

5-2009

# The Effects of Accountability On Leniency Reduction In Self Ratings

Brettney Smith

Clemson University, [brettns@clemson.edu](mailto:brettns@clemson.edu)

Follow this and additional works at: [http://tigerprints.clemson.edu/all\\_theses](http://tigerprints.clemson.edu/all_theses)



Part of the [Industrial and Organizational Psychology Commons](#)

---

## Recommended Citation

Smith, Brettney, "The Effects of Accountability On Leniency Reduction In Self Ratings" (2009). *All Theses*. Paper 555.

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact [awesole@clemson.edu](mailto:awesole@clemson.edu).

THE EFFECTS OF ACCOUNTABILITY ON LENIENCY REDUCTION IN  
SELF RATINGS

---

A Thesis  
Presented to  
the Graduate School of  
Clemson University

---

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science  
Applied Psychology

---

by  
Brettney DāSean Smith  
May 2009

---

Accepted by:  
Dr. Fred Switzer, III, Committee Chair  
Dr. Mary Anne Taylor  
Dr. Claudio Cantalupo

## ABSTRACT

The purpose of the present study was to assess the effects of accountability on leniency reduction in self-ratings. It was hypothesized that participants in both the upward and illegitimate accountability condition would have lower levels of leniency in their self-ratings than participants in the no accountability condition. Accountability was operationalized as participants being told that they would have to justify their self-ratings of driving performance to either a professor who specializes in driving research (upward accountability) or to an education graduate student who maintains the driving simulator (illegitimate accountability) via an audiotape. The results showed that accountability had a significant effect on leniency reduction in self-ratings of driving performance. The implications of these results, limitations, and ideas for future research are discussed.

## ACKNOWLEDGMENTS

First, I would like to give thanks to the one who created me, God. Additionally, I would like to thank my Mom for raising me up into the young, intelligent, and handsome man that I am today. Additionally, I would like to thank my advisors, Drs. Fred Switzer, III, and Mary Anne Taylor. Dr. Switzer has provided me with valuable time and a wealth of knowledge that will always stick with me. Additionally, I will always be in awe of how Dr. Taylor displays such zeal for guiding graduate students, like myself, to become future consultants with such clarity, ease, and tranquility.

I would also like to thank Dr. Claudio Cantalupo for his insights and comments which helped me develop my thesis into a more comprehensible one not filled with I-O jargon.

Last, but definitely not least, I would like to thank Dr. Johnell Brooks, Matt Crisler, and Reeve Goodenough for helping me with scheduling participants in the Driving Simulator and for helping me to properly design, operate, and manage the Driving Simulator. They have been my life savers.

## TABLE OF CONTENTS

	Page
TITLE PAGE .....	i
ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	iii
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
CHAPTER	
I. INTRODUCTION .....	1
Literature Review of Self-Appraisals .....	5
Model and Solutions for Self Assessments.....	18
Self-Appraisals and Accountability .....	22
The Present Study .....	33
Hypotheses .....	37
II. DESIGN AND METHOD .....	39
Participants.....	39
Apparatus .....	39
Measures .....	39
Design .....	41
Procedure .....	42
III. RESULTS .....	44
Initial Analyses .....	44
Effects of Accountability on Self-Evaluations of Driving Performance .....	45
IV. DISCUSSION .....	47
Limitations and Future Research .....	56

Table of Contents (Continued)

	Page
APPENDIX .....	59
Self-Rating Scale of Driving Performance .....	60
REFERENCES .....	66

## LIST OF TABLES

Table		Page
1	Raw data from the driving simulator .....	62
2	Raw data from the self-ratings .....	63
3	Difference scores for Dimensions of Driving Performance .....	64

## LIST OF FIGURES

Figure		Page
1.1	Results for the effects of accountability on self-ratings .....	65



## CHAPTER ONE

### INTRODUCTION

Much appraisal research has cited the incorporation of self-appraisals (i.e., self-evaluations or self-ratings) as a mechanism for increasing perceptions of fairness in the appraisal system. Although performance appraisal systems add value to many companies and organizations, political factors that supervisors, subordinates, and other workers employ to achieve their personal motives and objectives take away from the value of these systems as tools to achieve organizational goals (Longenecker, Sims, and Gioia, 1987). According to organizational researchers, many employees attempt to control their own outcomes and they attempt to either maintain or gain power through the appraisal system. Additionally, the way in which many supervisors appraise subordinates is not based on accuracy indexes more so than these political considerations; some of the reasons why politics supersede accuracy in performance appraisals are because of the desire for employees, to maintain good interpersonal relationships with each other—especially supervisors' desire to maintain good interpersonal relationships with their subordinates—and their acknowledgement of the importance of the appraisal as documentation that could either help advance or hinder an employee's growth within an organization (Longenecker et al., 1987).

As a result of political considerations that are employed in performance appraisals, many employees perceive these appraisal systems as being unfair and not based on their actual performance. In order to overcome these perceptions of unfairness,

self-appraisals have been employed as a participatory mechanism for all employees within the performance appraisal system. Employees like self-appraisals because they perceive them as being fair and being an accurate assessment of their job performance. The use of self-appraisals is one way that each employee has a voice (input) in the system and self-appraisals can be used to increase perceptions of fairness. According to Roberts (2002), performance appraisals that are perceived as unfair by employees have negative outcomes for the organization as a whole, including a lack of motivation to change performance behavior after their appraisals have been given, a rejection of the performance appraisal system as a whole, and a rejection to abide by human resources decisions that are influenced by performance appraisal information. As a result of these perceptions of unfairness, many industrial and organizational psychologists believe that employees' performance will decrease as a result of their viewing the appraisal system as an inaccurate measure of their job performance (Roberts, 1994).

Self-appraisals serve as a source of employee input into the appraisal system because employees assess their own level of performance on various dimensions of their jobs. These self-appraisals serve as physical documentation that can potentially be used as a basis of a supervisor's evaluation of that employee's job performance. As a result of employees being able to evaluate their own performance, self-appraisals increase perceptions of fairness with the performance appraisal system because supervisors can take these self-appraisals into consideration when they evaluate employees' job performance. Research has shown that the employment of self-evaluations in the performance appraisal system increases levels of satisfaction with the performance

appraisal system and with the overall organization; self-evaluations reduce interpersonal conflict between supervisors and subordinates when the latter party receives negative feedback; and the use of self-appraisals reduces the general anxiety and tension that employees have about the performance appraisal process (Roberts, 2002). Therefore, when self-ratings are employed within the performance appraisal system, it fosters an employee's maturation and progression throughout an organization (Roberts, 2002). When workers participate in the appraisal process via the use of self-ratings, they are able to have a voice as to the ratings that they receive from their supervisor because they are able to dispute these ratings that they received via physical or verbal feedback based on their own evaluations of their performance (Roberts, 2002). When workers have a voice in the appraisal process, they take more responsibility for their appraisal ratings that they receive by their supervisors (regardless of whether they are positive or negative performance appraisals), and they are more likely to think that the procedural and distributive processes of the appraisal system are just. Positive interpersonal relationships between supervisors and subordinates are fostered through this form of employee participation, which in turn lead to better supervisory strategies that aid in the development of their subordinates; this subordinate development then eliminates interpersonal disputes and discrepancies that may arise from supervisors' negative feedback or performance ratings of their subordinates (Roberts, 2002).

Although there are many positive benefits to the use of self-ratings, many organizations often avoid self-ratings because they are highly susceptible to leniency bias. Many organizations are wary of using self-ratings because they believe that

employees will not give an accurate assessment of their own job performance. Instead, and as research has shown, employees will be lenient in their self-appraisals of their job performance and they will also lack congruency with other sources of performance appraisal. As will be discussed later in more detail, researchers such as Thornton (1980) have found that employees tended to be more lenient in their self-appraisals of job performance and that these self-appraisals were incongruent with supervisor and peer ratings. Harris and Schaubroeck (1988) also reached this conclusion in their meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings, which is still referenced today by researchers examining the congruence between each source of performance appraisal ratings.

As mentioned earlier, within the performance appraisal system, many appraisal ratings are not based on completely accurate performance measures. Additionally, employees who use self-ratings tend to overrate their performance behavior. Recent research on accountability has shown it to be an effective mechanism in curbing leniency on performance appraisals. Accountability is a current buzzword in the popular press and media, and research on accountability has emerged in various fields from educational psychology to industrial and organizational psychology. In a psychological context, accountability is defined as the internal or external belief that people may be required to justify their behavior (Lerner & Tetlock, 1999).

As a result of the positive implications that self-appraisals and accountability have for the entire appraisal process along with the organization as a whole, it is important to examine research that has been done on self-appraisals, and it is important to examine

research that has assessed the accuracy and problems associated with the usage of this form of evaluation. Also, it is important to examine accountability, specifically to determine if it can be applied to lessen leniency in self-appraisals.

### *Literature Review of Self-Appraisals*

Muchinsky (2003) defines *self-assessments* as employees evaluating their own performance. Usually these self-assessments are used for the improvement of employees within organizations, but 5% of U.S. companies use these self-appraisals for job performance evaluations (Atwater, 1998). For instance, in the city of Bismarck, North Dakota, self-assessments are 25% of city employees' performance evaluations.

The usefulness of self-appraisals is reduced by various factors, such as rating purpose. Bretz, Milkovich, and Read (1992) conducted a literature review on the past appraisal literature and they found—from both laboratory and field studies—that when self-appraisals were used for evaluative purposes, they were prone to highly susceptible to leniency bias. Campbell and Lee (1988) believed that rather than using self-ratings for evaluative purposes, they should be used strictly for developmental purposes, and these authors believed that subsequent employee performance could be improved with the use of self-appraisals because they would form a self-fulfilling prophecy for the employees; in other words, the employees would feel as if their self-appraisals would be reflective of their future performance evaluations from their supervisors.

Early organizational researchers believed that self-assessments could be explained by cognitive mechanisms, particularly self-schemata (Markus, 1977). Markus (1977) found support for the concept of self-schemata, that when employees gathered, stored,

and retrieved information in a certain context, they created self-schemata; in other words, they created systems of cognition about themselves. This author defines self-schemata as a cognitive mechanism that aids one's gathering and synthesizing of behavioral information that is present in one's social environment, and this mechanism comes from generalizations that one makes about himself or herself based on previous behaviors (e.g., past performance; Markus, 1977). The author found that self-schemata help ease the complexity of synthesizing self-related information and they help people predict their own future behavior in an easier manner (Markus, 1977). Along with these findings, Markus (1977) found that self-schemata also made individuals more prone to resist information that did not align with information that was contained in their schemata. This finding was pivotal to self-appraisal research as it would pave the way for future research regarding the impact of self-esteem on leniency in self-appraisal ratings.

Although the previous authors felt that self-rating scales should be used only for developmental purposes rather than for evaluative purposes, these scales can yield accurate information about employees' performance within organizations. But more often than not, research has shown that many inadequacies exist with the usage of these types of performance appraisals.

Thornton (1980) conducted a literature review on the psychometric properties of self-appraisals of job performance by examining factors such as construct validity, variability, halo, and leniency that affect the accuracy of self-appraisals. The author found that self-appraisals often lack congruency with other sources of performance appraisal, such as supervisors and peers, and the author found that self-appraisals were

affected by leniency bias and a lack of variability and a lack of divergent validity (Thornton, 1980). Thornton's (1980) literature review also yielded inconclusive findings regarding whether self-appraisals could be improved via a psychometric property such as its scale format, or via rater training to eliminate potential leniency biases.

Mabe and West (1982) extended the previous literature review by examining the validity of self-evaluations in terms of employees' ability to rate themselves on the dimensions of their job performance. The authors conducted a meta-analytic review of 55 studies that examined self-evaluations of performance when compared with other more objective measures of performance. Mabe and West (1982) found that self-evaluations were correlated with ability and measures of performance when individuals possessed a high level of intelligence, a high need for achievement, and when individuals possessed a high level of internal locus of control. Potential reasons for the lack of alignment between self-ratings and other measures of performance revolved around a rater's past exposure to self-appraisals, and whether workers had to identify themselves when they submitted their self-evaluations (Mabe & West, 1982).

Although the aforementioned reasons were cited as primary reasons for the lack of validity for self-evaluations, other studies showed that the reasons for the inaccuracies for these self-assessments are because they are highly subject to self-esteem, leniency and self-enhancement biases.

Regan, Gosselink, Hubsch, and Ulsh (1975) conducted one of the earliest experiments on the impact of self-esteem on leniency in self-appraisals. The authors examined a need-for-self-esteem and they posited that participants would overrate their

abilities, reject counterfactual information to their self-schemata, and overreact to reverence from fellow participants (Regan et al., 1975). Participants were assigned as either actors or bystanders and the actors rated their own behavior along with bystanders on a task after they were complimented or scorned for their performance. The authors did not find evidence for a need-for-self-esteem notion; in fact, the authors found the contrary, a self-deflation bias in which they rated themselves lower after being criticized and they did not overrespond to praise (Regan et al., 1975). It is important to note that the authors failed to assess the participants' level of self-esteem prior to assigning the participants to the treatment condition and this threatens the internal validity of their study because there was no way to examine whether they obtained their results as a result of participants lacking self-esteem. Also, history effects may have interfered with the study. Subsequent studies did not replicate these authors' findings; indeed, future research they found quite contrary results.

Meyer (1980) tested engineers at General Electric Company for positive leniency bias and he found that they exhibited this characteristic; the majority of the engineers thought that their performance was at the top 25<sup>th</sup> percentile of the company. Although they may have believed this, this is not possible because it is highly unlikely that 90 percent of one's workforce is at top 25<sup>th</sup> percentile for job performance. Anderson, Warner, and Spencer (1984) conducted a study which showed that leniency and self-enhancement are prevalent in self-rating scales. These researchers made up a self-evaluation form for clerical tasks as well as tasks that sounded true but were really false, such as "operating a matriculation machine" and "circumscribing general meeting



registers” (Anderson et al., 1984). The authors of this study found that the applicants inflated their abilities for both the real and bogus tasks (Anderson et al., 1984).

Holzbach (1978) examined leniency bias in self-ratings by looking at performance appraisals of 107 managerial and 76 professional employees in a medium-sized manufacturing location. The author found that self-appraisals had greater levels of leniency than superior or peer assessments, and the authors found a lack of congruency between self-supervisor and self-peer assessments as compared to other forms of appraisals (e.g., supervisor and peer appraisals; Holzbach 1978). Felson (1981) took a different approach to the examination of self-ratings by examining bias in self-ratings among football players via looking at ambiguity and bias in the self-concept. College football players assessed their own performance on seven different dimensions of athletic skill. The coaches served the supervisory role in which they assessed the football players’ levels of performance along with their perceptions of how much confidence the football players possessed; football players rated their own levels of self-confidence by responding to one item that asked them to rate their self-confidence (Felson, 1981). Felson (1981) believed that ambiguity would influence the relationship of leniency levels in the football players’ self-ratings. The author found that ambiguity did influence levels of leniency in self-ratings because when football players were required to rate their own abilities on ambiguous athletic dimensions of skill, they were more prone to inflate their ratings of ability compared to less ambiguous ratings of ability (Felson, 1981). The author believed that the results indicated that ambiguous abilities contained on self-assessments yield responses from self-raters that are indicative of their levels of self-

esteem rather than their true level of ability on those particular ambiguous dimensions (Felson, 1981).

Podsakoff and Organ (1986) identified problems that arise with the use of self-reports in organizational research. The authors identified six uses for self-rating scales:

1. Obtaining demographic or otherwise factual data (such as age or sex of the respondent, years of tenure, etc.) that are, in principle, verifiable from other sources.
2. Assessing the effectiveness of experimental manipulations.
3. Gathering personality data (trait anxiety, need for achievement, locus of control, and so forth).
4. Obtaining descriptions of respondents' past or characteristic behavior (e.g., asking supervisors about their "structuring" behaviors), and/or seeking respondents' intentions of future behavior (e.g., to quit), or how they would behave under certain hypothetical conditions (i.e., various role-playing exercises).
5. Scaling the psychological states of respondents, such as job attitudes, tension, or motivation.
6. Soliciting respondents' perceptions of an external environmental variable (the supervisor's behavior, formalization of organization processes, climate).

Along with these uses of self-report data, other uses of self report data could include:

7. Organizational Citizenship Behaviors (OCBs)
8. Counterproductive work behaviors

The authors indicated that a major problem with these purposes is that they are not distinct from each other when the self-assessments are designed; so, these self-assessments place conflicting demands upon the rater because they ask participants about demographic information, past behavior, predicted future behavior, and external environmental variables, rather than the self-appraisal focusing on a clear and distinct purpose (Podsakoff & Organ, 1986). This, in turn, creates ambiguity (as shown in the previous study) and creates leniency in the self-ratings. Yet, the authors believed that rather than focusing on methods of reducing leniency as a solution to improve the accuracy of self-ratings, the focus should be on examining the way in which the self-assessment scales are created (Podsakoff & Organ, 1986) and they believed that these problems could be alleviated via statistical procedures and scale reordering. Yet, as will be seen later in this literature review, these problems were not and could not be eliminated by enhancing the psychometric properties of the self-appraisals.

Farh and Werbel (1986) conducted experimental research in which they tested for leniency by varying the purpose of the performance appraisals, such as using self-appraisals for administrative purposes (e.g., pay raises, promotions, etc.) or for non-administrative purposes (e.g., fostering employee growth and development). They found that when self-appraisals were used for administrative purposes, they were more lenient than when these appraisals were used for non-administrative purposes (Farh & Werbel, 1986). Yet, when self-appraisals are used for high-expectation validation—the belief that self-reported information will be verified against some other performance measures—and for administrative purposes, the authors of the study found that the appraisals were more

accurate and that they contained less leniency (Farh & Werbel, 1986). This study foreshadows the purpose of the current study because validation is very similar to accountability which was examined in the present study.

Harris and Schaubroeck (1988) conducted an important meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings, which is still referenced by researchers examining the congruence between different sources of performance appraisal ratings. The authors conducted a meta-analysis to determine whether self-ratings lacked congruency with other sources of appraisal ratings because they believed that previous studies yielded inconsistent findings (Harris and Schaubroeck, 1988). Harris and Schaubroeck (1988) found results that aligned with previous studies on the relationship between self- and other sources of performance ratings. They found that a low to moderate correlation existed between self-ratings and supervisor ratings ( $\rho = .35$ ) and between self- and peer ratings ( $\rho = .36$ ). It is important to note that the authors findings put to rest claims by Podsakoff and Organ (1986), along with other researchers who believed that the quality of self-assessments could be improved via improving their psychometric properties because the authors of this study found that rating format and rating scale had no impact on the lack of congruency between self and other sources of appraisals (Harris & Schaubroek, 1988). The authors believed that potential causes for the lack of congruency were a result of leniency within the self-ratings (which comes from the egocentric bias), levels of self-esteem among self-raters, and different conceptions of the job as a result of the source of the performance evaluation (Harris & Schaubroek, 1988).

Farh and Dobbins (1989) examined the effects of self-esteem on leniency bias in self-reports of performance using structural equation modeling by examining undergraduate students' self-appraisals of their grades in an organizational behavior course. The authors found that participants' self-assessments of their job performance exhibited leniency bias and the authors attributed these leniency effects to the participants' levels of self-esteem, and the authors found that higher levels of leniency existed in self-reports when these ratings were made on ambiguous job dimensions (Farh & Dobbins, 1989). The authors suggested that individuals who possessed a high level of self-esteem were more likely to reject negative feedback regarding their performance (Farh & Dobbins, 1989). This study aligns itself with previous research that showed that self-esteem and ambiguity affected self-ratings in the sense that they made them more prone to leniency bias.

Shore, Shore, and Thornton (1992) examined the construct validity of self-evaluations of performance dimensions in assessment centers by testing 394 employees from a large petroleum company on their future managerial abilities and these employees rated themselves on how well they did on all of the dimensions covered in exercises that they had to perform at the assessment center. Shore et al. (1992) found that self-evaluations were correlated with assessor ratings of potential managerial abilities; yet, the authors also found there was very low construct validity for the self-evaluations of performance dimensions in assessment centers. As previous research has indicated, the authors found that the reason for low construct validity was that amount of leniency found in self-assessments. Specifically, the authors found that only 6% of the participants

placed themselves in the bottom third of the assessment group, whereas roughly 57% of the participants indicated that they performed at the upper tier of performers within their assessment group (Shore et al., 1992). The authors indicated that the leniency and rating inflation that they found in self-appraisals aligned with previous research on the existence of leniency bias in self-rating scales (Shore et al., 1992).

John and Robins (1994) took a different approach to the leniency bias in self-ratings by examining individual differences in self-enhancement and the role of narcissism among self-ratings. The authors studied self-assessments of performance to determine their levels of accuracy and leniency bias. 102 MBA students participated in a group discussion that revolved around managerial duties and they assessed their own performance via their involvement in the discussion (John & Robins, 1994). Although individual differences did weaken the level of self-enhancement bias in the self ratings, participants inflated their ratings of their performance in the managerial discussion when compared to their peers and the staff (that oversaw the management discussion group). The authors found that self-enhancement bias had a moderate correlation with narcissism ( $r = .46$ ), a relatively general and constant individual difference variable. This finding suggested that the self-enhancement bias found in the self-ratings reflects an individualistic tendency for people to see themselves in an overly positive light (John & Robbins, 1994).

Similar to the previous study that dealt with a management group discussion, Furnham and Stringfield (1998) assessed the level of congruency between self, peer, and superior assessment assessments by assigning managers to seven teams that were task

oriented. Managers on these teams worked with each other between 4 and 6 months on an organizational project. Upon the completion of the organizational project, each team received ratings from three other sources including themselves (i.e., peers, superior, and a consultant). The authors' findings did not differ from previous literature on self-ratings because they found that self-ratings did not correlate with the other three measures of performance. The authors also found that self-ratings were more lenient than the other measures of performance (Furnham & Stringfield, 1998) which is consistent with previous research as well. Sala and Dwight (2002) conducted an assessment of executive performance among 276 senior executives for a global technological company by distributing self-assessments along with peer assessments to these executives to complete. Although the results showed a strong correlation between direct reports and self-reports, the overall findings of the study indicated that self-reports did not relate to the executives' actual measures of job performance (Sala & Dwight, 2002).

However, it is important to note that leniency is not a universal phenomenon. Farh, Dobbins, and Cheng (1991) who showed that Taiwanese employees tended to rate themselves modestly and hence are not lenient with their self-assessments. Yet, according to studies conducted by Arvey and Murphy (1998) and Yu and Murphy (1993), Chinese employees tended to exhibit leniency bias in their self-ratings of job performance.

Xie, Roy and Chen (2006) examined cultural and individual level differences in self-rating behavior in order to build upon research regarding the cultural relativity hypothesis—which posits that people's social preferences and interpersonal interactions can be explained in part by the individualism-collectivism dichotomy. The authors stated

that previous research has made a faulty assumption that individualism-collectivism lies on a continuum, and the authors believed that these two variables may be orthogonal (Xie et al, 2006). The study examined self-enhancement and general self-efficacy as potential mediators between self-assessments and individualism, and they sought to see whether individualism, specifically the individual assessment of individualism, is the major contributor to the leniency bias in self-ratings as compared to differences on a cultural level (i.e., individualism v. culturalism). The authors found that the individual assessment of individualism was the best predictor of self-rating leniency, and this study questioned the notion of examining leniency effects on a cultural level. Instead, and based on their findings, the authors believe that individual differences would predict one's tendency to overrate themselves on self-appraisals of job performance.

Atwater (1998) notes that research shows that self-assessments are the most accurate for job performance when these assessments are not used for salary increases or advances in job placements (i.e., when they are used for administrative purposes). According to Williams and Levy (1992), when employees are knowledgeable about the mechanisms involved in the job performance appraisal process and there is less ambiguity about the appraisal system, their self-appraisals are more likely to be accurate. Also, Farh and Werbel (1986) found that when employees are certain that their job performance will be based on an objective standard, then their self-appraisals will yield the most accurate results about their job performance.

These reasons for the inaccuracies of self-rating scales are not a surprise to many researchers and psychologists. Gosling, John, Craik, and Robins (1998) found that many



self-concept theorists believe that people strive to keep and heighten their self-worth to various phenomena. Taylor and Brown (1994) found that many people contain personal positive illusions about their job-related abilities and job performance. In fact, the self-enhancement theory is described by many people as if it were an inherent trait in all of us (Gosling et al., 1998). Although many people believe that self-enhancement is a natural human phenomenon, this belief has not been supported through a significant amount of studies (Gosling et al., 1998).

Along with these studies, Gosling et al. (1998) conducted a study in which they assessed the accuracy of self-rating scales by comparing these scales with online codings by observers. Participants had to complete a task and fill out a self-rating scale on how well they completed this task. Initially, the authors of this study found a correlation between the self-reports and the observer codings, and they found that the average relationship between the self-reports and the observer codings was low. But, the authors viewed this correlation as a false representation of the accuracy of the students' self-reports and they decided to only use 12 acts that were deemed as highly reliable by the observers. From this decision, the self reports were more accurate because the average correlation was moderately high ( $r = .40$ ; Gosling et al., 1998).

Although they are subject to leniency effects, it seems as if self-appraisals are viable future mechanisms for organizational assessment because these reports can be used to assess the job performance of workers within an organization (Gosling et al., 1998). Additionally, if they were used, organizations would have more buy-in from

employees about the appraisal system which could lead to increased employee satisfaction and increased employee job performance.

Thus, although self-assessments have been used in many organizations, the reliability and validity of self-assessments are weakened by leniency bias that may result from various factors such as ambiguity among the employees regarding the purpose of the appraisal or as a result of high levels of self-esteem. In order to effectively incorporate self-appraisals into the general appraisal system, these threats to the usefulness of self-appraisals must be eliminated.

#### *Model and Solutions for Self-Assessments*

As a result of the inadequacies that lie within self-assessments, researchers have attempted to develop a general model for self-assessments. Levy (1993) proposed a path-analytic model to predict the relationships among self-esteem, locus of control, self-appraisal, and attributions. 270 participants completed two individual difference measures (self-esteem and locus of control) and then worked on a test that they believed was a managerial selection instrument and made attributions for their performance. After they completed these measures, they then completed self-assessments of their performance. These measures pertained to the various factors that were hypothesized to affect self-assessments. The authors proposed a model that was shown to be significant as a result of their findings. The following are the components of the model:

1. a positive relationship between self-esteem and self-appraisal such that those higher in esteem reported higher self-appraisals than their low esteem counterparts.

2. a negative relationship between locus of control and self-appraisals such that those believing they had more control over their environment reported higher self-appraisal ratings than do who felt that they had less control over their environment.

3. a positive relationship between locus of control and attributional tendency such that those believing that they had more control over their environment made more internal attributions about their performance than those who believed that they had less control over their environment.

4. A negative relationship between self-appraisals and attributional tendency such that those reporting higher self-appraisals made more internal attributions for their performance than those who reported lower self-appraisals.

In the context of Levy's (1993) model of self-appraisals, self-appraisal research has focused on ways to reduce leniency bias as a result of self-esteem and locus of control. Locus of control can be tied to raters' ambiguity regarding the purpose of the appraisal system because if they are unclear about the purpose of the self-ratings, they may feel that they have less control over their environment and as a result be more lenient on these appraisals. Farh, Werbel, and Bedeian (1988) attempted to implement a self-appraisal based performance evaluation system (SABPE) among 88 faculty members at a large land grant university. The SABPE incorporated self-appraisals into traditional supervisory evaluation procedures and faculty members knew the dimensions of performance evaluation that supervisors used when evaluating their performance, such as instructional method (teaching methods/technique, teaching innovation, and curricula

development) and instructional support. The authors found that the SABPE significantly reduced leniency, increased the correlation between self and supervisory rating, and moderated the correlation between self and objective criterion measurements (Farh et al., 1988). Although this was a significant study as it showed a way to improve self-appraisals, it did not have significant generalizability because the SABPE was only focused within the teaching context.

Along with creating a self-appraisal evaluative system for faculty members, Williams and Levy (1992) assessed the effects of perceived system knowledge (PSK) on the agreement between self-ratings and supervisory ratings; PSK was defined as how much knowledge employees thought they had on dimensions of job performance. The study was conducted on 73 employees from two Midwestern financial institutions who completed a PSK questionnaire. The results indicated that self-ratings were correlated more with supervisors' ratings when the subordinates indicated that they had high levels of perceived system knowledge. Yet, it should be noted that in a dissertation, Groehler (1997) assessed whether PSK contributed to the level of agreement between self- and supervisor ratings sources and she found that it did not. Instead, Groehler (1997) stated that actual level of knowledge of performance appraisal system was a more important variable than PSK which would align with Farh and colleagues' (1989) study because faculty members received the information that supervisors used for performance appraisals therefore increasing their actual knowledge of the performance appraisal system.

Along with research on employee's level of knowledge of the performance appraisal system, other researchers assessed the effects of comparative performance information on the accuracy of self-ratings and the agreement between self- and supervisor ratings. Farh and Dobbins (1989) hypothesized that leniency effects would be reduced by increasing the correlation between self-ratings and objective measures of performance and between self-ratings and supervisory ratings when comparative information was used. 163 undergraduate students were either assigned to a control condition or the social comparison condition; regarding the latter condition, participants were provided with information about their hypothetical coworkers who performed the same editing task as they did. The authors found that correlations between self- and supervisor evaluations and between self-evaluations and objective performance indicators were significantly greater when self-raters were given the same comparative information that was available to supervisors. The results obtained in this study can be seen as an addition to the findings of the previous study because they highlight how self-appraisals can be used when employees are aware of the dimensions that supervisors use to evaluate their job performance (Farh & Dobbins, 1989).

Schrader and Steiner (1996) expanded research on comparison standards by assessing which comparison standards would be the most effective at improving the correlation between self and supervisory performance ratings. The authors examined five different types of comparison standards (ambiguous, internal, absolute, relative, and multiple) among 202 supervisors and subordinates from nine different organizations in a

large Southern city. The authors found that more explicit and objective comparative standards produced higher levels of interrater agreement (Schrader & Steiner, 1996).

Although the aforementioned studies on employees' level of knowledge and comparative information have shed light on ways to reduce leniency bias, it is important to revisit the Farh and Werbel (1986) study of the influence of purpose of the performance appraisal and the expectation of validation on self-appraisal leniency. The authors posited that self-appraisals would be less lenient when they were used for a research purpose as compared to a grading purpose, and the authors hypothesized that leniency would be drastically reduced when the 62 participants were told that their self-ratings, regarding their classroom participation, would be validated against some evaluative standard compared to when there was no validation. This study had major implications for the future of research regarding leniency in performance appraisals because researchers have begun to examine a variant of validation, accountability, in the context of how it could reduce leniency and increase accuracy in job performance.

#### *Self-Appraisals and Accountability*

Tetlock (1983) was one of the first researchers to examine the effects of accountability in social psychology. In this initial study, accountability was defined as the expectation of people to provide justification to someone else about their views, and the current definition of accountability is based off of this study (Tetlock, 1983). In their literature review of research on the effects of accountability, Lerner and Tetlock (1999) asserted that there are four specific dimensions of accountability: mere presence, identifiability, evaluation, and reason giving. Mere presence refers to the expectation that

a person will observe one's behavior regarding their performance; identifiability refers to participants belief that they will have to provide proof that they conducted ratings; evaluation refers to participants' belief that their behavior will be assessed by somebody else; and reason giving refers to the participants expectation that they will have to provide an explanation for their actions (Lerner & Tetlock, 1999).

According to Lerner and Tetlock (1999), eight different types of accountability exist: accountability to an audience with known views, accountability to an audience with unknown views, predecisional accountability, postdecisional accountability, outcome accountability, process accountability, legitimate accountability, and illegitimate accountability. Accountability to an audience with unknown views refers to the fact that participants will be expected to provide justification for their performance behavior to an audience whose views are unknown as compared to the contrary with accountability to an audience with known views. Predecisional accountability refers to participants being told that they will have to provide justification for their decisions prior to making them whereas postdecisional accountability refers to participants being told after they have made a decision that they have to justify their rationale for their decisions (Lerner & Tetlock, 1999). Outcome accountability refers to the effectiveness of participants' decisions being the primary criterion for their performance evaluation whereas process accountability refers to participants' decision processes being the criteria for their performance evaluation (Simonson & Staw, 1992). Legitimate accountability refers to participants' belief that they feel obliged to provide justification to a source because they feel that that source should be obeyed whereas illegitimate accountability entails

participants holding a view contrary to legitimate accountability (Tyler, 1997). It is also important to note that Tetlock and Kim (1987) manipulated accountability in terms of preexposure-accountability and post-exposure accountability; preexposure-accountability refers to participants justifying their initial impressions of a test-takers prior to receiving the test-takers' responses versus postexposure-accountability referring to participants justifying their impressions of the test-takers after they received the test-takers' responses (Tetlock & Kim, 1987). Also, Harris (1994) conceptualized accountability in terms of upward versus downward accountability. Upward accountability refers to employees providing justification of their ratings of a subordinate to their supervisor whereas downward accountability refers to employees providing the actual subordinates with justification for their ratings of those subordinates (Harris, 1994). Therefore, because of these many different operationalizations of accountability, it is important to examine research on the effectiveness of accountability on reducing biases.

As mentioned earlier, Tetlock was one of the first social psychologist to study accountability and in one particular study (1983), he examined the effects of accountability on people's stances regarding social issues. It was hypothesized that participants in the accountability conditions would engage in more cognitive thought processing; it was also hypothesized that participants in the accountability condition who had to report to an individual's with known social views would more likely to shift their own views in line with those individuals and engage in less thought processing; participants who had to justify their views to individuals whose views were unknown were more likely to engage in more cognitive thought processing so that they would be



able to justify their decisions to individuals regardless of the stance that those individuals took. Forty-eight participants had to describe their opinions on three social issues—affirmative action, capital punishment, and temporary issues—and the participants were assigned to one of four conditions: accountability to an individual with liberal views, accountability to an individual with conservative views, accountability to an individual with unknown views, and anonymity of their thought processes. The results indicated that participants who reported to individuals with known views were more likely to shift their stances on the social issues to the views of that individual; but when participants were accountable to individuals with unknown views, participants were more likely to engage in more cognitive thought processing (Tetlock, 1983). From Tetlock's (1983) initial study, it seemed as if accountability was most effective when people had to justify their views to people with unknown views.

Tetlock and Kim (1987) extended the previous line of research by examining the effects of accountability on participants' cognitive processing on a personality task. The researchers hypothesized that participants' levels of confidence would decrease on items on a personality measure that could be argued as either having true or false predictions. Sixty undergraduate students were told that they would participate in a person-perception process—how people created opinions of others based off of certain types of information. They were given Personality Research Form (PRF) responses from three persons and they had to write a biographical description of each person based off of their PRF responses. After they completed this task, participants had to predict the three test-takers' likely responses to additional sets of PRF questions and rate their level of confidence of

whether their predictions would be true (likely to occur) or false (unlikely to occur). Participants were assigned to one of three accountability conditions: preexposure-accountability (participants had to describe how they formed impressions of the test-takers' prior to receiving their PRF scores), postexposure accountability (participants had to explain how they formed impressions of the test-takers and how they wrote their test-takers' biographies after receiving the test-takers' PRF scores), and no accountability condition; participants in the accountability condition were told that their interviews would be audiotaped for data-analysis purposes. The results supported the original hypothesis, that participants' levels of confidence would decrease for PRF items that could either have true or false predictions and participants engaged in more cognitive thought processing. Thus, this research supported existing literature that suggested that accountability may be the most effective when participants have to justify their behaviors to an individual with unknown views but this occurred more in the preexposure accountability condition than in either the postexposure- or no-accountability conditions.

Antonioni (1994) took a different approach to the study of accountability from the previous social psychological background by applying the previous research to the workforce. Antonioni (1994) studied the effects of feedback accountability on upward appraisal ratings because the author was interested in whether employees assigned different ratings to their managers based on whether they were held accountable. The author of this study focused on one particular dimension when defining accountability, identifiability, by making the subordinates identify themselves on the upward appraisals. It was hypothesized that managers who knew the names of subordinates rating them

would view the appraisal process more positively than managers who received appraisals from anonymous subordinates; it was hypothesized that subordinates who have to identify themselves on their appraisals of their managers will have a less positive view about the upward appraisal system than subordinates who anonymously appraised their managers; and it was hypothesized that subordinates who were accountable for their upward appraisals of their managers would make more positive ratings than subordinates who provided upward appraisals of their managers anonymously. Thirty-eight managers and 183 subordinates participated in the study, and these participants were either assigned to the accountability condition or the anonymity condition. Subordinates completed the Upward Leadership Behavior Assessment (ULBA) of their managers in either condition, and managers received either a summary of the ULBA report from the anonymous subordinates or complete ULBA reports from accountable subordinates. The results supported the original hypotheses, that managers would support the upward appraisal system more when subordinates were accountable for their ratings, that subordinates in the anonymity condition would feel more positive about the upward appraisal process when they were anonymous (than accountable), and that accountable subordinates were more likely to significantly inflate their ratings of their managers compared to anonymous subordinates. Although this study would seem to imply that accountability would cause more leniency amongst subordinates in the appraisal systems within organizations, caution should be warranted for the author's limited operationalizations of accountability as it only entails one of the four dimensions of accountability, identifiability (Antonioni, 1994).

In comparison to the previous study, Mero and Motowidlo (1995) broadened the scope of accountability by investigating the effects of accountability on the accuracy and favorability of performance ratings. It was hypothesized that raters who are held accountable for their ratings in a motivational context which there are no special pressures to achieve a certain rating outcome will rate more accurately than raters in the same motivational context who are not held accountable, and the authors also hypothesized that motivational contexts that do exert special pressures to achieve certain outcomes. Accountable raters in these situations will feel the personal implications of their ratings more acutely than nonaccountable raters and should be more motivated to avoid personal consequences that might be aversive for them. In comparison to Antonioni's (1994) limited definition of accountability, Mero and Motowidlo (1995) operationalized accountability as the participants being informed that they would have to justify their ratings to the researchers. 247 undergraduate students performed an in-basket task and watched a videotaped simulation over two-weeks during two sessions. The videotape contained vignettes that showed information about 4 simulated subordinates' performance. After assigning the ratings to the simulated subordinates, participants in the accountability condition had to provide their ratings to their supervisors (the researchers) in either a motivational or nonmotivational context; or participants were assigned to the nonaccountability condition. The authors' assessed participants' accuracy by creating a variation of the ratio of positive and negative performance vignettes of the subordinates' performance. Both hypotheses were supported in this study because participants who were held accountable with no motivational contexts rated the simulated subordinates

more accurately than nonaccountable participants. Also, participants who were held accountable with a motivational context, specifically that the subordinates received low performance ratings in the past, were more favorable on their ratings as compared to raters who were not held accountable (Mero & Motowidlo, 1995). The authors showed the potential positive benefits of the use of accountability within the appraisal system, mainly that employees would possibly provide more accurate ratings.

Frink and Ferris (1998) continued to examine the effects of accountability in a performance appraisal process in both a laboratory and field setting. The authors hypothesized that participants would set higher goals in a high-accountability condition as compared to a low or no accountability condition; it was hypothesized that participants would have higher levels of context attentiveness and task attentiveness in the high accountability condition with the task outcomes are the primary criteria of accountability as compared to participants in the low or no-accountability condition; and it was hypothesized that performance would be influenced by the interaction between accountability and goals, in which the correlation between goals and performance in a high accountability condition would be significantly different than in a low versus no accountability condition, and this correlation would be significantly lower in a high accountability-condition in comparison to a low accountability condition. In the laboratory experiment, 115 undergraduate students were assigned to either a high or no accountability condition; in the high accountability condition, participants completed questionnaires dealing with felt accountability, attentiveness, and a mathematical problem set and afterwards they were told that they would have meet with a team leader

to discuss their goals as compared to participants in the no accountability condition being dismissed after conducting the questionnaires. Results from this laboratory experiment supported the three original hypotheses. Along with a laboratory experiment, the authors conducted a field study in which 27 telemarketers completed the same questionnaires as the participants did in the laboratory experiment; yet, rather than being assigned to conditions, the authors performed a median split on the felt accountability questionnaire as a basis of placing telemarketers into a low or high accountability condition. Similar to the laboratory experiment, all of the hypotheses were supported except for a part of the second original hypothesis, that high accountability will result in higher levels of task attentiveness as compared to the other accountability conditions. There were several implications from this study, that participants goal-directed behavior may be reduced by accountability conditions; although this finding differed from previous research findings, the authors results still indicated that accountability caused a greater amount of cognitive thought processing among the telemarketers and participants which falls in line with previous research and it showed how accountability could have positive benefits for organizations.

Beckner, Highhouse, and Hazer (1998) conducted a field study that examined the effects of upward accountability and rating purpose on peer-rater inflation and delay. The authors defined upward accountability as raters' expectation that they would have to provide justification for peers' ratings to their supervisor. 93 clerical, technical, client service, and administrative employees completed a peer-appraisal instrument and they were randomly assigned to a 2 (upward accountability v no accountability) by 2

(administrative purpose versus research purpose) experimental design. The results showed that when workers were held accountable to their supervisors and conducted the peer-appraisals for research purposes only, they were more likely to delay their ratings. The authors found no significant effect for administrative purpose on rater delay, and they also found that purpose had no significant effect on peer-rating inflation. Surprisingly, the authors found no significant effects of upward accountability on peer-rating inflation (Beckner et al., 1998). Although they authors did not find a significant effect for upward accountability on peer-rating inflation, upward accountability may be more effective for subordinates and managers providing ratings of each other.

Gordon and Stuecher (2001) continued the former line of research on upward accountability by examining the effects of accountability and anonymity on the linguistic complexity of teaching evaluations. The authors hypothesized that participants who had to describe their evaluation of their instructor to a faculty member (i.e., upward accountability) would have more complex evaluations and when participants had to describe their instruct evaluations to another student, those evaluations would exhibit less complexity. The authors also hypothesized that the condition in which participants had to provide an explanation of their evaluation to a faculty member along with sign the form would yield the most complex instructor evaluation. 157 undergraduate students were assigned to a two (anonymous v signed) by 3 (low student v high student v high-faculty accountability between subjects design). The authors found that there was no significant main effect for anonymity on the complexity of teaching evaluations but they found that participants in the high-faculty accountability condition exhibited the highest complexity

on their evaluations of their teachers as compared to low and high student accountability conditions. As predicted, the authors also found that condition in which upward accountability was coupled with an identifiable form yielded the most evaluation complexity as compared to the other conditions. Several factors from this study should be noted, that the authors were using anonymity and accountability as two independent variables. This should be of interest to the reader because accountability entails a form of anonymity within the definition, mainly identifiability. Also, in contrast to the Beckner et al. (1998) study and in conjunction with previous accountability research, this study's upward accountability condition did help aid in more complex instructor evaluations which could be argued to produce more accurate student evaluations of instructors.

Instead of looking at the effects of accountability based off of its previous conceptualizations, Brtek and Motowidlo (2002) examined the effects of accountability in terms of procedure and outcome accountability on interview validity. 338 undergraduates were assigned the role of an interviewer and they were assigned to one of four conditions: procedure accountability only, procedure and outcome accountability, outcome accountability only, and no accountability. The authors hypothesized that holding participants procedurally accountable would increase the validity of the interview whereas holding participants accountable for accuracy of the outcome of their interview ratings would lower the validity of the interview. It is also important to note that the positive effects that procedure accountability had on interview validity was mediated by participants' attentiveness (Brtek & Motowidlo, 2002).



Mero, Guidice, and Brownless (2007) studied the effects of audience and form of accounting on rater response and behavior were studied on 197 MBA students, in which these students rated an undergraduate team of students performing a task, and the MBA students had to provide the ratings to either the team of students (downward accountability) or to the session administrator (upward accountability); it is important to note that they also had a mixed accountability condition in which the participants had to justify their ratings to both the session administrator and the team of students (Mero et al., 2007). The participants also had to justify their ratings to each audience either by face-to-face or by justifying their ratings via a written evaluation of why they believed the team of students received the ratings that the participants assigned to them. The authors found that when the participants had to provide their ratings of the student team's performance to the session administrator, these ratings were less inflated and lenient in comparison to the team of students or a no accountability condition; the authors also found that raters who were required to meet face-to-face to provide justification for their ratings were more accurate than participants who had to provide a written justification of their ratings (Mero et al., 2007). It is important to realize that this study is just one of many studies that demonstrated how accountability can reduce inflation and leniency in performance appraisals to yield more accurate ratings.

### *The Present Study*

Although previous research has demonstrated the effects of accountability on performance appraisals, scant literature exists regarding the effects of accountability on reducing leniency of self-appraisals. To the author's knowledge, only one study has

directly examined the effects of accountability on reducing inflation in self-appraisals and this study will be covered in more detail below.

Sedikides, Herbst, Hardin, and Dardis (2002) studied whether accountability was a deterrent to self-enhancement. The authors argued that although self-enhancement is related to many positive social and psychological benefits, it also is associated with social and personal conflict with other people (Sedikides et al., 2002). They conducted four experiments to determine whether accountability could help curb self-enhancement bias in one's self-evaluations. Participants were required to complete an essay about whether the United States should study the planet Mars via space exploration and the participants had to grade their own essay on dimensions that were given to them by the experimenters. After grading their own essays, participants in the accountability (accountability versus no accountability) condition were told that they would have to "explain, justify, and defend" their responses to "an accomplished writer," *Chris Becker*. The authors found that when participants were held accountable for their self-evaluations of their essays, the self-enhancement bias was reduced among the self-evaluations of the essays compared to when the participants were not held accountable. The authors also found from their multiple experiments that the identification category had the most profound effect on curtailing the self-enhancement bias.

Although the authors in the previous study were making a step in the right direction regarding assessing whether accountability could curtail self-enhancement bias in self-evaluations, their study does not directly pertain to performance appraisal. The authors of the previous study limited their generalizability when they made their

participants write an essay about Mars because this topic is not applicable to a performance appraisal context or, arguably, to any daily activity in the participants' lives. Unless people are employees at NASA, they usually do not have to complete essay assignments about a planet within the solar system. The generalizability of this study is further limited because the participants may not have been knowledgeable about the United States' pursuit exploring Mars; more importantly, the participants may not have held much weight to the essay assignment itself because they may have felt that they would not really have to be accountable to another person in order to gain course credit for participation in the study; it seems as if these researchers would have attempted to achieve accountability in a way that would make participants feel that their course credit would depend on how well they justify their ratings to either a member of a high or low audience in order to increase the generalizability of their study. Generally speaking, because the topic in this study had no implications or ties to the introductory psychology course, the participants may not have taken the assignment seriously.

The fact that the study consisted of the participants providing a written justification to "an accomplished writer" was problematic for their operationalization of accountability because they assumed that the "accomplished writer" would be a member of the "high-audience". In reality, it could be argued that the participants saw the accomplished writer as an illegitimate source of accountability because although the fictional "Chris Becker" was well versed in Logic and English, he did not play a pivotal role in the participants fulfilling their introductory psychology course option. Furthermore, the participants may have thought that he was an illegitimate source of

accountability because he was merely a graduate student and they may have believed that he would not play an instrumental role to their hypothetical final essay grade. In addition to these issues, the accountability manipulation checks are questionable. For the first accountability manipulation check, “*My grading of the essay will be \_\_\_\_*,” participants answered this question on a Likert scale that ranged from 1, *completely confidential*, to 7, *attributable to me personally*; it seems as if this manipulation check was illogical because if participants did not feel that the essay grade would not be attributable to them, then a rating of 1 should have stated this exactly rather than *completely confidential*. It seems as if *completely confidential* could have evoked a perception from the participants that their grades would be private and not disclosed publicly (or to their professors). As a result of this illogical flaw on the first accountability manipulation check, the authors could not truly determine how accountable the participants felt for the grade that they received on their essay assignment. Regarding the second manipulation check, “*How well would you assess Chris Becker’s status as an essay reviewer?*” it seems that this question did not address whether participants felt that they had to justify and explain their ratings to this supposed member of the high audience; instead, participants may have thought that this manipulation check merely asked them to rate how qualified he was to review their papers. Therefore, it seems as if the Sedikides et al. (2002) study’s operationalization of accountability is questionable.

Therefore, the current study attempted to replicate the findings from the previous study of the effects of accountability on leniency reduction in self-appraisals. The present study attempted to apply accountability to a more realistic setting because it dealt with

self-appraisals of driving performance. Although driving performance may not be as generalizable as another job dimension, driving performance is multidimensional and it can be argued that driving performance is a job dimension that many companies and organizations must assess from their workers, workers who range from commercial truck drivers, bus drivers, and/or white-collar workers who are required to travel to various locations via an automobile. Accountability was operationalized as one being told that he or she would have to justify his or her driving performance ratings to either a professor who specializes in driving research (Upward Accountability) or to an education graduate student who is interested in driving performance (Illegitimate Accountability) via an audiotape. It is important to note that the current operationalization of Upward Accountability was strengthened because not only did participants in this condition have to report to a member of the higher audience, but also report to a member of the higher audience who had credibility (i.e., professor who has credibility because she specializes in driving research). This operationalization optimized the saliency of accountability.

### *Hypotheses*

Based on the literature concerning self-appraisals and accountability, for the present study, the following hypotheses were developed:

*Hypothesis 1a:* Participants in both accountability conditions would make significantly more accurate ratings of their driving performance than participants in the no-accountability condition.

*Hypothesis 1b:* Participants in the no-accountability condition would have more lenient ratings as compared to participants in both accountability conditions.

*Hypothesis 2a:* Participants in the illegitimate accountability condition would have significantly different ratings as compared to participants in the upward accountability condition.

*Hypothesis 2b:* Participants in the illegitimate accountability condition would have more lenient ratings as compared to participants in the upward accountability condition.

## CHAPTER TWO

### DESIGN AND METHOD

#### *Participants*

Participants were 57 university students from various psychology courses at a public southeastern land-grant university. Students participated in the study voluntarily and they received course credit for their participation in the study.

#### *Apparatus*

The study occurred in a state of the art Driving Simulator Lab, located at the Southeastern University. This Simulator Lab includes a GlobalSim Drive Safety automotive simulator with five forward visual channels and three rearward channels (two side mirrors and one rearview mirror); each channel displays 50 degrees field-of-view and high resolution (1024 x 768) textured graphics. The variables that were measured by the driving simulator were as follows: participant's speed variability while driving, lane position, brake usage at stop signs, following distance, and acknowledgement of traffic lights. The major advantage of using a driving simulator was that it produced objective data that could be compared to the participants' self-ratings of driving performance.

#### *Measures*

*Self-evaluation questionnaire.* Participants assessed their own driving performance on an 11 item self-rating scale (see Appendix A). Subjects were asked to rate their driving performance on a 5-point Likert scale anchored by 5 = Excellent, 1 = Poor (e.g., "How well did you stay in your lane while driving on the curvy roads?").

The questionnaire was developed as follows. The initial questions were drafted based on the facets of driving performance that could be objectively measured in the driving simulator. After the questions were created, pilot studies were conducted in which students completed the driving scenario. After they have completed the driving scenario, students completed the self-rating questionnaires. After students completed the self-rating questionnaires, interviews were conducted with these students regarding the clarity, wording, and general structure of the self-rating questionnaire (and these interviews were taken into consideration if changes needed to be made on the self-rating questionnaire). Data from these pilot studies were collected and basic descriptive statistics were run to ensure that there were no irregularities in the self-rating questionnaire and to ensure that the questionnaire was reliable.

The dimensions of driving performance that was measured were as follows: speed, lane position, red traffic light acknowledgment, braking, and following distances/times. The driving simulator produced data based off of a scenario that was created for the participants that lasted no more than 5 minutes. Participants rated how well they thought they followed the speed regulations throughout the driving scenario and the driving simulator produced data that indicated whether they followed the speed regulations by looking at their average speeds throughout the driving scenario; participants rated how well they thought they stayed in their lane throughout the scenario and the driving simulator produced data that indicated their average lane position throughout the scenario; participants rated how well they thought they acknowledged a red traffic light (via stopping) within the driving scenario and the driving simulator



produced data that indicated whether they acknowledged the red traffic light (via stopping), and the braking dimension was also based off of the red traffic light acknowledgment; participants rated how well they thought they followed a car in front of them and the driving simulator produced data that indicated the following distance of the participants during the scenario.

### *Design*

This study was a post-test only factorial design in which accountability (Upward Accountability, Illegitimate Accountability, and No Accountability) was the independent variable and difference scores (e.g., z-score for the simulator for braking minus z-score for the self-rating score of braking) for the dimensions of driving performance were the dependent variables. In contrast to the Sedikides et al. (2002) study, Upward Accountability entailed participants being required to complete and submit an audiotape in which they explained to the professor—who specializes in driving research—why they rated themselves the way they did (along with the actual self-ratings) for their self-ratings. The participants in the Upward Accountability condition were told that they had to justify their self-ratings of their driving performance on an audiotape that would be sent to a professor who specializes in driving research. Illegitimate Accountability entailed all of the facets that were within the Upward Accountability operationalization but instead of participants justifying their ratings to professor (who specializes in driving research), they justified their self-ratings to a graduate student who maintains the driving simulator. The No Accountability condition entailed participants completing their self-

assessments and giving an explanation of their ratings on an audiotape **AFTER** they completed their self-assessments.

### *Procedure*

The participants entered the lab and they were asked to fill out a consent form. Participants were given 3-4 minutes to practice a warm up scenario in the driving simulator. After the warm up, the participants got into the car and got comfortable with the controls and seating within the car; after they got comfortable, the participants were prompted to begin the driving scenario—which consisted of the variables that are listed in the apparatus section. After completing the scenario, the participants were told to stop. Once the participants stopped the scenario, they got out of the car and if the participants were in the Upward Accountability condition, they were told, “You will be audiotaped [point to the audiotape] about your driving performance after you have completed your self-rating scales. On the audiotape, in two minutes or less, please explain to Dr. Brooks, a professor who specializes in driving research, why the ratings you gave yourself were an accurate indication of your performance in the driving simulator, and this audiotape will be sent to Dr. Brooks”. If the participants were in the Illegitimate Accountability condition, they were told “You will be audiotaped [point to the audiotape] about your interview after you have completed your self-rating scales. On the audiotape, in two minutes or less, please explain to the Matt Crisler, an education graduate student who is interested in driving performance, why the ratings you gave yourself were an accurate indication of your performance in the driving simulator, and this audiotape will be sent to Matt”. If the participants were in the No Accountability condition, they were told, “I will

now give you a self-rating scale in which you will assess your driving performance. After I have handed you the self-rating scale, please begin filling it out. Please let me know if you have any questions.[ After the participant has completed the self-rating scale]In two minute or less, please explain on this audiotape why the ratings you gave yourself were an accurate indication of your performance in the driving simulator”. After each participant completed both the questionnaire, they were given a chance to ask questions, and they were debriefed and dismissed.

## CHAPTER THREE

### RESULTS

#### *Initial Analyses*

Descriptive analyses were conducted to evaluate the characteristics of the independent variable, accountability (upward, illegitimate, and no-accountability), and the dependent variables (the difference scores for acknowledgement of traffic lights, lane keeping, acknowledgement of speed limit, following distance, and braking; see Tables 1, 2, and 3). These statistics were also examined for errors and outliers. These data indicated that 14 of 57 participants did not stop during the stoplight/braking portion of the simulation. The ramifications of this finding will be discussed below.

The dependent variables were calculated as follows:

- (1) For each of the five dimensions/areas of driving (lane-keeping, speed vs. speed limit, following distance, braking, and traffic light compliance) a z-score was calculated for the simulator performance data for each participant.
- (2) For these same dimensions a z-score was calculated for the self-assessment of performance for each participant.
- (3) A difference score between the simulator-rated performance and the self-rated performance was calculated for each dimension for each participant.
- (4) These difference scores are the “raw” dependent variables in this study.

If the participants' difference scores were positive, then these scores indicated that the participants underestimated their driving performance, so they were modest. If the participants' differences scores were negative, then these scores indicated that

the participants they overestimated their performance, they were overconfident and lenient on their self-ratings of their driving performance. We did not use absolute values or squares.

### *Effects of Accountability on Self-Evaluations of Driving Performance*

Two primary inferential analyses were conducted, one using all five of the simulator-generated variables and all participants, and one using a subset of the participants, and one using a subset of both participants and dependent variables. First, a multiple dependent measures analysis of variance (MANOVA) was run using the difference scores for all five dimensions of driving performance. The accountability condition served as the independent variable with three levels: upward v. illegitimate v. no accountability. Note that a step-down/stepwise procedure for selecting the dependent variables was not used here. Choices of dependent variables for MANOVAs were made on conceptual and theoretical bases. The results from the MANOVA indicated that the accountability condition did not have a significant effect on leniency reduction of participants' self-ratings of their driving performance. However, a substantial number of the participants ran the red light during the braking and traffic light portion of the simulation. While this behavior was one of the variables in the first MANOVA, we considered the possibility that this behavior indicated that these participants were not taking the simulation seriously—in other words this behavior may have indicated that these participants were outliers. Therefore, the MANOVA was run again without these participants which left a remainder of 43 data points in the dataset. The results from this MANOVA indicated that the accountability condition had a significant effect on the

leniency reduction of self-rating scales for dimensions of driving performance ( $F(8, 74) = 2.067, p = .05$ ).

Figure 1 indicates that accountability had a significant effect on leniency reduction in self-ratings of driving performance. Overall, the figure indicates that participants in the upward accountability condition exhibited the least leniency in their self-ratings of driving performance across all dimensions. The participants' leniency levels fluctuated across the illegitimate and no accountability conditions across the four dimensions of driving performance.

As shown in Figure 1, the results for the effects of accountability on the accuracy of self-rating scales are mixed. Participants in the no accountability condition tended to most accurately assess their own driving performance on the lane and following dimensions in comparison to participants in the upward and illegitimate accountability conditions. Yet, participants in both upward and illegitimate accountability conditions tended to most accurately assess their own driving performance on the speed and braking dimensions in comparison to participants in the no accountability condition. As an exploratory variable, an analysis of gender differences in response to authority was done and no significant gender differences were found.

Therefore, accountability had a significant effect on the leniency reduction of the participants' self-ratings of their driving performance.

## CHAPTER FOUR

### DISCUSSION

The current study attempted to determine whether accountability could help to curb leniency in self-ratings of driving performance. Although one previous research study attempted to examine the effects of accountability on self-appraisals, the experimental design of that study was questionable. In attempt to make the current study more generalizable to an applied setting, a driving simulator was used in order to have an objective measure of driving performance, a multidimensional construct that is relevant for various applied work settings. The main benefit of using a driving simulator in the current study was that it was an objective measure that could be used to compare the participants' driving performance, per this objective measure, with the participants' self-ratings of driving performance to determine how accurate the participants were and to determine whether or not they overestimated how well they thought they performed in the driving simulator.

The results showed support for an effect of accountability on leniency reduction of self-ratings of driving performance. It is important to note that the operational definitions of accuracy and leniency. Accuracy was defined as how much the Driving Simulator z-scores and self-rating z-scores converged with each other and this is indicated through the difference score; thus, the larger the difference-score (in either a positive or negative direction) the more inaccurate the participants were with their self-ratings of driving performance; the smaller the difference score (in either the positive or negative direction), the more accurate the participants were with their self-ratings of

driving performance. Leniency is defined as the participants overestimating their driving performance by giving themselves higher performance ratings on their self-rating scales (this would be indicated by a negative difference score).

The first hypothesis stated that participants in both accountability conditions would make significantly more accurate ratings of their driving performance than participants in the no-accountability condition. For the four main dimensions of driving performance (speed, following, lane position, and braking) that were measured, this hypothesis was partially supported. The participants in the upward and illegitimate accountability conditions made more accurate ratings of driving performance for the speed and braking dimensions than the participants in the no-accountability condition. The results did not yield similar findings with regards to lane and following dimensions.

The second hypothesis stated that participants in the no-accountability condition would have more lenient ratings as compared to participants in both accountability conditions. The results generally supported this finding (with the exception of the following dimension).

The third hypothesis stated that participants in the illegitimate accountability condition would have significantly different ratings as compared to participants in the upward accountability condition and this hypothesis was generally supported (as indicated in Figure 1). Additionally, it was expected that participants in the upward accountability condition would have the least lenient ratings in comparison to both the illegitimate and no accountability conditions and this finding was supported.



The fourth and final hypothesis stated that participants in the illegitimate accountability condition would have more lenient ratings as compared to participants in the upward accountability condition and the results generally tended to support that finding as indicated in Figure 1. Therefore, the results show that accountability changes the nature of self-ratings/self-appraisals.

It is interesting to note how the difference scores between the no accountability and illegitimate accountability conditions fluctuated between three out of the four dimensions of driving performance, the braking, following, and lane dimensions. Two implications can be drawn from this finding. First, indeed, these three dimensions of driving performance served as manipulation checks for the operationalizations of the illegitimate accountability condition. If participants were assigned to the illegitimate accountability condition, they were informed that they would have to justify their ratings to an education graduate student who was interested in studying their ratings of driving performance. As shown in Figure 1, participants in the illegitimate accountability condition exhibited the most lenient self-ratings of driving performance for these three dimensions, and it seems probable that they exhibited the most leniency because they thought that having to justify their ratings of driving performance for these three measures to an education graduate student was bogus (which was the point of this condition). It seems as if they did not believe that they would be held accountable for their self-ratings when they had to justify them to an education graduate student in comparison to participants who had to justify their self-ratings to a driving researcher or

to no one (as the difference scores for participants in these two conditions were the more modest).

The point that these three dimensions of driving performance served as a manipulation check for the illegitimate accountability condition ties into the second point that these dimensions may not have been taken seriously by the participants in this condition. It could be argued that the speed variable of driving performance was more meaningful to these participants because they realized that more tangible consequences result from poor performance on this dimensions. For instance, if people fail to acknowledge the speed limits in real life, then they can face the possibility of a speeding violation, the loss of their license, or even jail time. But for the lane and following dimensions, it may have been hard for the participants to see any tangible consequences that could result from their poor performance on these two dimensions and so they may have been more prone to give themselves higher ratings on these performance dimensions in the illegitimate accountability dimension than the other two accountability conditions. If people are out of their lane then they attempt to get back in it; even if people are following a car too closely, many of them believe that there is enough room between their own car and the car in front of them and they believe that they are effectively able to handle prevent serious injuries to themselves or to other drivers. In fact, research has shown that people gauge how well they are driving by how well they can maintain their lane position (i.e., how well they can stay in the center of the lane; J. O. Brooks, personal communication, February, 12, 2009). If people use their brakes to stop at a traffic light or prevent themselves from getting into an accident, then they may feel that they are able to

effectively utilize their brakes (the braking pedal). If this logic is applied to the difference scores for these two dimensions, then it is easy to see why participants in the illegitimate accountability condition were more prone to make more lenient self-ratings of their driving performance on these two dimensions; it seems probable that they thought that they could defend their performance on these three dimensions to an education graduate student because of the fact that they would know that the education graduate student would not be as knowledgeable about what constitutes good driving performance in comparison to a driving researcher.

In addition to the noted fluctuations that occurred across the illegitimate and no accountability conditions for the following and lane difference scores, an additional observation was found for participants' performance on the speed dimension of driving performance. The data showed that participants misjudged their speed, specifically they underestimated their velocity (i.e., they thought they were going slower than they actually were). It is also important to provide an explanation for why the traffic light variable was removed from the data analyses. The traffic light variable was thrown out of the data analyses because of some issues that existed with this measurement. The traffic light variable was coded as a dichotomous variable and this type of measurement was inappropriate for the purpose of my experiment. The fact that this variable was dichotomous made it very difficult to gauge how well participants acknowledged the stop light because it may have been the case where the participants stopped at the traffic light but at the last minute (which in real life could have caused an accident or interrupted traffic flow).

The aforementioned findings regarding participants' in the illegitimate accountability condition, self-ratings of driving performance have implications for the use of accountability in applied setting. If organizations implement self-appraisals and hold employees accountable for these appraisals, they must make the employees accountable to a legitimate source. If it is the case where employees would have to justify their ratings to peers or to someone whom the employees would feel that they would not feel obliged to provide justification of their self-ratings of their job performance, then they would be more likely to inflate their ratings of their own job performance. If organizations would make this error in holding participants accountable to an illegitimate source, then their implementation of the self-appraisals within the general performance appraisal system would be counterproductive because they would end up obtaining bogus self-appraisals that would be filled with inflation and leniency bias.

In addition to the manipulation check, one may argue against the need for accountability in self-ratings of job performance based on the results that showed that participants in the no accountability condition were more accurate on some dimensions of driving performance than participants in the upward accountability and at times participants in this condition underestimated their ability of driving performance. But it is important to note that previous research has shown that self-rating scales are inaccurate and prone to self-inflation and leniency biases. Not only is this argument misleading because of previous research, but also is misleading in the context of the current study due to the ambiguity of some of the dimensions of driving performance that were measured. When the participants provided justification for their self-ratings of driving

performance via audiotape, many indicated that the driving simulator felt odd or weird and they felt that it affected their driving ability as compared to a real-time driving environment. For instance, many participants mentioned that both the brake pedal and steering wheel were extremely sensitive than what they were accustomed to in their own cars; the fact that participants made these comments could imply that their self-ratings on lane position, following, and braking may have been affected so that they would be prone to underestimate and rate themselves more poorly on their ability to perform on these three dimensions of driving performance (as Figure 1 indicates). Additionally, the author of the current study created segments of the driving scenario for these three measures of driving performance that produced uncertainty for the participants. For instance, for the following dimension of driving performance, the author created a segment of the scenario where participants were driving on a highway and then they suddenly traversed a two-lane roadway that contained cars driving extremely slow when they switched from the highway to this roadway. Because of the way that this segment (as well as other segments) of the scenario was created, participants in the upward and no accountability conditions may have been prone to underestimate their driving performance for the braking, following and speed dimensions (but not for the illegitimate accountability condition for reasons mentioned in the previous paragraph). Yet, participants in the upward accountability condition may have underestimated their performance even more so than participants in the no-accountability condition because of the fact that they were told that they would have to justify their self-ratings to a professor who specializes in driving research.

The author felt the need to incorporate accuracy into the study because accuracy and leniency usually go hand in hand in the performance appraisal literature; if self-ratings are lenient, then they are most likely to be inaccurate. Yet, many of the dimensions of driving performance that were measured in the current study were plagued with ambiguity and variability within these dimensions which took away from the ability to measure the accuracy of the self-rating scales. Future research should contain cleaner and more precise dimensions of a performance construct that are very comprehensible to the participants in order to measure the accuracy of their self-appraisals. But the aforementioned findings do not take away from the significance of the results that were yielded in the current study. The primary purpose of the study was to determine whether accountability could help reduce leniency in self-ratings of driving performance and the results showed just that. The results showed that accountability can help curb leniency in self-rating scales of driving performance and, in fact, participants in the upward accountability were more likely to underestimate themselves on their driving performance on their self-rating scales in comparison to participants in the other two accountability conditions. Yet, the author believes that this underestimation (overcompensation) effect was an artifact of the current study and that this effect would not exist in an applied setting. If employees were held accountable to their supervisors or to an employee higher in the leader hierarchy of an organization in comparison to participants who were not held accountable to any higher level employee, then there would be a significant gap in the leniency on the self-appraisals that exist between these two groups. Again, as research has shown, it would be most likely that participants who

were not held accountable would exhibit extremely inflated self-ratings of their job performance and participants who were held accountable to a member of a higher audience would not exhibit leniency in their self-appraisals; so, in other words, they would most likely give an accurate assessment of their job performance.

The findings from this study have numerous implications for the use of accountability and self-rating scales in applied settings. As indicated earlier, many employees perceive performance appraisal systems as being unfair and not based on their actual performance. These perceptions also could lead to employees having feelings of negative affect and job dissatisfaction towards their organizations. Recently, self-appraisals have been employed as a participatory mechanism and self-improvement mechanism for all employees because the employees perceive this appraisal as being fair and an accurate assessment of their performance. The use of performance appraisals is one way in which employees can have a voice or input into the performance appraisal process. As a result of them having input in the performance appraisal process, these employees are more likely to perceive the system as being fair as well as the organization as a whole (Roberts, 2002). Yet, research has shown that self-appraisals are highly susceptible to leniency bias as employees tend to inflate their ratings of their job performance. Although the fact that self-appraisals exhibit leniency, the omission of this piece of job performance data in 360 feedback is troubling as the current sources of performance appraisals are flawed. When supervisors provide appraisals of employees' job performance, they are basing them only on hardcore outcomes, such as productivity or sales quotas, more so than the employees' actual job performance. Additionally, peer

ratings of subordinates' job performance are questionable as well because research has shown that many peers (coworkers) are lenient on their appraisals of subordinates' job performance (Muchinsky, 2003). So, it seems most fitting to attempt to increase the accuracy and reduce the leniency of self-appraisals so that they can be incorporated into the performance appraisal system in order to obtain more descriptive and precise information about employees' job performance within organizations.

Based on the findings from the current study, it seems as if accountability seems to be a potential mechanism for curbing leniency and inflation in self-appraisals of job performance. In previous studies accountability has been shown to increase the accuracy of performance appraisals. Accountability can also help curb leniency in self-appraisals of job performance because when employees know that someone else is observing their behavior of their job performance (e.g., coworkers, supervisors, etc.); when employees know that they will be identified on their self-appraisals; when employees know that their self-ratings will be evaluated by a higher authority, and, finally, when employees have to justify their ratings of their own performance to a higher authority (e.g., supervisor or manager), then it seems as if they will be more likely to make accurate ratings of their self-appraisals. Thus, if organizations employ both self-appraisals and accountability into their performance appraisal system, then it seems most likely that they will have happier employees and will believe that they will be appraised on their actual job performance.

#### *Limitations and Future Research*

It is important to note the limitations that exist in the current study. Although my study exhibited minimal threats to internal validity, it faced a few threats to external



validity. One main limitation to the current study was the saliency of the accountability manipulation. Although the current study made a significant improvement to the experimental design and operationalization of accountability over and above those in the Sedikides et al. (2002) study, the operationalization was still weak. The fact that the participants received course credit regardless of the accountability condition may have weakened the effect that accountability had on leniency reduction. Future research should examine the effects of accountability on leniency reduction in self-appraisals with a more hardcore tangible outcomes (promotions, pay raises, etc.) that would be entailed within the upward accountability condition that would be dependent upon them justifying their ratings of their driving performance to a higher authority.

In addition to the saliency of the definition of accountability, the observed effect that was found in the present study for accountability may only be found in laboratory experiments and may not be generalizable to the work setting. Although this may be the case, it seems as if the argument could be made that the effect of accountability on leniency reduction of self-appraisals would be more pronounced in the applied setting because, in comparison to a laboratory setting, there are various financial and hardcore outcomes that could be tied to the accountability condition to make the employees more likely to be more accurate on their self-appraisals if these self-appraisals were incorporated into the performance appraisal system. Future research should examine the effects of accountability on leniency reduction of self-appraisals in the workplace. Additionally, with the use of a laboratory research design, many of the perceived causes

of leniency bias were controlled for, such as ambiguity of the purpose of the appraisal system, and levels of self-esteem.

## APPENDIX

## Self-Rating Scale of Driving-Performance

### Self-Evaluation of Driving Performance Form

Participant #: \_\_\_\_\_

**Instructions: For the following statements please rate the quality of your driving performance in the driving scenario by circling a number from “1”, POOR, to “5”, EXCELLENT (please feel free to make comments under the statements):**

1) How would you rate your performance in staying in your lane while driving on curvy roads during the driving scenario?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

2) How good were you at driving at the speed limit?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

3) How would you rate your acknowledgment of traffic lights in the driving scenario?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

4) How would you rate your performance in maintaining a safe distance between your car and the car in front of you during the driving scenario?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

5) How would you rate your performance in using your brakes when required (e.g., at a stoplight) during the driving scenario?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

6) How would you rate your performance in controlling the steering wheel of the vehicle during the driving scenario?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

7) How would you rate your performance in using your brakes when necessary during the driving scenario?

1	2	3	4	5
Poor	Barely Acceptable	Good	Fairly Well	Excellent

8) How would you rate your performance on not following the car in front of you too closely during the driving scenario?

1                      2                      3                      4                      5  
Poor      Barely Acceptable      Good      Fairly Well      Excellent

9) How would you rate your performance in driving at speeds that are at or around the speed limits in the driving scenario?

1                      2                      3                      4                      5  
Poor      Barely Acceptable      Good      Fairly Well      Excellent

10) How would you rate your performance in stopping at the red stoplight in the driving scenario?

1                      2                      3                      4                      5  
Poor      Barely Acceptable      Good      Fairly Well      Excellent

11) How would you rate your overall driving performance?

1                      2                      3                      4                      5  
Poor      Barely Acceptable      Good      Fairly Well      Excellent

Table 1: Raw Data From the Driving Simulator

Accountability Condition	Lane Keeping: M (SD)	Following: M (SD)	Speed: M (SD)	Traffic Light: M (SD)	Braking: M (SD)
Upward Accountability	.004 (0.16)	1.69 (0.76)	4.07 (2.10)	.68 (0.48)	.03 (0.08)
Illegitimate Accountability	-.03 (0.17)	1.47 (0.42)	3.95 (1.69)	.84 (0.37)	.09 (0.06)
No Accountability	-.02 (0.15)	1.65 (0.84)	3.75 (2.14)	.74 (0.45)	.05 (0.06)

Table 2 Raw Data From The Self-Ratings

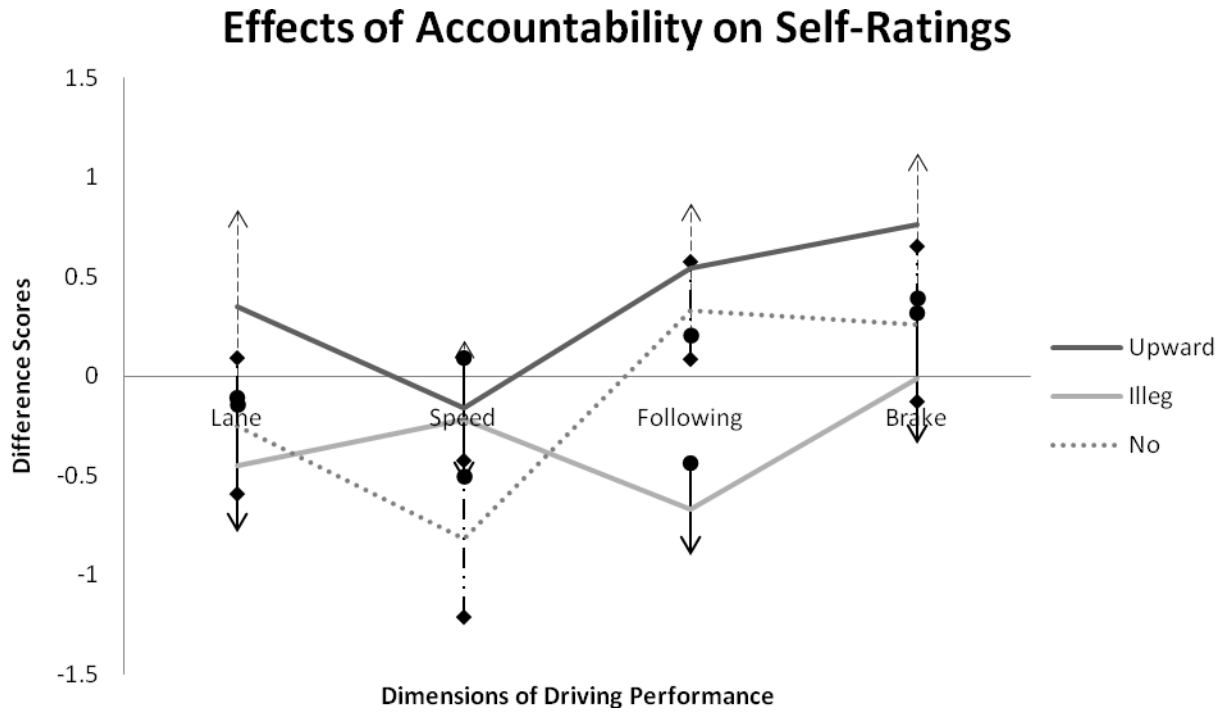
Accountability Condition	Lane Keeping: M (SD)	Following: M (SD)	Speed: M (SD)	Traffic Light: M (SD)	Braking: M (SD)
Upward Accountability	3.32 (0.93)	3.92 (0.77)	3.37 (0.86)	3.95 (0.72)	3.34 (0.71)
Illegitimate Accountability	3.58 (0.56)	4.39 (0.76)	3.68 (0.58)	3.97 (1.09)	3.84 (0.65)
No Accountability	3.42 (0.73)	3.89 (0.97)	3.61 (0.79)	4.08 (0.80)	3.68 (0.65)

Table 3 Difference Scores for Dimensions of Driving Performance

Accountability Condition	Lane Keeping:	Following:	Speed:	Traffic Light:	Braking:
Upward Accountability	.35	.54	-.16	.43	.76
Illegitimate Accountability	-.45	-.67	-.22	.57	-.01
No Accountability	-.25	.33	-.82	.36	.26



Figure 1. The effects of accountability on self-ratings of driving performance (as measured by the difference scores for the lane, speed, following, and braking dimensions).



## REFERENCES

- Anderson, C.C., Warner, J.L., & Spencer, C.C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology, 69*, 574-580.
- Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology, 47*, 349-356.
- Atwater, L.E. (1998). The advantages and pitfalls of self-assessment in organizations. In J.W. Smither (Ed.), *Performance appraisal* (pp. 331-369). San Francisco: Jossey-Bass.
- Beckner, D., Highhouse, S., & Hazer, J.T. (1998). Effects of upward accountability and rating purpose on peer-rater inflation and delay: a field experiment. *Journal of Organizational Behavior, 19*, 219-214.
- Bretz Jr., 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*, 321-352.
- Brtek, M.D., & Motowidlo, S.J. (2002). Effects of procedure and outcome accountability on interview validity. *Journal of Applied Psychology, 87*, 185-191.
- Campbell, D.J., & Lee, C. (1988). Self-appraisal in performance evaluation: Development versus evaluation. *Academy of Management Review, 13*, 302-314.
- Farh, J.L., & Dobbins, G.H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology, 42*, 835-850.
- Farh, J.L., Dobbins, G.H., & Cheng, G.S. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. Workers. *Personnel Psychology, 44*, 129-147.
- Farh, J.L., & Werbel, J.D. (1986). Effects of purpose of the appraisal and expectation of validation on self-appraisal leniency. *Journal of Applied Psychology, 71*, 527-529.
- Farh, J.L., Werbel, J.D., & Bedeian, A.G. (1988). An empirical investigation of self-appraisal-based performance evaluation. *Personnel Psychology, 41*, 141-156.
- Felson, R.B. (1981). Ambiguity and bias in the self-concept. *Social Psychology Quarterly, 44*, 64-69.

- Frink, D.D., & Ferris, G.R. (1998). Accountability, impression management, and goal setting in the performance evaluation process. *Human Relations, 51*, 1259-1283.
- Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360° feedback examining self, manager, peers, and consultant ratings. *Human Relations, 51*, 517-530.
- Gordon, R.A., & Stuecher, U. (1992). The effect of anonymity and increased accountability on the linguistic complexity of teaching evaluation of job applicants. *Journal of Psychology, 126*, 639-649.
- Groehler, L.R. (1997). The effects of performance appraisal system knowledge on the agreement between self- and supervisor ratings of performance (Doctoral dissertation, Illinois Institute of Technology, 1997). *Dissertation Abstracts International, 58*, 6853.
- Harris, M.M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management, 20*, 737-756.
- Harris, M.M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-58.
- Holzbach, R.L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology, 63*, 579-588.
- John, O.P., & Robins, R.W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206-219.
- Lerner, J.S., & Tetlock, P.E. (1999). Accounting for effects of accountability. *Psychological Bulletin, 125*, 255-275.
- Levy, P.E. (1993). Self-appraisal and attributions: A test of a model. *Journal of Management, 19*, 51-62.
- Longenecker, C.O., Sims, H.P., & Gioia, D.A. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive, 1*, 183-193.
- Mabe, P.A. III, & West, S.G. (1988). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280-296.
- Markus, M. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35*, 63-78.

- Mero, N.P., Guidice, R.M., & Brownlee, A.L. (2007). Accountability in a performance appraisal context: The effect of audience and form of accounting on rater response and behavior. *Journal of Management*, 33, 223-252.
- Mero, N.P., & Motowidlo, S.J. (1995). Effects of rater accountability on the accuracy and favorability of performance ratings. *Journal of Applied Psychology*, 4, 517-524.
- Meyer, H.H. (1980). Self-appraisal of job performance. *Personnel Psychology*, 33, 291-295.
- Muchinsky, P.M. (2003). *Psychology applied to work: An introduction to industrial and organizational psychology*. Belmont, CA: Wadsworth/Thomson Learning.
- Podsakoff, P.M., & Organ, D.W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, 12, 531-544.
- Regan, J.W., Gosselink, H., Hubsch, J., & Ulsh, E. (1975). Do people have inflated views of their own ability? *Journal of Personality and Social Psychology*, 31, 295-301.
- Roberts, G.E. (1994). Maximizing performance appraisal system acceptance: Perspectives from municipal government personnel administrators. *Public Personnel Management*, 23, 525-542.
- Roberts, G.E. (2002). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, 31, 333-339.
- Sala, F., & Dwight, S.A. (2002). Predicting executive performance with multirater surveys: Whom you ask makes a difference. *Consulting Psychology Journal: Practice and Research*, 54, 166-172.
- Schrader, B.W., & Steiner, D.D. (1996). Common comparison standards: An approach to improving agreement between self and supervisory performance ratings. *Journal of Applied Psychology*, 81, 813-820.
- Sedikides, C., Herbst, K.C., Hardin, D.P., & Dardis, G.J. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, 83, 592-605.
- Schoon, H. (2008). Person-supervisor fit: Implications for organizational stress, organizational commitment, and job satisfaction. Unpublished master's thesis, Clemson University, Clemson, SC.

- Shore, T.H., Shore, L.M., & Thornton, G.C. III (1992). Construct validity of self- and peer evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology, 77*, 42-54.
- Simonson, I., & Staw, B.M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of Applied Psychology, 77*, 419-426.
- Taylor, S.E., & Brown, J. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*, 21-27.
- Tetlock, P.E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology, 45*, 74-83.
- Tetlock, P.E., & Kim, J.I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52*, 700-709.
- Thornton, G.C., III (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33*, 263-270.
- Tyler, T. (1997). The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Personality and Social Psychological Review, 1*, 323-345.
- Williams, J.R., & Levy, P.E. (1992). The effects of perceived system knowledge on the agreement between self-ratings and supervisor ratings. *Personnel Psychology, 45*, 835-847.
- Xie, J.L., Roy, J.P., & Chen, Z. (2006). Cultural and individual differences in self-rating behavior: An extension and refinement of the cultural relativity hypothesis. *Journal of Organizational Behaviour, 27*, 341-364.