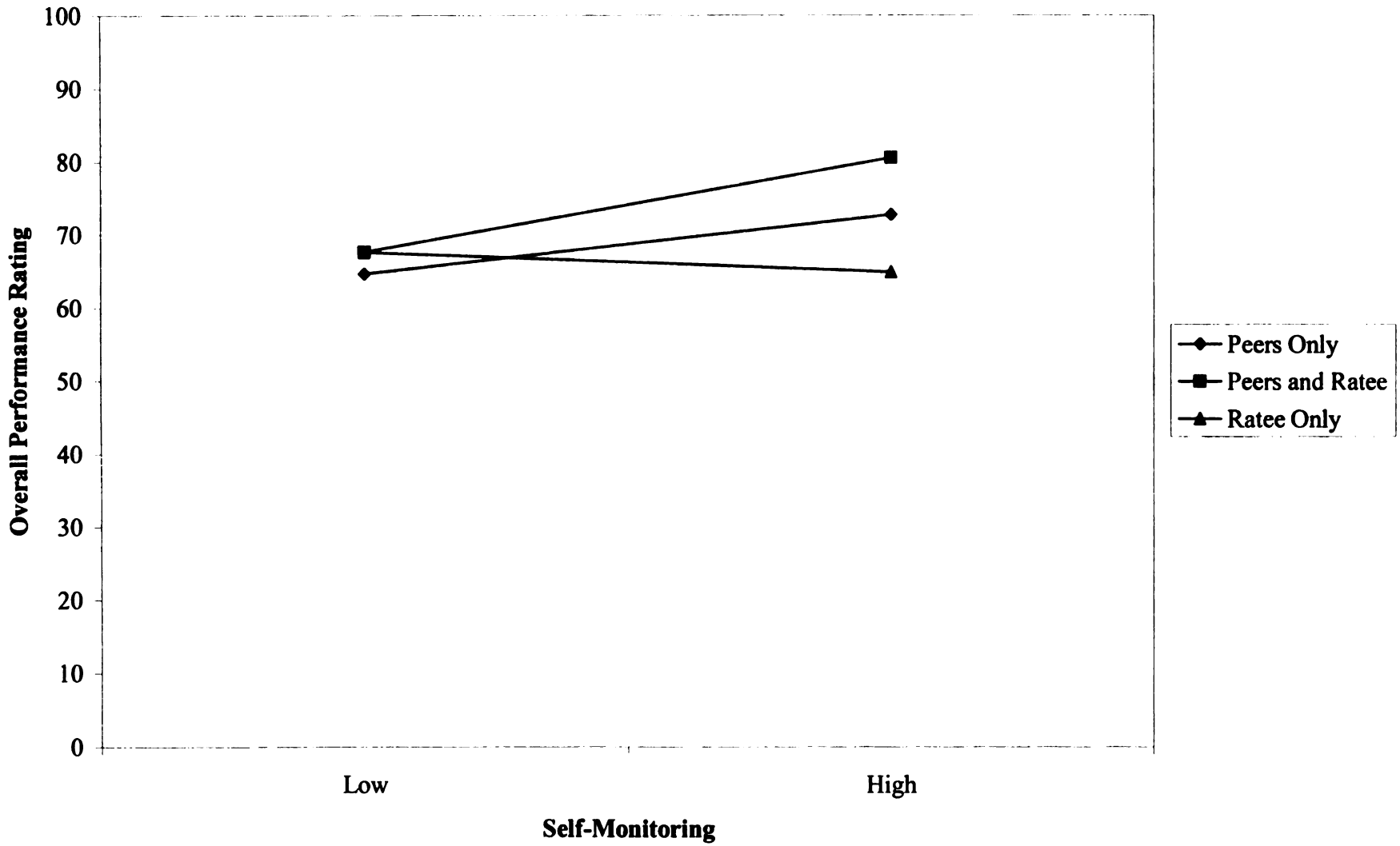


Figure 24. Hypothesis 6 (Results): Accountability X Self-Monitoring Interaction On Overall Performance Ratings



1.59, $p > .05$; Overall Performance: F (2,195) = 1.44, $p > .05$). Regardless, Hypotheses 7a (grievance policy present) and 7b (grievance policy absent) were examined for exploratory purposes.

Grievance policy present (Hypothesis 7a). With regard to ratings of Cooperation, neither accountability (F (2,24) = .0.00, $p > .05$) nor raters' levels of ego concern (F (1,101) = 0.00, $p > .05$) had significant main effects on performance ratings provided. Further, counter to Hypothesis 7a, these two variables did not interact significantly (F (2,99) = 0.78, $p > .05$) in their effects on performance ratings provided. These results are displayed in Table 22 and Figure 25.

Similarly, neither accountability (F (2,24) = 0.16, $p > .05$) nor raters' levels of ego concern (F (1,101) = 1.28, $p > .05$) had significant main effects on ratings of Contributions, nor did these two predictors interact significantly (F (2,99) = 0.78, $p > .05$) in their effects on performance ratings provided. These results are displayed in Table 23 and Figure 26.

Finally, with regard to ratings of Overall Performance, neither accountability (F (2,24) = 0.10, $p > .05$) nor raters' levels of ego concern (F (1,101) = 0.01, $p > .05$) had significant main effects on performance ratings provided. Similar to Cooperation and Contributions and counter to expectations, accountability and raters' levels of ego concern did not interact significantly (F (2,99) = 0.08, $p > .05$) in their effects on ratings of Overall Performance. These results are displayed in Table 24 and Figure 27.

Grievance policy absent (Hypothesis 7b). Neither accountability (F (2,24) = 0.20, $p > .05$) nor raters' levels of ego concern (F (1,98) = 0.54, $p > .05$) had significant main effects on ratings of Cooperation. As expected, a significant interaction (F (2,96) = 3.22,

Figure 25. Hypothesis 7a (Results): Accountability X Ego Concern Interaction On Cooperation Ratings (Grievance Policy Present)

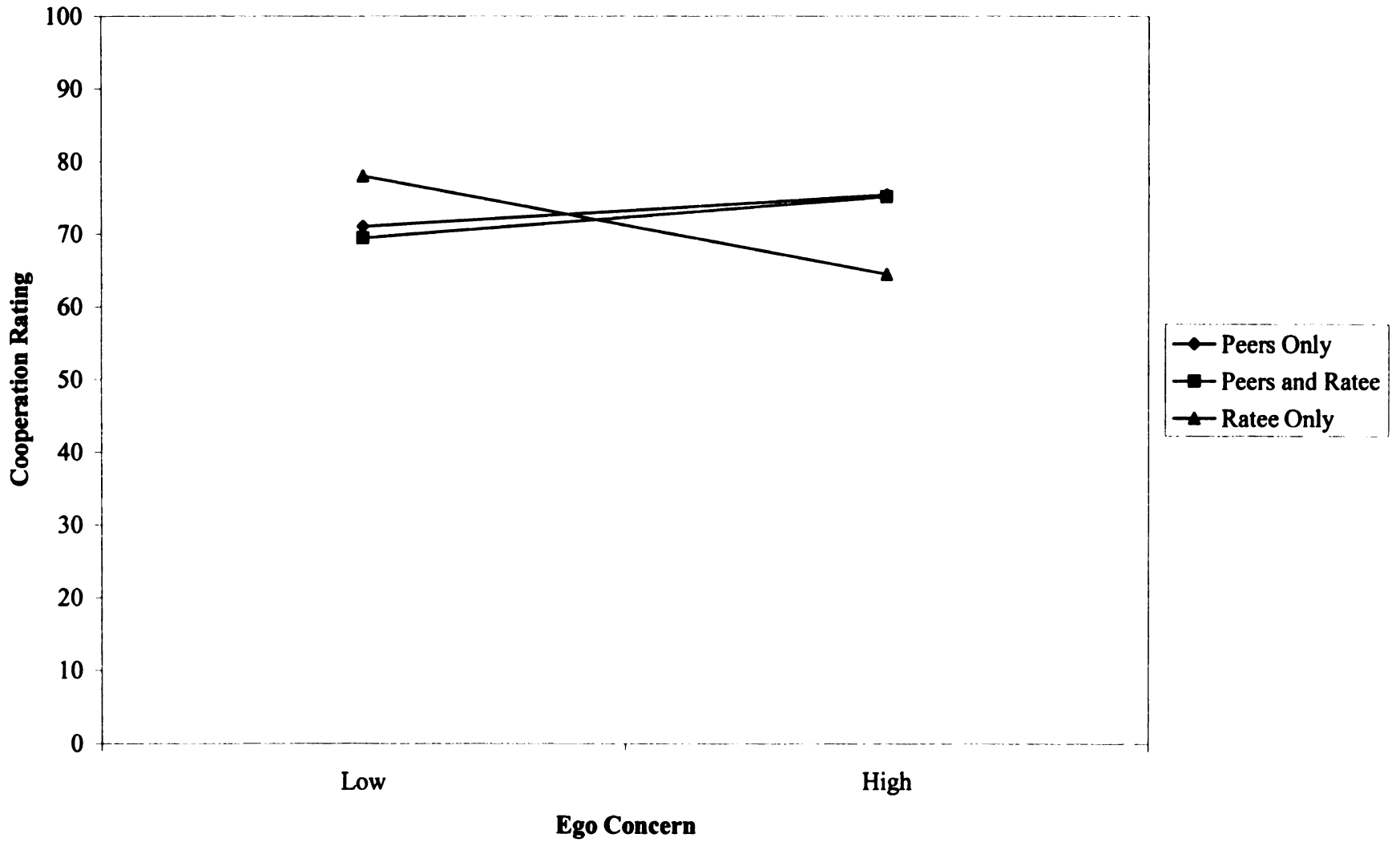


Figure 26. Hypothesis 7a (Results): Accountability X Ego Concern Interaction On Cooperation Ratings (Grievance Policy Present)

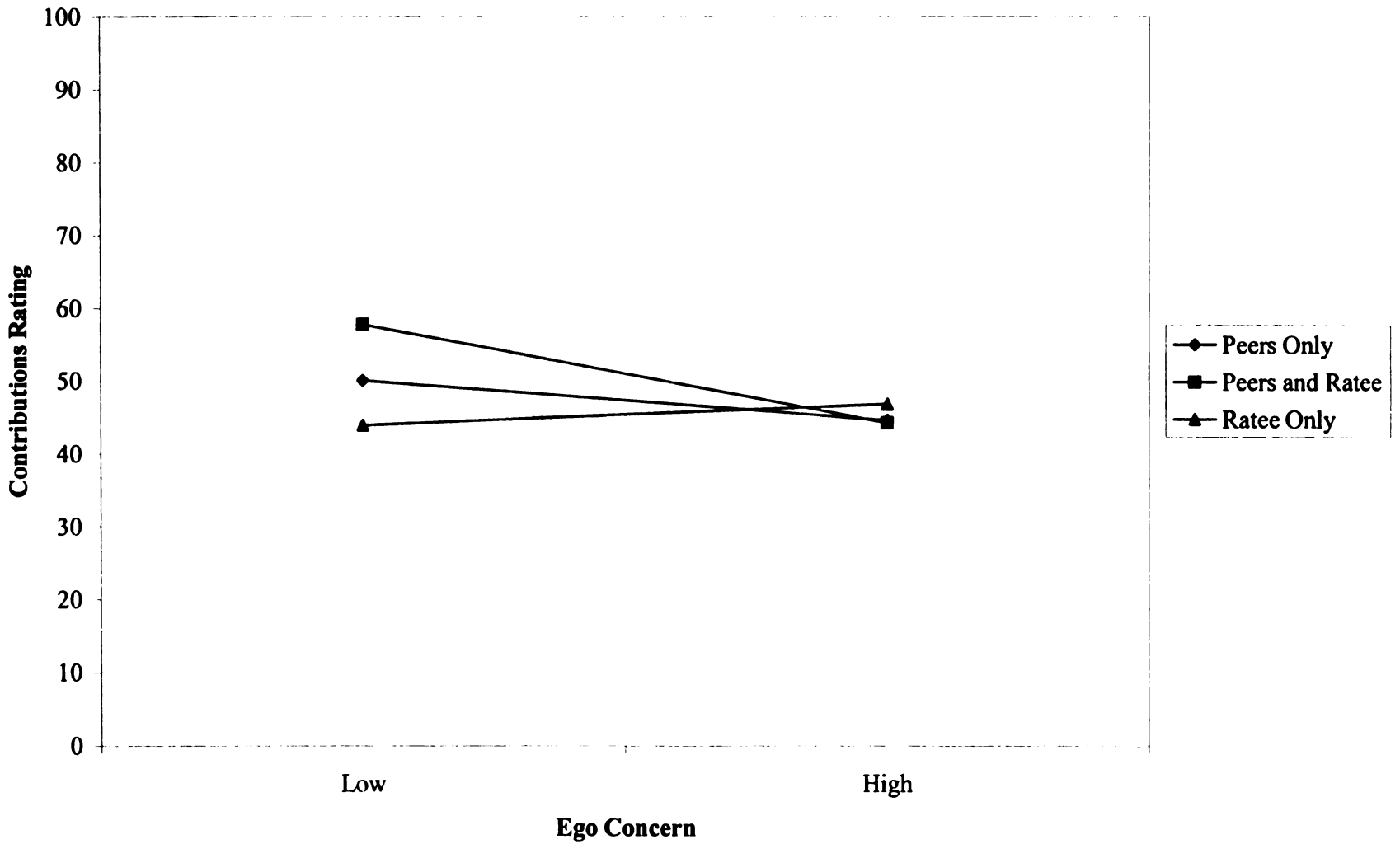
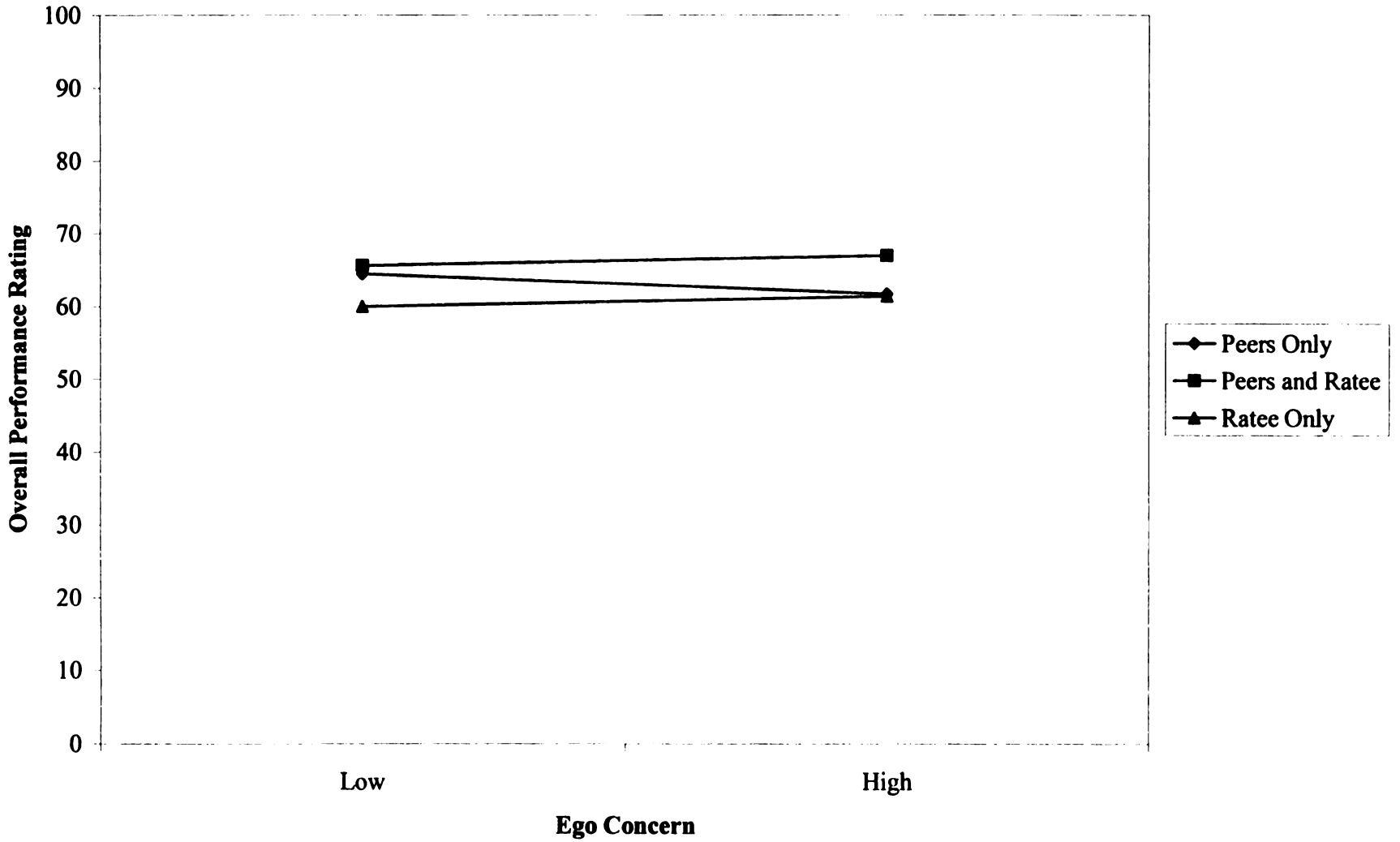


Figure 27. Hypothesis 7a (Results): Accountability X Ego Concern Interaction On Overall Performance Ratings (Grievance Policy Present)



$p < .05$) between accountability and raters' levels of ego concern in their effects on ratings of Cooperation was observed. However, an examination of Figure 28 reveals that the nature of this interaction is not as specified in Hypothesis 7b (see Figure 9). These results are displayed in Table 22.

With regard to ratings of Contributions, neither accountability (F (2,24) = 0.12, p > .05) nor raters' levels of ego concern (F (1,98) = 0.10, p > .05) had significant main effects on ratings provided. Counter to expectations, these two predictors also did not interact significantly (F (2,96) = 1.11, p > .05) in their effects on ratings of Contributions. These results are displayed in Table 23 and Figure 29.

Finally, with regard to ratings of Overall Performance, neither accountability (F (2,24) = 0.32, p > .05) nor raters' levels of ego concern (F (1,98) = 0.06, p > .05) had significant main effects on performance ratings provided. As expected, a significant interaction (F (2,96) = 3.66, p < .05) between accountability and raters' levels of ego concern in their effects on ratings of Overall Performance was observed. An examination of Figure 30 reveals that although the nature of this interaction is not as specified in Hypothesis 7b (see Figure 9), it is consistent with results obtained with regard to ratings of Cooperation in the absence of a grievance policy reported above (see Figure 28). These results are displayed in Table 24. Thus, Hypothesis 7 was not supported.

Figure 28. Hypothesis 7b (Results): Accountability X Ego Concern Interaction On Cooperation Ratings (Grievance Policy Absent)

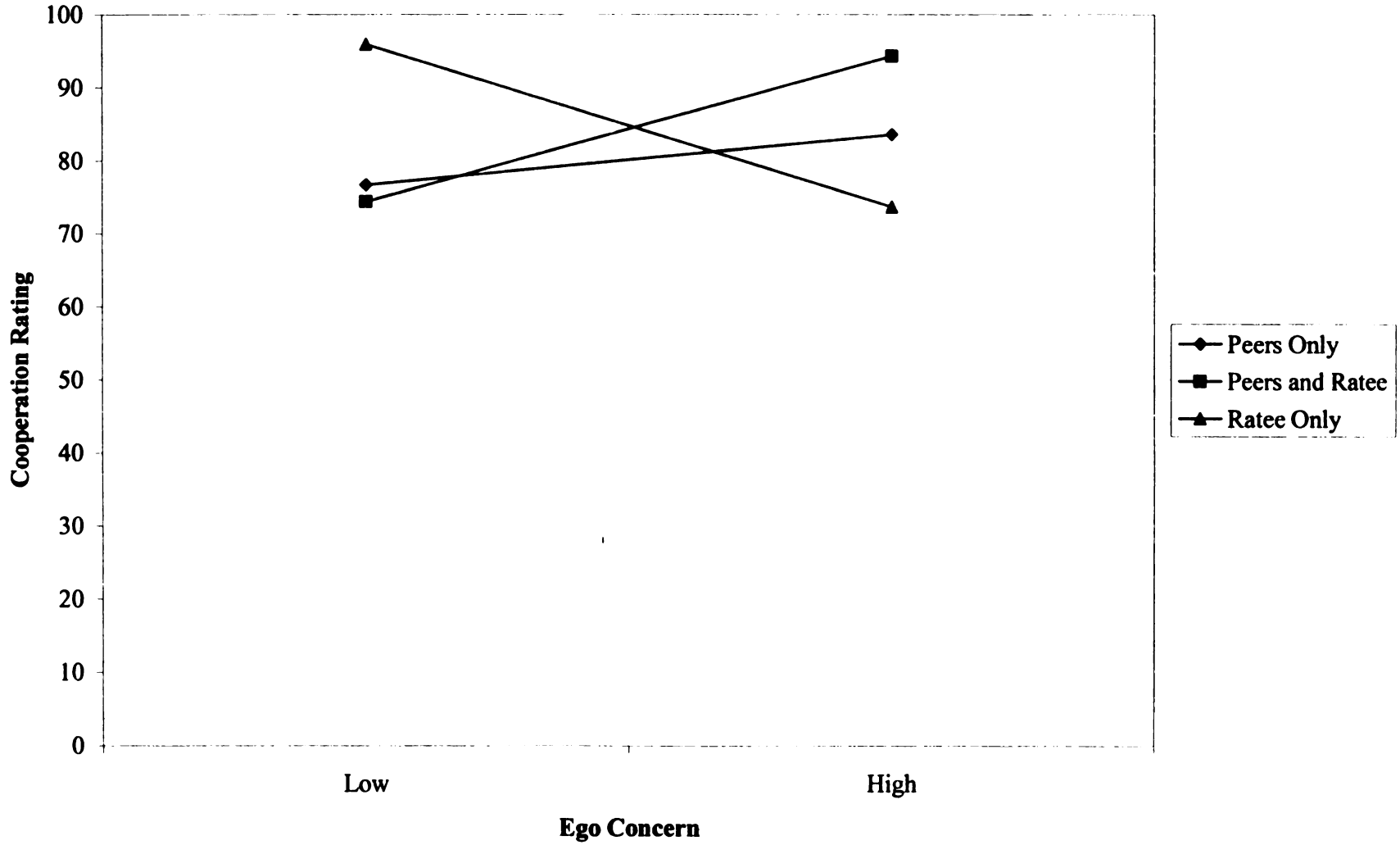


Figure 29. Hypothesis 7b (Results): Accountability X Ego Concern Interaction On Contributions Ratings (Grievance Policy Absent)

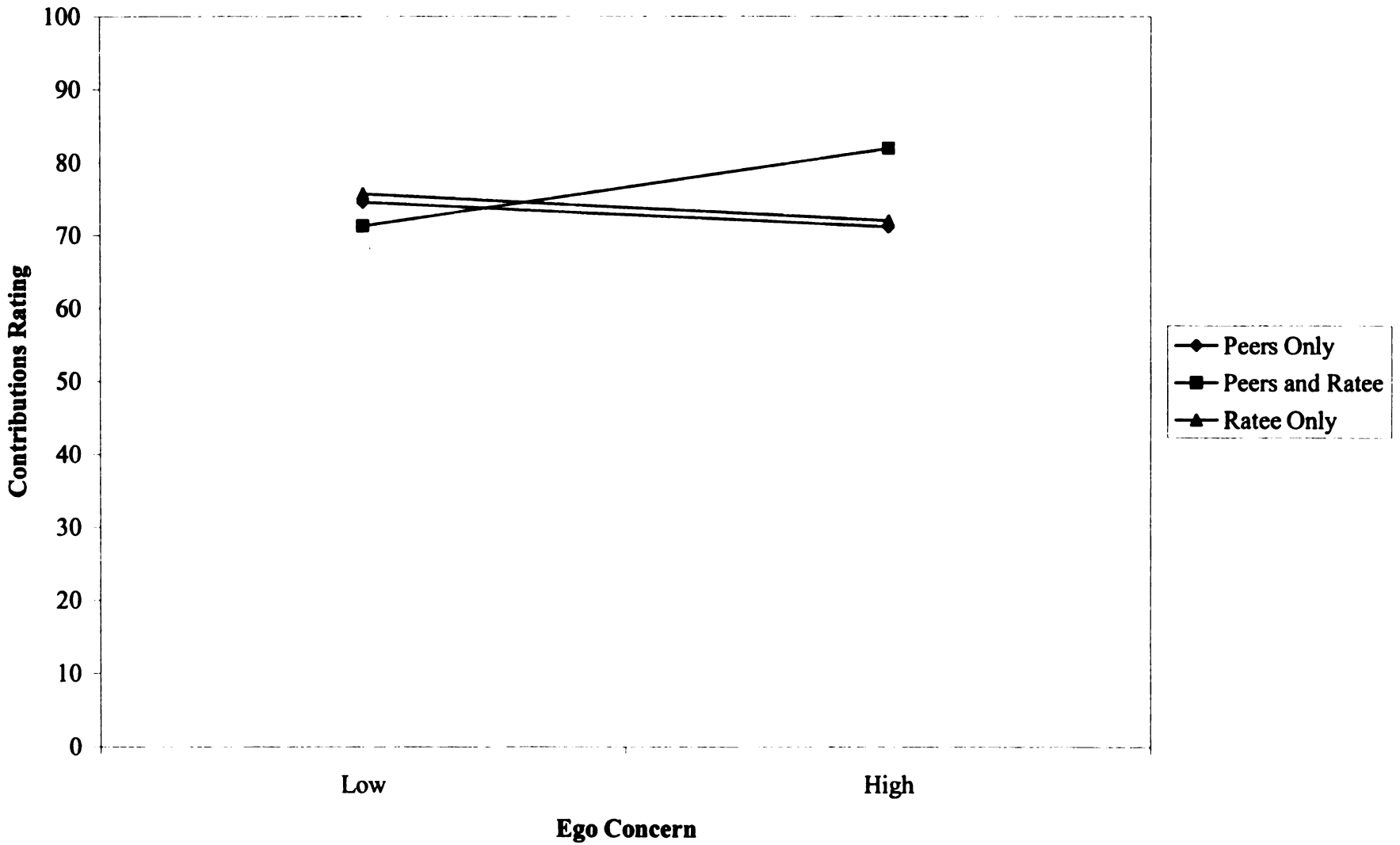
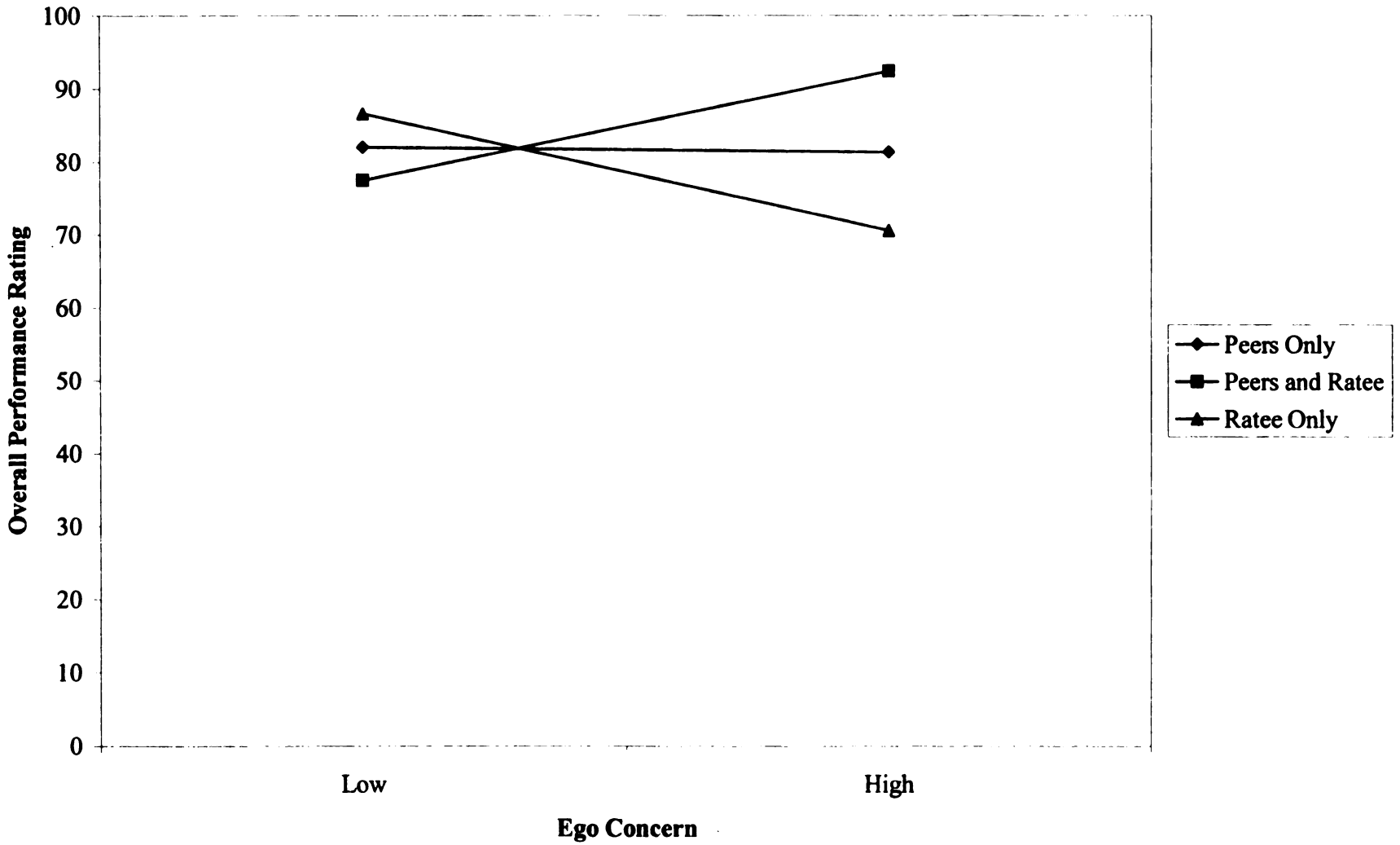


Figure 30. Hypothesis 7b (Results): Accountability X Ego Concern Interaction On Overall Performance Ratings (Grievance Policy Absent)



DISCUSSION

Review of Research Goals

The primary goal of the present research was to examine the joint influences of both situational characteristics and rater individual difference characteristics on the performance ratings provided for poorly performing others. Specifically, the present research examined person X situation interactions on rating leniency. The two situational characteristics examined were rater accountability and whether a formal grievance policy was in force. These particular characteristics were chosen for their likely presence in actual rating scenarios in organizations, and for their potential to produce interpersonal conflict in the context of rating poorly performing others. Further, five individual difference characteristics were chosen for their theoretical and empirical relevance in performance rating contexts. These variables include agreeableness, empathy, conscientiousness, self-monitoring, and ego concern. As noted previously, performance appraisal research has not examined the interactive effects of such situational characteristics and rater individual difference characteristics on rating leniency.

With regard to the rater accountability, a secondary goal of this research was to disentangle the effects of accountability per se from the effects of anonymity. Previous research (e.g., Fisher, 1979; Ilgen & Knowlton, 1980; Klimoski & Inks, 1990; Shapiro, 1975) confounds these variables by making ratings anonymous when raters are not accountable and not anonymous when raters *are* accountable. In real rating contexts, rating anonymity and rater accountability are likely two very different things. In order for a rater to be accountable to a ratee, the rating cannot be anonymous to the ratee. However, it possible for a rater to lack accountability for a rating even when the rating

itself is not anonymous. The current research sought to move one step beyond existing research confounding rater accountability and anonymity by holding lack of anonymity constant and manipulating rater accountability.

Results Pertaining to Research Goals

With regard to the goals that raters are likely to possess in the rating context, Hypotheses 1 and 2 received partial support. Hypothesis 1 predicted that raters who were accountable to their peers would endorse a fairness goal to a greater extent than a liking goal. As shown in Table 8, this hypothesized difference was observed. However, counter to expectations, this same pattern of means was observed when raters were accountable to ratees as well as when they were accountable to both their peers *and* ratees. Given that subjects endorsed a fairness goal to a significantly greater extent than a liking goal regardless of accountability condition, this suggests that the manipulation failed to make them feel accountable toward ratees. Another possibility, however, as discussed in a later section, is that the confederate's behavior was so poor in the current study that fairness (i.e., low ratings) was the only option in the minds of subjects.

Hypothesis 2 predicted that liking and fairness goal valence would relate to performance ratings provided such that liking goal valence would be positively related with performance ratings, and fairness goal valence would be negatively related with performance ratings. As shown in Table 9, these relationships were in the hypothesized directions for all instances except between fairness goal valence and ratings of cooperation. Further, these relationships were significant ($p < .05$) between liking goal valence and ratings of contributions and between liking goal valence and ratings of overall performance. These findings suggest that as Murphy and Cleveland (1995)

suggest, the goals that raters pursue in the performance appraisal context do indeed relate to the ratings that they provide for others.

As summarized in the Results section, none of the hypotheses predicting interactive effects of situational characteristics and rater individual difference characteristics (i.e., Hypotheses 3-7) received convincing support. Despite this lack of support, however, one should not necessarily conclude that these hypotheses are incorrect. Rather, for reasons summarized in a subsequent section, I believe that other problems with the current data should preclude such conclusions.

Rater individual differences. Some previous research examines the relationship between rater individual difference characteristics and rating leniency. For example, research demonstrates that both agreeableness (e.g., Bernardin, Cooke, & Villanova, 2000) and self-monitoring (e.g., Jawahar, 2001) relate positively to performance ratings provided, and that raters' levels of conscientiousness relate negatively to the ratings that they provide (e.g., Bernardin et al., 2000).

Data from the current study do not replicate previously reported results obtained for agreeableness (e.g., Bernardin; Cooke, & Villanova, 2000) and conscientiousness (e.g., Bernardin et al., 2000); these variables did not relate significantly to any of the three performance ratings. However, consistent with the work of Jawahar (2001), a significant positive relationship (controlling for the nested data structure through HLM) was observed between raters' levels of self-monitoring and ratings of cooperation in the presence of a grievance policy (t (130) = 2.51, $p < .05$). Similarly, a marginally significant positive relationship between self-monitoring and ratings of overall performance was observed in the presence of a grievance policy (t (130) = 1.84, $p = .06$).

Previous research has not examined the effects of raters' levels of empathy and ego concern as they relate to rating leniency.

Rater accountability. Some previous research examines the effects of rater accountability on performance ratings, and the *leniency* of these ratings more specifically. For example, Fisher (1979), Ilgen and Knowlton (1980), Klimoski and Inks (1990), and Shapiro (1975) all demonstrate that holding raters accountable to poorly performing ratees leads to rating leniency. Rater accountability in these studies was operationalized as informing raters that they would have to justify, or rationalize performance ratings to ratees.

Data from the current study do not replicate these results. Examining the two conditions in which 1) raters were held accountable only to ratees, and 2) raters were not held accountable to anyone (i.e., the two conditions most similar to past research on rater accountability), rater accountability did not relate significantly to any of the three performance ratings, regardless of the presence or absence of a formal grievance policy. Even with the inclusion of the remaining two rater accountability conditions, no effects were observed for rater accountability on any of the three performance ratings, regardless of the presence or absence of a formal grievance policy. Raw means and standard deviations (i.e., not controlling for the nested data structure) for each of these conditions are displayed in Table 3.

Thus, regarding the goal of disentangling the effects of rater accountability and rating anonymity on leniency, one might argue that the current data render the conclusion that holding raters accountable (i.e., beyond anonymity) for their ratings has no effect on rating leniency, and instead that any effects observed in prior research should be

attributed to anonymity rather than accountability. This conclusion, however, is premature. As noted with regard to the unsupported research hypotheses, I believe that problems with the current data preclude any firm conclusions. These problems are reviewed in the next section.

Problematic Performance Data

An examination of the means and standard deviations for the three performance ratings presented in Tables 2 and 3 reveal problematic distributions for these variables. Specifically, the patterns of means and standard deviations for these variables (Cooperation: $M = 50.61$, $SD = 40.31$; Contributions: $M = 24.63$, $SD = 32.81$; Overall Performance: $M = 33.58$, $SD = 33.89$) suggest non-normal distributions of ratings. Frequency distributions for these variables are presented in Figures 31-33. Indeed, the Contributions (Figure 32) and Overall Performance (Figure 33) distributions are positively skewed (Contributions: *Skewness* = 1.11, *Kurtosis* = -.23; Overall Performance: *Skewness* = .65, *Kurtosis* = -1.02), and the distribution of Cooperation ratings is bimodal (*Skewness* = -.01, *Kurtosis* = -1.65). Such non-normal distributions are problematic and violate assumptions associated with both univariate and multivariate statistics (Tabachnick & Fidell, 2001).

In search of normal distributions. Given that subjects were asked to rate a poorly performing confederate, one would expect a low mean performance rating; however, there is no reason to expect a non-normal distribution of ratings around this mean. Therefore, because non-normal distributions of performance ratings are problematic and obscure the results obtained in the current study, nine sets of additional analyses were conducted in an attempt to eliminate these problems and produce normally distributed

Figure 31. Frequency Distribution of Cooperation Ratings

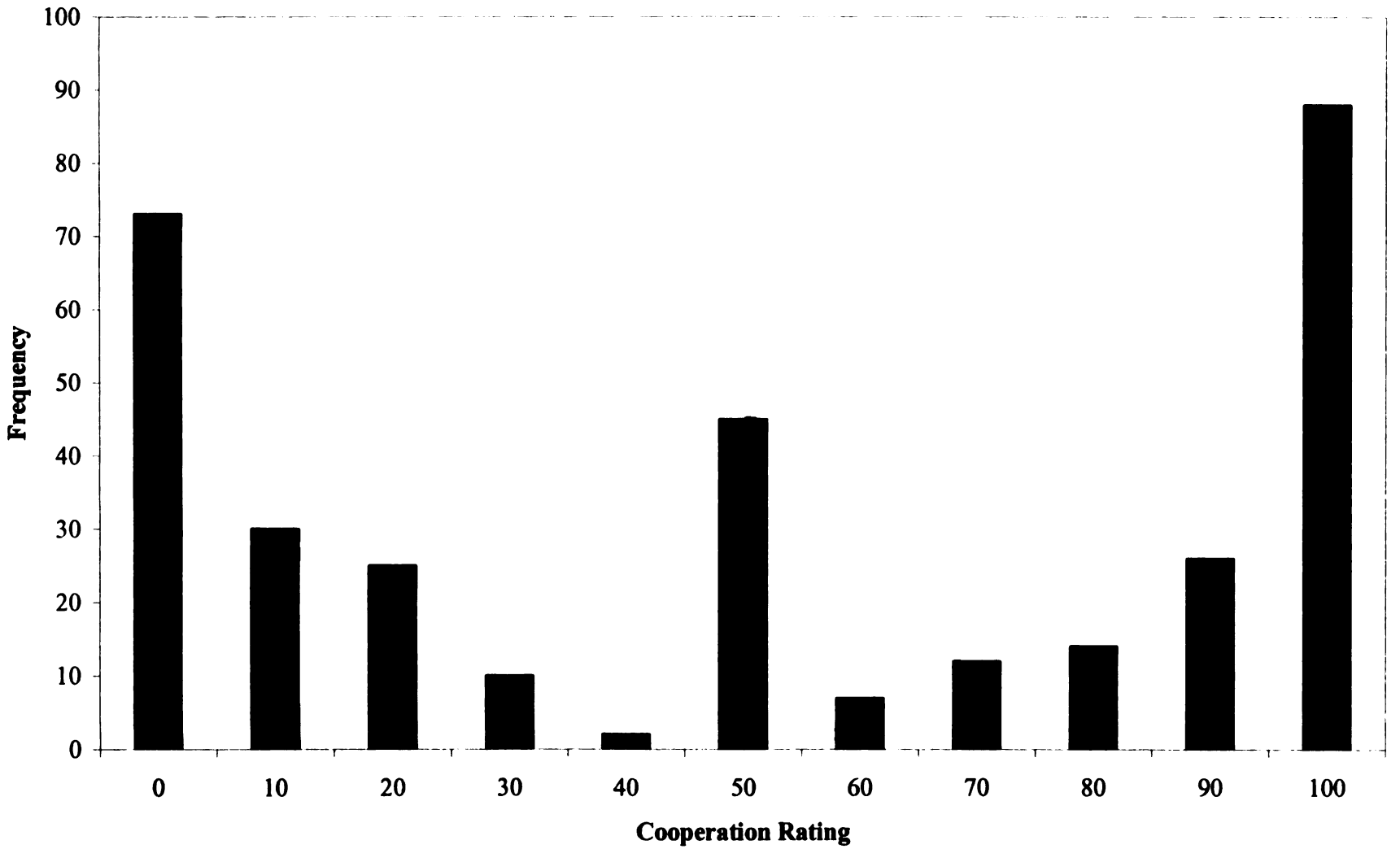


Figure 32. Frequency Distribution of Contributions Ratings

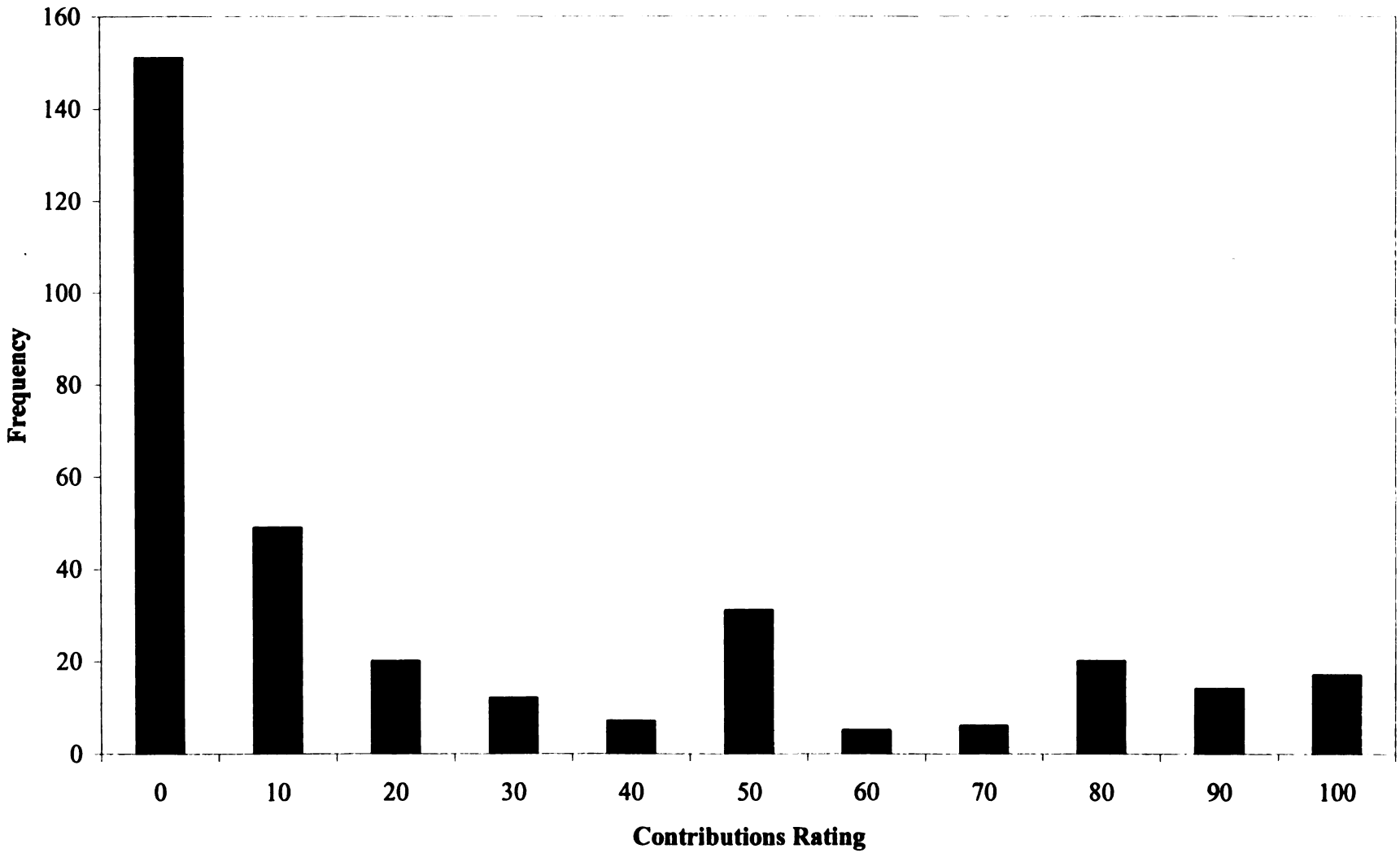
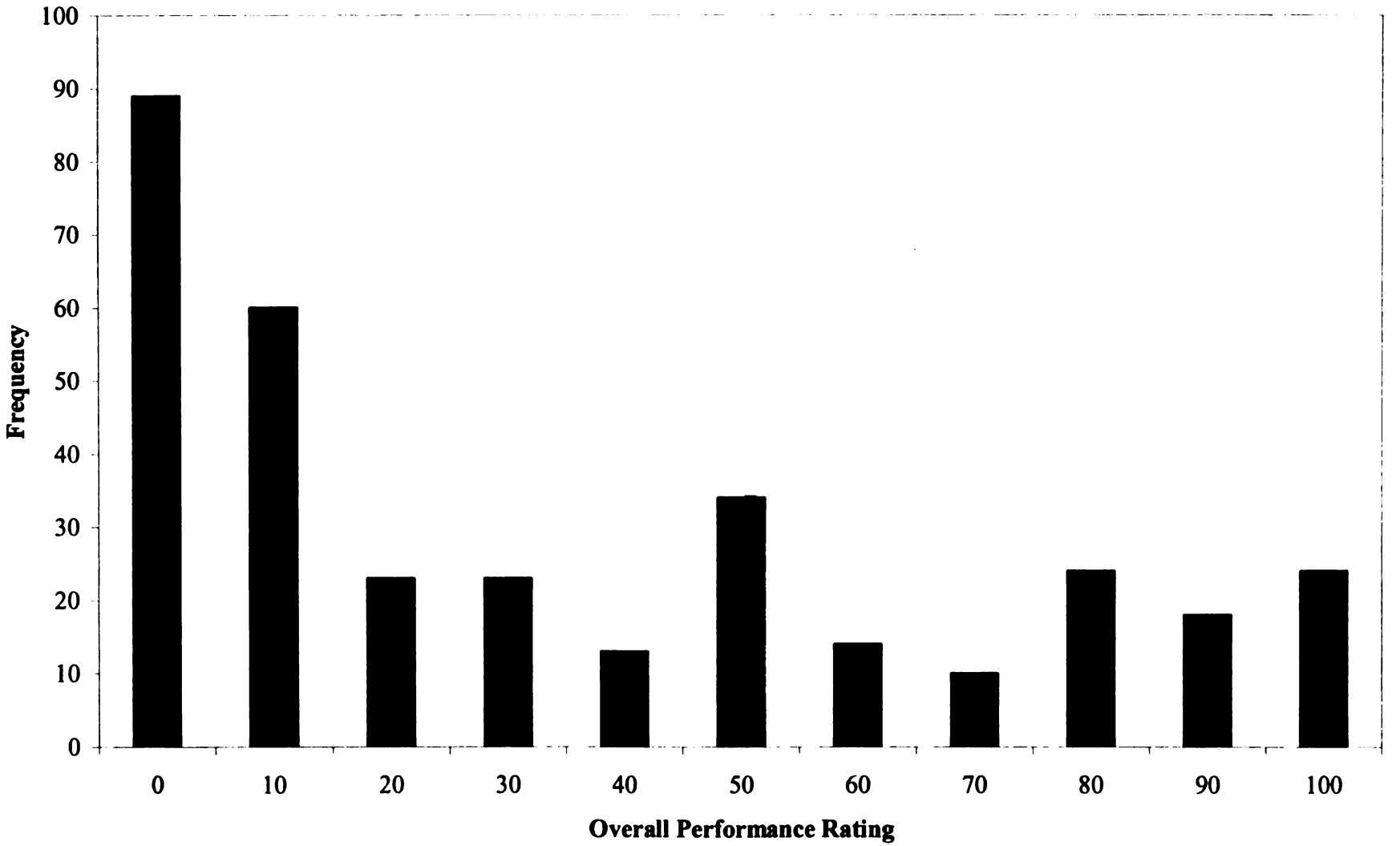


Figure 33. Frequency Distribution of Overall Performance Ratings



dependent variables. Each of these sets of analyses is discussed only briefly, primarily because of their failure to produce more supportive results. Tables 25-27 provide means, standard deviations, skewness, and kurtosis values for these alternative datasets.

Tabachnick and Fidell (2001) summarize a variety of transformations for non-normal data. They suggest that one simply try all possible transformations and use the one that produces the most normal distributions. Using the entire dataset, a LOG10 transformation produced the most normal distributions for the three sets of performance ratings. Using these transformed distributions, those hypotheses specifying effects on performance ratings (i.e., Hypotheses 2-7) were re-analyzed. The results of this set of analyses were no more supportive of the hypotheses than were the original analyses.

Second, it was reasoned that any subject who assigned ratings of 100 for the poorly performing confederate either did not understand what he/she was being asked to do, or he/she did not take the experiment seriously. Thus, all subjects who assigned a 100 for the poorly performing confederate on any of the three dependent variables were excluded from analyses. This resulted in the elimination of 82 subjects. Using this reduced dataset ($N = 250$), Hypotheses 2-7 were re-analyzed, but the hypotheses remained unsupported.

Third, a LOG10 transformation was applied to the reduced dataset just described (i.e., excluding subjects who rated the confederate 100 on any of the performance dimensions). Using this reduced and transformed dataset ($N = 250$), Hypotheses 2-7 were re-analyzed. The results of this set of analyses also failed to support the hypotheses.

Fourth, it was reasoned that any subject who responded incorrectly to any of the manipulation check items (see Appendices E and F) either did not experience the

Table 25

Summary of Means, Standard Deviations, Skewness, and Kurtosis Values for Reduced and Transformed Datasets: Cooperation

Dataset	*N*	*M*	*SD*	*Skewness*	*Kurtosis*
1. Original	332	50.61	40.31	-.01	-1.65
2. LOG10 Transformation of #1	332	1.63	.49	-1.58	1.96
3. Deletion of Subjects Who Assigned the Confederate A Rating of 100	250	34.62	33.30	.48	-1.23
4. LOG10 Transformation of #3	250	1.47	.50	-1.22	.86
5. Deletion of Subjects Who Responded Incorrectly to Any Manipulation Check Items.	209	37.29	34.29	.36	-1.39
6. LOG10 Transformation of #5	209	1.68	.45	-1.70	2.40
7. Deletion of Subjects Who Responded Incorrectly to Any Confederate Behavior Items.	261	45.06	40.04	.22	-1.61
8. LOG10 Transformation of #7	261	1.56	.52	-1.32	1.02
9. Deletion of Subjects Who Responded Incorrectly to Any Manipultion Check or Confederate Behavior Items	159	49.22	40.57	.04	-1.68
10. LOG10 Transformation of #9	159	1.62	.49	-1.47	1.46

Table 26

Summary of Means, Standard Deviations, Skewness, and Kurtosis Values for Reduced and Transformed Datasets: Contributions

Dataset	*N*	*M*	*SD*	*Skewness*	*Kurtosis*
1. Original	332	24.63	32.81	1.11	-.23
2. LOG10 Transformation of #1	332	1.40	.54	-.85	-.07
3. Deletion of Subjects Who Assigned the Confederate A Rating of 100	250	15.88	24.96	1.59	1.37
4. LOG10 Transformation of #3	250	1.25	.53	-.62	-.40
5. Deletion of Subjects Who Responded Incorrectly to Any Manipulation Check Items.	209	17.39	26.38	1.46	.87
6. LOG10 Transformation of #5	209	1.42	.53	-.80	-.27
7. Deletion of Subjects Who Responded Incorrectly to Any Confederate Behavior Items.	261	18.22	27.83	1.49	.97
8. LOG10 Transformation of #7	261	1.28	.56	-.61	-.51
9. Deletion of Subjects Who Responded Incorrectly to Any Manipultion Check or Confederate Behavior Items	159	19.26	28.87	1.43	.72
10. LOG10 Transformation of #9	159	1.28	.55	-.49	-.73

Table 27

Summary of Means, Standard Deviations, Skewness, and Kurtosis Values for Reduced and Transformed Datasets: Overall Performance

Dataset	*N*	*M*	*SD*	*Skewness*	*Kurtosis*
1. Original	332	33.58	33.89	.65	-1.02
2. LOG10 Transformation of #1	332	1.43	.51	-.94	.26
3. Deletion of Subjects Who Assigned the Confederate A Rating of 100	250	23.64	27.33	1.08	-.03
4. LOG10 Transformation of #3	250	1.30	.49	-.71	.00
5. Deletion of Subjects Who Responded Incorrectly to Any Manipulation Check Items.	209	23.97	27.71	1.09	-.01
6. LOG10 Transformation of #5	209	1.45	.50	-.96	.35
7. Deletion of Subjects Who Responded Incorrectly to Any Confederate Behavior Items.	261	26.42	29.88	.94	-.42
8. LOG10 Transformation of #7	261	1.33	.52	-.75	-.10
9. Deletion of Subjects Who Responded Incorrectly to Any Manipultion Check or Confederate Behavior Items	159	26.89	30.26	.94	-.42
10. LOG10 Transformation of #9	159	1.35	.51	-.79	.06

experiment in the manner intended or did not take the experiment seriously. Therefore, any data provided by such subjects is suspect. Thus, all subjects who responded incorrectly to any of the manipulation check questions were excluded from analyses. This resulted in the elimination of 123 subjects. Using this reduced dataset ($N = 209$), Hypotheses 207 were re-analyzed. The results of this set of analyses did not support the hypotheses any more than the original analyses.

Fifth, a LOG10 transformation was applied to the reduced dataset just described (i.e., excluding subjects who responded incorrectly to any of the manipulation check items). Using this reduced and transformed dataset ($N = 209$), Hypotheses 2-7 were re-analyzed, but the results failed to support the hypotheses.

Sixth, it was reasoned that any subject who responded incorrectly to any of the confederate behavior questions (see Appendix H) either did not experience the experiment in the manner intended or did not take the experiment seriously. Therefore, any data provided by such subjects is suspect. Thus, all such subjects were excluded from analyses. This resulted in the elimination of 71 subjects. Hypotheses 2-7 were analyzed using this reduced dataset ($N = 261$). These hypotheses remained unsupported.

Seventh, a LOG10 transformation was applied to the reduced dataset just described (i.e., excluding subjects who responded incorrectly to any of the manipulation check items). Using this reduced and transformed dataset ($N = 261$), Hypotheses 2-7 were re-analyzed, but the results of this set of analyses still did not support the hypotheses.

Eighth, the decision rules from the 4^{th} and 6^{th} set of analyses described above were combined to eliminate a total of 173 subjects who responded incorrectly to any of

the manipulation check (see Appendices E and F) or confederate behavior (see Appendix H) questions. Hypotheses 2-7 were analyzed using this reduced dataset ($N = 159$), but the results were no more supportive of the hypotheses than were the original analyses.

Finally, a LOG10 transformation was applied to the reduced dataset just described (i.e., excluding subjects who responded incorrectly to any of the manipulation check or confederate behavior items). Using this reduced and transformed dataset ($N = 159$), Hypotheses 2-7 were re-analyzed. Similar to other attempts, the results of this final set of analyses failed to provide support for the hypotheses.

Future Directions

Recommendations for future research fall into three categories. First, researchers should pursue person by situation interactions in leniency research. Research focusing solely on individual differences or situational factors lacks comprehensiveness and completeness. Individuals do not behave in vacuums, and it is unreasonable to assume that different individuals respond to the same situations in the exact same manner. Rather, different individuals are likely to respond to different situations in *different* manners. Regarding rating leniency, we know quite a bit about the influence of rater characteristics on rating leniency, just as we understand how some situational characteristics can influence rating leniency. Research on rating leniency is lacking, however, with regard to simultaneous examinations of both rater and situational characteristics.

Second, future researches should attempt to disentangle the effects of rater accountability and rating anonymity on leniency. The current study attempted to do so by holding lack of rating anonymity constant—i.e., regardless of rater accountability, no

ratings were completely anonymous in the current study. An examination of Table 3 reveals that with the exception of Overall Performance ratings in absence of a grievance policy, subjects who received the fourth level of the rater accountability manipulation (i.e., not accountable to anyone) actually rated the confederate *higher* (in terms of raw means) than other accountability conditions; this is counter to expectations. This suggests that subjects who received the fourth level of the accountability manipulation actually felt *more* accountable for their ratings than subjects who received other levels of this manipulation. Future research might replicate the current manipulation with a modification to the fourth level. Specifically, subjects who received this level of the manipulation in the current study were required to go around in a circle and reveal their ratings of everyone (but they did not have to justify these ratings to anyone). A modification to this level of the manipulation might be to simply tell subjects that ratees will *see* their ratings (i.e., rather than make raters speak their ratings).

Finally, the current study should be replicated after modifying the confederate's behavior. One potential reason for the skewed distributions of performance ratings (for Contributions and Overall Performance) is that the confederates' performance in the current study was too poor. By choosing useless items as "important" in the Winter Survival Task, and by not offering any other input when asked to do so, it was reasonable for subjects to provide such low ratings of the confederates' behavior. The current study may have been successful if confederates had chosen more reasonable items as important, still refusing to offer other assistance or input when asked to do so. Thus, future research using poorly performing confederates should ensure that the confederates' behavior is not so low that it will lead to a floor effect in performance ratings.

Limitations

Research using college undergraduates as subjects often notes the limitation of making generalizations to non-college populations. This limitation is particularly important, however, with regard to performance appraisal research. College freshmen simply do not have much (if any) experience rating the performance of others. Therefore, asking them to rate one of their peers (regardless of whether they are instructed to think of him/her as a subordinate or as a peer) may be a difficult and unnatural request. Future research should employ either field sample or perhaps MBA students—i.e., individuals who are accustomed to the performance appraisal context. Obtaining access to field settings in which one can *manipulate* variables such as accountability and the presence/absence of a formal grievance policy may prove to be difficult and unrealistic given the sensitive nature of performance appraisals. To the extent that this is true, future researches should at least attempt to utilize MBA students rather than undergraduate students.

Second, the current research examined subjects rating their peers as peers as opposed to rating them as subordinates. Therefore, one must only generalize to situations in which raters are asked to rate their peers—e.g., faculty rating prospective job candidates, peer ratings in 360 degree feedback programs, etc. One should not generalize to supervisors rating their subordinates or even subordinates rating their supervisors. The rating relationship explored in the current study was lateral. As such, one should only generalize from this study to other lateral rating relationships.

Concluding Comments

Past research offers much insight into rating leniency from both situational perspectives as well as individual difference perspectives. However, researchers have long recognized the importance of considering person by situation interactions when attempting to understand and predict human behavior. Eysenck recently reiterated this importance in a 1997 article when he called for more person by situation interaction research in psychology. The rating leniency literature is lacking in this regard. Instead, past research has relied solely on one perspective or the other (i.e., person vs. situation) rather than joining the two in simultaneous explorations of rating leniency.

How the situation and individual rater affect the rating process has important implications for designing effective, non-biased rating systems and selecting the managers who will eventually use such systems. The more we understand how individual and situational characteristics influence performance ratings, the better equipped we will be to make recommendations to human resource management professionals regarding the use of performance appraisal in their organizations. In order to better understand these influential individual and situational characteristics, we should explore their interactive effects; to date, such research is non-existent in the performance appraisal literature. By examining both individual *as well as* situational characteristics, this research is one step toward filling this gap in the performance appraisal literature.

Appendix A
Power Analysis Results

Analysis	Effect Size		
	Small	Medium	Large
n = 25 (N = 150)			
3-way interaction	0.063	0.207	0.537
2-way interaction	0.063	0.217	0.563
t-test	0.075	0.335	0.755
n = 30 (N = 180)			
3-way interaction	0.067	0.260	0.659
2-way interaction	0.067	0.270	0.680
t-test	0.081	0.399	0.839
n = 35 (N = 210)			
3-way interaction	0.070	0.313	0.756
2-way interaction	0.071	0.323	0.772
t-test	0.087	0.460	0.896
n = 40 (N = 240)			
3-way interaction	0.074	0.365	0.830
2-way interaction	0.075	0.376	0.842
t-test	0.093	0.517	0.934

Appendix B

Winter Survival Task

After your small light aircraft crashes, your group, wearing business/leisure clothing, is stranded on a forested mountain in appalling winter weather (snow covered, sub-freezing conditions), anything between 50 and 200 miles from civilization (you are not sure of your whereabouts, and radio contact was lost one hour before you crashed, so the search operation has no precise idea of your location either). The plane is about to burst into flames and you have a few moments to gather some items. Aside from the clothes you are wearing which does not include coats, you have no other items. It is possible that you may be within mobile phone signal range, but unlikely.

Your group's aim is to survive as a group until rescued. From the following list choose just 10 items that you would take from the plane, after which it and everything inside is destroyed by fire. Splitting or only taking part of items is not permitted. First, you will have three minutes by yourself to consider and draw up your own individual list of what the group should have. Then, you will have 10 minutes as a group to discuss and agree a list of 10 items on behalf of the group. You will need to make sure that everyone helps out with the final list.

Choose 10 items from the following - splitting or only taking part of items is not permitted

1) Pack of 6 boxes x 50 matches
2) Roll of polythene sheeting 3m x 2m
3) Crate of beer (12 liters in total)
4) Blockbuster rental card
5) Bottle of brandy
6) Crate of bottled spring water (twelve liters in total)
7) Small toolbox containing hammer, screwdriver set, adjustable wrench, hacksaw and large pen-knife
8) Box of distress signal flares
9) Small basic first-aid kit containing plasters, bandages, antiseptic ointment, small pair of scissors and pain-killer tablets
10) VHS tape of top 10 Seinfeld episodes
11) Tri-band mobile phone with infrared port and battery half-charged
12) Clockwork transistor radio
13) Milli Vanilli compact disc
14) Gallon container full of fresh water
15) Box of 36 x 50gm chocolate bars
16) Map of the New York Metropolitan Museum of Art
17) Shovel
18) Short hand-held axe
19) Package of magic markers
20) Hand-gun with magazine of 20 rounds
21) 20m of 200kg nylon rope

22) Box of 24 x 20gm bags of peanuts
23) Bag of 10 mixed daily newspapers
24) Box of tissues
25) Bag of 20 fresh apples
26) Electronic calculator
27) Laptop computer with infrared port, modem, unknown software and data, and unknown battery life
28) Inflatable 4-person life-raft
29) Compass
30) Large full Aerosol can of insect killer spray
31) Small half-full aerosol can of air freshener spray
32) Notebook and pencil
33) Jumper cables
34) Box of size 8 women's promotional pink 'Barbie' branded fleece-lined track-suits (quantity is half of each team/group size)
35) Gift hamper containing half-bottle champagne, large tin of luxury biscuits, box of 6 mince pies, 50gm tin of caviar without a ring-pull, a 300gm tin of ham without a ring-pull, and a 500gm Christmas pudding
36) Box of travel games, including chess, back-gammon and draughts
37) Sewing kit
38) Soccer ball
39) Surge protector
40) Whistle
41) Torch with a set of spare batteries
42) Box of 50 night-light 6hr candles
43) Bag of 6 large blankets
44) Can of coffee grounds
45) Roll of toilet paper

Appendix C
Confederate Training

Experiment Overview

Subjects in this study will work as a group to complete the Winter Survival Exercise. This exercise asks individuals to role-play a situation in which they are onboard a plane that crashes in the middle of the forest in the winter. They are told that their plane is about to burst into flames and that they have time (as a group) to keep only 10 items. Subjects are then given time to look through the list of items onboard and pick out the 10 items that they think are most important. Subjects then work with their group to decide upon a final list of 10 items.

Confederates In Research

A confederate is someone who participates in an experiment who is actually *part of* the experiment itself. The confederate's behavior creates a realistic situation for real subjects. In this study, you will be a poorly performing confederate so that the primary question of interest—how individuals respond to poorly performing people—can be examined.

Specific Confederate Behaviors In This Experiment

There are three specific behaviors that you must exhibit in each and every session of this experiment. First, you should ***choose bad items***. Whenever it is time for everyone in the group to reveal the items that they chose to survive, you should reveal your list. It is extremely important, however, that your lists are the exact same across confederates. Therefore, I have selected the 10 items that I would like for you to choose each time.

#4 Blockbuster rental card
#10 VHS tape of top 10 Seinfeld episodes
#13 Milli Vanilli compact disc
#16 Map of the New York Metropolitan Museum of Art
#19 Package of magic markers
#26 Electronic calculator
#33 Jumper cables
#38 Soccer ball
#39 Surge protector
#44 Can of coffee grounds

Second, you should ***be hesitant to offer any suggestions to the group***. After you reveal your initial list of 10 items, you should not offer any assistance or input to the group. If you are asked at any other time what you think about something, you should not offer any suggestions or thoughts. Respond simply by saying, "I really don't care what you guys pick."

Remember

This section highlights some important things for you to remember as well as what to do in certain situations.

1. Do not reveal that you are a confederate to any of the subjects during or after the experiment.
2. If asked to justify the items that you have picked, you should respond by saying, "I don't know; I really don't care."
3. Be sure to behave consistently across trials.
4. Be sure to exhibit the behaviors described above in ***each and every*** session.
5. You should provide the same performance rating for each person; specifically, you should indicate "90" for each and every rating.
6. When it comes time to justify your performance ratings, it does not matter what you say exactly, but try to behave in a manner consistent with your behavior during the rest of the experiment.
7. If it is cold outside, be sure to wear your coat each time you show up (when participating in multiple sessions in a row).
8. It does not matter how you fill out the questionnaires, but be sure to fill them out.

Appendix D
Performance Ratings

Now you will rate each of your group members in terms of their cooperativeness, contributions to the group, and overall performance. You will make these ratings on a scale from 0 to 100. You should provide ratings for each group member other than yourself. ***Note: You should not write anything next to your own name.***

Cooperativeness
How cooperative was the group member?
Scale: 0 = extremely uncooperative, 100 = extremely cooperative

Group Member A ____

Group Member B ____

Group Member C ____

Group Member D ____

Group Member E ____

Group Member F ____

Group Member G ____

Contributions to the group
How much did the group member contribute to the group?
Scale: 0 = did not contribute anything, 100 = contributed a great deal to the group

Group Member A ____

Group Member B ____

Group Member C ____

Group Member D ____

Group Member E ____

Group Member F ____

Group Member G ____

Overall performance
How would you rate the overall performance of the group member?
Scale: 0 = extremely poor performance, 100 = excellent performance

Group Member A ____

Group Member B ____

Group Member C ____

Group Member D ____

Group Member E ____

Group Member F ____

Group Member G ____

Appendix E

Accountability and Presence of Grievance Policy Manipulations

Condition 1

Accountability: Accountable only to peers

Grievance Policy: Present

Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you rate each of your group members' performance you will inform each group member of his/her rating, and then you will justify these ratings to the rest of your group. For example, Group Member A will leave the group for a few moments while you and your group members justify your ratings of Group Member A. After you justify this rating to the rest of the group, everyone *except* the group member being rated will have the opportunity to challenge—or question—your ratings. After you resolve any disputes, you will move on to Group Member B, and so on."

Reminder Before Assigning Ratings: "Remember, after you rate each of your group members' performance you will inform each group member of his/her rating, and then you will justify these ratings to the rest of your group. For example, Group Member A will leave the group for a few moments while you and your group members justify your ratings of Group Member A. After you justify this rating to the rest of the group, everyone *except* the group member being rated will have the opportunity to challenge—or question—your ratings. After you resolve any disputes, you will move on to Group Member B, and so on."

Condition 2

Accountability: Accountable only to ratee

Grievance Policy: Present

Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you rate each of your group members' performance you will inform the entire group of the ratings you provided for each group member, and then you will justify these ratings to each member individually. For example, you will inform the entire group how you rated Group Member A, and then you will justify this rating to Group Member A in a one-on-one discussion. After you justify this rating to the specific group member, he/she will have the opportunity to challenge—or question—your ratings. After you resolve any disputes, you will move on to Group Member B, and so on."

Reminder Before Assigning Ratings: "Remember, after you rate each of your group members' performance you will inform the entire group of the ratings you provided for each group member, and then you will justify these ratings to each member individually. For example, you will inform the entire group how you rated Group Member A, and then you will justify this rating to Group Member A in a one-on-one discussion. After you justify this rating to the specific group member, he/she will have the opportunity to challenge—or question—your ratings. After you resolve any disputes, you will move on to Group Member B, and so on."

Condition 3
Accountability: Accountable to peers and ratee
Grievance Policy: Present
Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you rate each of your group members' performance you will have to go around in a circle and justify these ratings to everyone in your group. After you justify each rating, everyone in your group (including the particular group member being rated) will have the opportunity to challenge—or question—your ratings. After you resolve any disputes, you will move on to Group Member B, and so on."

Reminder Before Assigning Ratings: "Remember, after you rate each of your group members' performance you will have to go around in a circle and justify these ratings to everyone in your group. After you justify each rating, everyone in your group (including the particular group member being rated) will have the opportunity to challenge—or question—your ratings. After you resolve any disputes, you will move on to Group Member B, and so on."

Condition 4
Accountability: Not accountable to anyone
Grievance Policy: Present
Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you make your ratings, your group members will have the opportunity to challenge—or question—your ratings."

Reminder Before Assigning Ratings: "Remember, after you make your ratings, your group members will have the opportunity to challenge—or question—your ratings."

Condition 5
Accountability: Accountable only to peers
Grievance Policy: Absent
Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you rate each of your group members' performance you will inform each group member of his/her rating, and then you will justify these ratings to the rest of your group. For example, Group Member A will leave the group for a few moments while you and your group members justify your ratings of Group Member A. But, keep in mind that no one will be allowed to challenge—or question—your ratings. After you justify Group Member A's rating, you will move on to Group Member B, and so on."

Reminder Before Assigning Ratings: "Remember, after you rate each of your group members' performance you will inform each group member of his/her rating, and then you will justify these ratings to the rest of your group. For example, Group Member A will leave the group for a few moments while you and your group members justify your ratings of Group Member A. But, keep in mind that no one will be allowed to challenge—or question—your ratings. After you justify Group Member A's rating, you will move on to Group Member B, and so on."

Condition 6
Accountability: Accountable only to ratee
Grievance Policy: Absent
Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you rate each of your group members' performance you will inform the entire group of the ratings you provided for each group member, and then you will justify these ratings to each member individually. For example, you will inform the entire group how you rated Group Member A, and then you will justify this rating to Group Member A in a one-on-one discussion. But, keep in mind that no one will be allowed to challenge—or question—your ratings. After you justify Group Member A's rating, you will move on to Group Member B, and so on."

Reminder Before Assigning Ratings: "Remember, after you rate each of your group members' performance you will inform the entire group of the ratings you provided for each group member, and then you will justify these ratings to each member individually. For example, you will inform the entire group how you rated Group Member A, and then you will justify this rating to Group Member A in a one-on-one discussion. But, keep in

mind that no one will be allowed to challenge—or question—your ratings. After you justify Group Member A's rating, you will move on to Group Member B, and so on."

Condition 7
Accountability: Accountable to peers and ratee
Grievance Policy: Absent
Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group. After you rate each of your group members' performance you will have to go around in a circle and justify these ratings to everyone in your group. But, keep in mind that no one will be allowed to challenge—or question—your ratings. After you justify Group Member A's rating, you will move on to Group Member B, and so on."

Reminder Before Assigning Ratings: "Remember, after you rate each of your group members' performance you will have to go around in a circle and justify these ratings to everyone in your group. But, keep in mind that no one will be allowed to challenge—or question—your ratings. After you justify Group Member A's rating, you will move on to Group Member B, and so on."

Condition 8
Accountability: Not accountable to anyone
Grievance Policy: Absent
Wording: "You and your group are going to complete what is known as the Winter Survival Task. You'll receive some specific information on this task in just a few moments. After you and your group complete the Winter Survival Task, you will be asked to rate the performance of each of your group members. When you rate your group members, you should consider each group members' cooperativeness and contributions to the group, but keep in mind that no one will be allowed to challenge—or question—your ratings."

Reminder Before Assigning Ratings: "Remember that no one will be allowed to challenge—or question—your ratings."

Appendix F

Rater Accountability Manipulation Check

INSTRUCTIONS: Please read and answer the following question.

1. I am going to have to ***justify*** my rating of each specific group member to the ***rest of*** the group (for example, I will have to justify my rating of Member A to ***everyone except*** Member A).
   a) Yes
   b) No

2. I am going to have to ***justify*** my rating of each specific group member ***to*** each specific group member (for example, I will have to justify my rating of Member A ***to*** Member A).
   a) Yes
   b) No

3. I am going to have to ***justify*** my rating of each specific group member to the ***entire*** group (for example, I will have to justify my rating of Member A to ***everyone*** in the group, including Member A).
   a) Yes
   b) No

Appendix G

Grievance Policy Manipulation Check

INSTRUCTIONS: Please read and answer the following question.

1. Will group members have the opportunity to object to the ratings that you provide?

Appendix H

Confederate Behavior Check

INSTRUCTIONS: Please think about the group member listed below and answer each question below by choosing the alternative that ***best*** describes the group member.

Please think about Group Member **A** when answering these questions.

1. The group member...
   a) openly contributed to the group without resistance
   b) resisted making contributions to the group

2. The group member...
   a) took the experiment seriously
   b) took the experiment as a joke

3. The group member...
   a) chose reasonable items as "important items"
   b) chose ridiculous items as "important items"

Appendix I
Goal Valence

INSTRUCTIONS: Please read the following statements carefully, and indicate the extent to which you agree or disagree with each statement using the scale provided. There are no right or wrong answers to these questions.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. It is important to me that I provide fair and honest ratings of my group members. **(F)**
2. I really do not care whether my group members think that I'm fair and honest. **(F-)**
3. I am not very concerned whether my group members think that I provide fair and honest ratings. **(F-)**
4. I want to make sure that my group members see me as a fair and honest person. **(F)**
5. It is important to me that my group members like me. **(L)**
6. I really do not care whether my group members think that I'm a nice person. **(L-)**
7. I am not very concerned whether my group members like me. **(L-)**
8. I want to make sure that my group members see me as a nice person. **(L)**

F = Being Fair
L = Being Liked

Appendix J
Agreeableness

INSTRUCTIONS: Please read the following statements carefully, and indicate the extent to which you agree or disagree with each statement using the scale provided. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. There are no right or wrong answers to these questions.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I feel little concern for others. (-)
2. I am interested in people.
3. I insult people. (-)
4. I sympathize with others' feelings.
5. I am not interested in other people's problems. (-)
6. I have a soft heart.
7. I am not really interested in others. (-)
8. I take time out for others.
9. I feel others' emotions.
10. I make people feel at ease.

Appendix K

Conscientiousness

INSTRUCTIONS: Please read the following statements carefully, and indicate the extent to which you agree or disagree with each statement using the scale provided. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. There are no right or wrong answers to these questions.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I am always prepared.
2. I leave my belongings around. (-)
3. I pay attention to details.
4. I make a mess of things. (-)
5. I get chores done right away.
6. I often forget to put things back in their proper place. (-)
7. I like order.
8. I wiggle out of my duties. (-)
9. I follow a schedule.
10. I am exacting in my work.

Appendix L
Empathy

INSTRUCTIONS: Please read the following statements carefully, and indicate the extent to which you agree or disagree with each statement using the scale provided. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. There are no right or wrong answers to these questions.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I often have tender, concerned feelings for people less fortunate than me.
2. Sometimes I don't feel very sorry for other people when they are having problems. **(-)**
3. When I see someone being taken advantage of, I feel kind of protective towards them.
4. Other people's misfortunes do not usually disturb me a great deal. **(-)**
5. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. **(-)**
6. I am often quite touched by things that I see happen.
7. I would describe myself as a pretty soft-hearted person.

Appendix M
Self-Monitoring

INSTRUCTIONS: Please read the following statements carefully, and indicate the extent to which you agree or disagree with each statement using the scale provided. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. There are no right or wrong answers to these questions.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I find it hard to imitate the behavior of other people. (-)
2. My behavior is usually an expression of my true inner feelings, attitudes, and beliefs. (-)
3. At parties and social gatherings, I do not attempt to do or say things that others will like. (-)
4. I can only argue for the ideas that I already believe. (-)
5. I can make impromptu speeches even on topics about which I have almost no information.
6. I guess I put on a show to impress or entertain people.
7. When I am uncertain how to act in a social situation, I look to the behavior of others for cues.
8. I would probably make a good actor.
9. I laugh more when I watch a comedy with others than when alone.
10. In a group of people I am rarely the center of attention. (-)
11. In different situations and with different people, I often act like a very different person.
12. I am not particularly good at making other people like me. (-)
13. Even if I am not enjoying myself, I often pretend to be having a good time.
14. I'm not always the person I appear to be.
15. I would not change my opinions (or the way I do things) in order to please someone else of win their favor. (-)
16. I have considered being an entertainer.
17. In order to get along and be liked, I tend to be what people expect me to be rather than anything else.
18. I have never been good at games like charades or improvisational acting. (-)
19. I have trouble changing my behavior to suit different people and different situations. (-)
20. At a party I let others keep the jokes and stories going. (-)
21. I feel a bit awkward in company and do not show up quite so well as I should. (-)
22. I can look anyone in the eye and tell a lie with a straight face (if for a right end).
23. I may deceive people by being friendly when I really dislike them.

Appendix N
Ego Concern

INSTRUCTIONS: Please read the following statements carefully, and indicate the extent to which you agree or disagree with each statement using the scale provided. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. There are no right or wrong answers to these questions.

1 = strongly disagree
2 = disagree
3 = neutral
4 = agree
5 = strongly agree

1. I do not like to engage in activities that may make me appear incompetent.
2. Before I decide to do something, I consider what others will think of me.
3. I am concerned with protecting my ego.
4. I like it when others recognize that I am good at something.
5. I do not like to hear negative information about myself.
6. When I do something foolish, I like to hide it from others so they do not find out.

Appendix O
Demographics

INSTRUCTIONS: We would like to learn a little more about you. Please read each of the following questions carefully and provide the most accurate answer that you can.

1. What is your sex?
    1) Male
    2) Female

2. What is your age?

3. What is your year in school?
    1) Freshman
    2) Sophomore
    3) Junior
    4) Senior

4. What is your major?
    1) Business
    2) Communications
    3) Education
    4) Engineering
    5) Mathematics
    6) Psychology
    7) Sociology
    8) Other

5. What is your overall GPA? (Indicate "No GPA" if you do not have a GPA yet)
    1) 0.0-0.5
    2) 0.6-1.0
    3) 1.1-1.5
    4) 1.6-2.0
    5) 2.1-2.5
    6) 2.6-3.0
    7) 3.1-3.5
    8) 3.6-4.0
    9) No GPA

6. What was your ACT or SAT score? (If you took both tests, please report your SAT score.)

7. What are your plans after graduation?
    1) Graduate school
    2) Work
    3) Don't know

8. How much work experience do you have working in full-time or part-time jobs?
    2) None (I have never had a full-time or part-time job)
    3) 1 month - 1 year
    4) 1-2 years
    5) 3-4 years
    6) More than 4 years

9. How much experience have you had rating the performance of others?
    1) None (I have never rated the performance of another person)
    2) A little (I have rated the performance of another person one or two times in the past)
    3) Some (I have rated the performance of another person three of four times in the past)
    4) A lot (I have rated the performance of another person five or six times in the past)
    5) Very much (I have rated the performance of another person more than six times in the past)

Appendix P
Informed Consent

Project Title:	Winter Survival Experiment
Investigators' Names:	Brad Chambers and Dr. Neal Schmitt
Description and Explanation of Procedure:	This study investigates how individuals use multiple sources of information when making decisions. In this experiment, you will be asked to work with others on a group task and then provide ratings of their performance. You will also answer a series of questionnaires throughout the experiment.
Time Commitment and Compensation:	This experiment will last 2 hours; you will be compensated with 4 research participation points.
Risks and discomforts:	None

Thank you for participating in our study! If you have any questions or concerns about this study, please feel free to contact Brad Chambers at 355-2171. If, at any time, you feel your questions have not been adequately answered, you may speak with the Head of the Department of Psychology (Dr. Neal Schmitt, 355-9563), or the University Committee on Research Involving Human Subjects (Dr. Ashir Kumar, 355-2180). Your participation in this research is voluntary, and your privacy will be protected to the maximum extent allowable by law. You are free to withdraw this consent and discontinue participation in this project at any time without penalty. If you choose to withdraw from the study prior to its completion, you will receive credit only for the time you have spent in the study. You can be removed from the study for disruptive behavior. If you are removed from the study, you will not receive credit for your participation. Within thee years after your participation, a copy of this signed consent form will be provided to you upon request. You will also receive a copy of this consent form (unsigned) immediately after you participate today.

Your signature below indicates your voluntary agreement to participate in this study.

_____

Name (please print)

_____

Signature

_____

Student Number (This will be used only for giving credit; your identification will not be paired with the data that you provide in any reports.)

_____

Date

Appendix Q
Debrief

The experiment you just completed was concerned with how people use multiple sources of information when making decisions. Specifically, we were interested in seeing how you would rate the performance of your group members after engaging in the survival task. We will examine the answers that subjects provided for the questionnaires that filled out prior to the study, and see if there are particular individual difference characteristics that relate to the ratings provided.

Please do not discuss this experiment with anyone else, for it is important that future subjects know nothing about the experiment before they begin.

The data you provided today is important to us, and we appreciate your help. If you have any questions or comments about today's experiment, please talk to the experimenter now or contact Brad Chambers, Department of Psychology, 22 Baker Hall, (517) 355-2171.

REFERENCES

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299). New York: Academic Press.

Ajzen, I., Timko, C., & White, J. B. (1982). Self-monitoring and the attitude-behavior relation. *Journal of Personality and Social Psychology, 42*(3), 426-435.

Barrett, R. S. (1966). *Performance rating*. Oxford, England: Science Research Assoc., Inc.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review, 3*, 439-449.

Bentler, P. M. (1972). A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1*, 343-357.

Berkowitz, L. (1972). Social norms, feelings, and other factors affecting helping and altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 63-108). New York: Academic Press.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85*(2), 232-236.

Bernardin, H. J., & Villanova, P. (1986). Performance appraisal. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management* (pp. 43-62). Lexington, MA: Lexington Books.

Bjerke, D. G., Cleveland, J. N., Morrison, R. F., & Wilson, W. C. (1987). *Officer fitness report evaluation study* (Navy Personnel Research and Development Center Report, TR). San Diego, CA: NPRDC.

Bliese, P. D. (2002). Multilevel random coefficient modeling in organizational research: Examples using SAS and S-PLUS. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 401-445). San Francisco, CA: Jossey-Bass.

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

Brown, J. D., & Gallagher, F. M. (1992). Coming to terms with failure: Private self-enhancement and public self-effacement. *Journal of Experimental Social Psychology, 28*(1), 3-22.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147-158.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.

BusinessBalls.com. (2002). *Team Building Games*. Available: http://www.businessballs.com/teambuilding.htm [2002, September 18].

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. *Journal of Applied Psychology, 71*(4), 672-678.

Carver, C. S., Lawrence, J. W., & Scheier, M. F. (1996). A control-process perspective on the origins of affect. In L. L. Martin & A. Tesser (Eds.), *Striving and feeling: Interactions among goals,. affect, and self-regulation*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior* (2nd ed.). New York, NY, US: Cambridge University Press.

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74*(1), 130-135.

Coens, T., & Jenkins, M. (2000). *Abolishing performance appraisals: Why they backfire and what to do instead*. San Francisco: Berrett-Koehler Publishers.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90*(2), 218-244.

Crocker, J., Thompson, L. L., McGraw, K. M., & Ingerman, C. (1987). Downward comparison, prejudice, and evaluations of others: Effects of self-esteem and threat. *Journal of Personality and Social Psychology, 52*(5), 907-916.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113-126.

Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisals. *Academy of Management Review, 10*(1), 116-127.

Dipboye, R. L., Smith, C. S., & Howell, W. C. (1994). *Understanding industrial and organizational psychology: An integrated approach*. Fort Worth: Harcourt Brace.

Dobbins, G. H., & Russell, J. M. (1986). The biasing effects of subordinate likeableness on leaders' responses to poor performers: A laboratory and a field study. *Personnel Psychology, 39*(4), 759-777.

Durham, C. C., Locke, E. A., Poon, J. M. L., & McLeod, P. L. (2000). Effects of group goals and time pressure on group efficacy, information-seeking strategy, and performance. *Human Performance, 13*(2), 115-138.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*(1), 91-119.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology, 73*(6), 1224-1237.

Fandt, P. M., & Ferris, G. R. (1990). The management of information and impressions: When employees behave opportunistically. *Organizational Behavior and Human Decision Processes, 45*(1), 140-158.

Farr, J. L. (1973). Response requirements and primacy-recency effects in a simulated selection interview. *Journal of Applied Psychology, 57*(3), 228-232.

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*(1), 31-44.

Ferris, G. R., & Mitchell, T. R. (1987). The components of social influence and their importance for human resources research. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resource management* (Vol. 5, pp. 103-128). Greenwich, CT: JAI Press.

Fisher, C. D. (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation. *Journal of Applied Psychology, 64*(5), 533-540.

Flavell, J. (1985). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Folger, R., Konovsky, M., & Cropanzano, R. (1992). A due process metaphor for performance appraisal. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 14, pp. 127-148). Greenwich, CT: JAI Press.

Gerstein, L. H., Ginter, E. J., & Graziano, W. G. (1985). Self-monitoring, impression management, and interpersonal evaluations. *Journal of Social Psychology, 125*(3), 379-389.

Gibbons, F. X., & Gerrard, M. (1991). Downward social comparison and coping with threat. In J. M. Suls & T. A. Wills (Eds.), *Social comparison: Theory and research* (pp. 317-345). Hillsdale, NJ: Erlbaum.

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

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Harris, M. M., Smith, D. E., & Champagne, D. (1995). A field study of performance appraisal purpose: Research- versus administrative-based ratings. *Personnel Psychology, 48*(1), 151-160.

Ilgen, D. R., & Favero, J. L. (1985). Limits in generalization from psychological research to performance appraisal processes. *Academy of Management Review, 10*(2), 311-321.

Ilgen, D. R., & Knowlton, W. A. (1980). Performance attributional effects on feedback from superiors. *Organizational Behavior and Human Decision Processes, 25*(3), 441-456.

International Personality Item Pool. (2001). *A scientific collaboratory for the development of advanced measures of personality traits and other individual differences*. Available: http://ipip.ori.org [2002, October 9].

Jawahar, I. M. (2001). Attitudes, self-monitoring, and appraisal behaviors. *Journal of Applied Psychology, 86*(5), 875-883.

Jawahar, I. M., & Stone, T. H. (1997). Appraisal purpose versus perceived consequences: The effects of appraisal purpose, perceived consequences, and rater self-monitoring on leniency or ratings and decisions. *Research and Practice in Human Resource Management, 5*, 33-54.

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50*(4), 905-925.

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

Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal, 38*(4), 1036-1051.

Kanfer, R. (1990). Motivation and individual differences in learning: An integration of developmental, differential and cognitive perspectives. *Learning and Individual Differences, 2*(2), 221-239.

Kernan, M. C., & Lord, R. G. (1990). Effects of valence, expectancies, and goal-performance discrepancies in single and multiple goal environments. *Journal of Applied Psychology, 75*(2), 194-203.

Klein, H. J. (1989). An integrated control theory model of work motivation. *Academy of Management Review, 14*(2), 150-172.

Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes, 45*(2), 194-208.

Littlepage, G., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. *Organizational Behavior and Human Decision Processes, 69*(2), 133-147.

Longenecker, C. O., Jaccoud, A. J., Sims, H. P., & Gioia, D. A. (1992). Quantitative and qualitative investigations of affect in executive judgment. *Applied Psychology: An International Review, 41*(1), 21-41.

Longenecker, C. O., Sims, H. P. J., & Gioia, D. A. (1987). Behind the mask: The politics of employee appraisal. *The Academy of Management Executive, 1*(3), 183-193.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49*, 710-721.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology, 69*(1), 147-156.

McNeely, B. L., & Meglino, B. M. (1994). The role of dispositional and situational antecedents in prosocial organizational behavior: An examination of the intended beneficiaries of prosocial behavior. *Journal of Applied Psychology, 79*(6), 836-844.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*(4), 525-543.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*(2), 246-268.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal.* Thousand Oaks: Sage Publications.

Murphy, K. R., Kellam, K. L., Balzer, W. K., & Armstrong, J. G. (1984). Effects of the purpose of rating on accuracy in observing teacher behavior and evaluating teaching performance. *Journal of Educational Psychology, 76*(1), 45-54.

Nathan, B. R., Mohrman, A. M., & Milliman, J. F. (1991). Interpersonal relations as a context for the effects of appraisal interviews on performance and satisfaction: A longitudinal study. *Academy of Management Journal, 34*(2), 352-369.

Naylor, J. C., Pritchard, R. D., & Ilgen, D. R. (1980). *A theory of behavior in organizations*. New York, NY: Academic Press.

Pashler, H. (2000). Task switching and multitask performance. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance* (pp. 277-307). Cambridge, MA: The MIT Press.

Pervin, L. A. (1989). Persons, situations, interactions: The history of a controversy and a discussion of theoretical models. *Academy of Management Review, 14*(3), 350-360.

Ralston, R. W., & Waters, R. O. (1996). The impact of behavioral traits on performance appraisal. *Public Personnel Management, 25*(4), 409-421.

Reichers, A. E. (1986). Conflict and organizational commitments. *Journal of Applied Psychology, 71*(3), 508-514.

Schmidt, A. M. (2000). *Prioritization among conflicting goals: The role of mastery and performance goal orientation.*Unpublished manuscript, East Lansing, MI.

Shapiro, E. G. (1975). Effect of expectations of future interaction on reward allocations in dyads: Equity or equality. *Journal of Personality and Social Psychology, 31*(5), 873-880.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*(4), 526-537.

Snyder, M. (1979). Self-monitoring processes. *Advances In Experimental Social Psychology, 12*, 85-128.

Snyder, M., & Gangestad, S. (1982). Choosing social situations: Two investigations of self-monitoring processes. *Journal of Personality and Social Psychology, 43*(1), 123-135.

Snyder, M., Gangestad, S., & Simpson, J. A. (1983). Choosing friends as activity partners: The role of self-monitoring. *Journal of Personality and Social Psychology, 45*(5), 1061-1072.

Snyder, M., & Kendzierski, D. (1982a). Acting on one's attitudes: Procedures for linking attitude and behavior. *Journal of Experimental Social Psychology, 18*(2), 165-183.

Snyder, M., & Kendzierski, D. (1982b). Choosing social situations: Investigating the origins of correspondence between attitudes and behavior. *Journal of Personality, 50*(3), 280-295.

Snyder, M., & Monson, T. C. (1975). Persons, situations, and the control of social behavior. *Journal of Personality and Social Psychology, 32*(4), 637-644.

Stone, E. F., Stone, D. L., & Dipboye, R. L. (1992). Stigmas in organizations: Race, handicaps, and physical unattractiveness. In K. Kelley (Ed.), *Issues, theory, and research in industrial and organizational psychology* (pp. 385-457). Amsterdam: Elsevier Science.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn and Bacon.

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

Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*(1), 74-83.

Thomas, L. T., & Ganster, D. C. (1995). Impact of family-supportive work variables on work-family conflict and strain: A control perspective. *Journal of Applied Psychology, 80*(1), 6-15.

Tunnell, G. (1980). Intraindividual consistency in personality assessment: The effect of self-monitoring. *Journal of Personality, 48*(2), 220-232.

Vancouver, J. B., & Schmitt, N. W. (1991). An exploratory examination of person-organization fit: Organizational goal congruence. *Personnel Psychology, 44*(2), 333-352.

Villanova, P., Bernardin, H. J., Dahmus, S. A., & Sims, R. L. (1993). Rater leniency and performance appraisal discomfort. *Educational and Psychological Measurement, 53*(3), 789-799.

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Waldman, D. A., & Thornton, G. C. (1988). A field study of rating conditions and leniency in performance appraisal. *Psychological Reports, 63*(3), 835-840.

Wexley, K. N., Alexander, R. A., Greenawalt, J. P., & Couch, M. A. (1980). Attitudinal congruence and similarity as related to interpersonal evaluations in manager-subordinate dyads. *Academy of Management Journal, 23*(2), 320-330.

Wexley, K. N., & Klimoski, R. (1984). Performance appraisal: An update. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 2, pp. 35-79). Greenwich, CT: JAI Pess.

Wexley, K. N., & Yountz, M. A. (1985). Rater beliefs about others: Their effects on rating errors and rater accuracy. *Journal of Occupational Psychology, 58*, 265-275.

White, M. J., & Gerstein, L. H. (1987). Helping: The influence of anticipated social sanctions and self-monitoring. *Journal of Personality, 55*(1), 41-54.

Zanna, M. P., Olson, J. M., & Fazio, R. H. (1980). Attitude-behavior consistency: An individual difference perspective. *Journal of Personality and Social Psychology, 38*(3), 432-440.