

The Determinants and Performance Effects of Managers' Performance Evaluation Biases

Jasmijn C. Bol

University of Illinois at Urbana–Champaign

ABSTRACT: This study examines the determinants and performance effects of centrality bias and leniency bias. The results show that managers respond to their own incentives and preferences when subjectively evaluating performance. Specifically, information-gathering costs and strong employee-manager relationships positively affect centrality bias and leniency bias. The findings also indicate that performance evaluation biases affect not only current performance ratings, but also future employee incentives. Inconsistent with predictions based on the agency perspective, the results show that managers' performance evaluation biases are not necessarily detrimental to compensation contracting. Although centrality bias negatively affects performance improvement, the evidence does not reveal a significant negative relation between leniency bias and performance. Rather, leniency bias is positively associated with future performance, which is consistent with the behavioral argument that bias can improve perceived fairness and, in turn, employee motivation.

Keywords: *subjectivity; performance evaluation; centrality bias; leniency bias.*

Data Availability: *Data used in this study cannot be made public due to a confidentiality agreement with the participating firm.*

I. INTRODUCTION

The objective of this study is to examine the determinants and performance effects of managers' performance evaluation biases. Understanding the causes and performance implications of bias is important for executives in charge of incentive design, as it can help them determine how much discretion to allow managers in performance-based compensation plans and, hence, can lead to improved incentive contracting.

I am grateful to the participating company for allowing me access to the data. I also thank Rashad Abdel-khalik, Stan Baiman, Antonio Dávila, Henri Dekker, Chris Ittner, Steven Kachelmeier (senior editor), Ranjani Krishnan (editor), Justin Leiby, Petro Lisowsky, Kenneth Merchant, Frank Moers, Mark Peecher, Fernando Peñalva, Joan Enric Ricart, Steven Smith, Stan Veuger, two anonymous reviewers, and seminar participants at the 2006 Global Management Accounting Research Symposium in Copenhagen, the 2006 AAA Annual Meeting in Washington, D.C., University of Pennsylvania, London School of Economics, University of Colorado at Boulder, University of Illinois at Urbana–Champaign, London Business School, Harvard Business School, University of Southern California, University of Lausanne, and University of Maryland for their comments and suggestions. This research was partly funded by the Prins Bernhard Cultuurfonds.

Editor's note: Accepted by Ranjani Krishnan.

Submitted: September 2008

Accepted: March 2011

Published Online: May 2011

I focus on two important forms of performance evaluation bias: centrality bias and leniency bias (Prendergast 1999). Centrality bias is the tendency to compress performance ratings, resulting in less variance in ratings than is warranted by variance in performance, while leniency bias is the tendency to inflate employees' performance ratings (Saal et al. 1980). Although numerous studies document the existence of both forms of bias (Kingsbury 1922; Kneeland 1929; Landy and Farr 1980; Jawahar and Williams 1997; Moers 2005), empirical evidence on the determinants and performance effects of managers' performance evaluation biases is scarce. This study extends the current literature by focusing on (1) what causes managers to display bias in performance ratings, and (2) how biased ratings influence employees' incentives.

To address the above questions, I examine the incentive plan of a financial service provider. In particular, I study the contract design and performance data of five branch offices for 2003 and 2004. I start my analysis by examining whether centrality bias and leniency bias are driven by managers' incentives and preferences. More specifically, I examine whether a factor based on economic theory—information-gathering costs—and a factor rooted in the behavioral view—the strength of the employee-manager relationships—increase centrality bias and leniency bias.¹ Second, I analyze how centrality bias and leniency bias affect employee incentives. Economic (agency) theory predicts that both biases decrease employee performance, as they negate the incentive effect of performance-based compensation contracts. In contrast, behavioral theory predicts that either bias can have a positive effect on employee motivation and performance if it enhances the perceived fairness of the incentive system.

The results show that managers respond to their own incentives and preferences when they subjectively evaluate performance. Specifically, findings indicate that information-gathering costs and strong employee-manager relationships positively affect centrality bias and leniency bias. The results also show that bias influences the incentives provided by performance-based compensation contracts. Contrary to predictions based on the agency perspective, the evidence indicates that managers' performance evaluation biases are not necessarily detrimental to compensation contracting. That is, although centrality bias negatively affects performance improvement, the evidence does not reveal a significant negative relation between leniency bias and performance. Rather, leniency bias positively affects performance improvement. This result is consistent with the behavioral argument that bias can positively influence incentives by enhancing the perceived fairness of an incentive system.

This study makes several contributions to the performance evaluation and compensation literature. First, this study answers several calls for empirical evidence on the effect of managers' personal rating incentives on rating behavior (e.g., Murphy and Cleveland 1991; Harris 1994). I show that the personal costs associated with rating employee performance decrease the accuracy of performance ratings. This finding indicates that subjective performance ratings are explained in part by managers' incentives and preferences. Yet, the discretion in the performance evaluation system allows managers to maintain that their evaluations were only based on employee performance. Consequently, in order to improve rating accuracy, research should focus not only on the rating ability of managers, but also on their rating incentives.

Second, to my knowledge, this is the first study to empirically investigate how biased performance ratings influence the effectiveness of performance-based incentive contracts. Prior studies establish the existence of centrality bias and leniency bias, but do not analyze the

¹ Since examining the effect of information-gathering costs on bias from the behavioral perspective would not lead to different hypotheses, and examining the effect of the strength of the employee-manager relationship on bias from an economic perspective would not lead to alternative predictions, I do not present the information-gathering hypotheses as exclusively economic hypotheses or the effects of the strengths of the employee-manager relationship as exclusively behavioral hypotheses.

consequences of these biases on future performance. My analysis shows that managers' performance evaluation biases are not necessarily detrimental for compensation contracting, as is assumed in previous literature (e.g., [Landy and Farr 1980](#); [Rynes et al. 2005](#)). Rather, I find that leniency bias positively affects employees' incentives consistent with the behavioral prediction that improved fairness perceptions can enhance the effectiveness of performance-based compensation contracts.

Finally, this study provides a detailed description of the salient features of an incentive system for non-executive employees, as well as both managers' and employees' reactions to this system. This analysis sheds light on how non-management employees are compensated, an area in which detailed information is generally scarce ([Indjejkian 1999](#)).

Section II reviews prior research on performance evaluation biases in compensation contracting and develops the study's testable hypotheses. The research setting and empirical design are presented in Section III. Section IV presents the results. Finally, Section V concludes and discusses directions for future research.

II. HYPOTHESES

Most studies that deal with subjectivity in compensation contracts focus on the determinants of subjectivity. These studies examine the role of subjectivity in incentive systems and show that introducing subjectivity improves contracts by mitigating incentive distortions or reducing risk ([Baker et al. 1994](#); [Baiman and Rajan 1995](#); [Hayes and Schaefer 2000](#); [Gibbs et al. 2004](#)). However, the introduction of manager discretion can also give rise to a number of problems, the most prominent of which is rating inaccuracy.²

A long line of research dating back to the 1920s examines rating accuracy in subjective performance appraisal (e.g., [Kingsbury 1922](#); [Kneeland 1929](#); [Thorndike 1949](#); [Barrett 1966](#); [Landy and Farr 1980](#)). These studies show that subjective performance ratings are susceptible to random error, systematic biases (e.g., halo and stringency), and bias due to race, sex, ethnicity, and other personal attributes of employees ([Feldman 1981](#)). Two systematic biases that have especially received a lot of attention in this stream of literature are centrality bias and leniency bias. Centrality bias refers to the tendency of managers to provide ratings that fail to adequately discriminate among subordinates in terms of their respective performance level ([Motowidlo and Borman 1977](#)). Leniency bias refers to the tendency of managers to provide their subordinates with higher ratings than is warranted by their performance ([Saal and Landy 1977](#)). For example, [Bretz et al. \(1992\)](#) document that while most organizations' evaluation systems comprise five levels, generally only three levels are used, with 60 to 70 percent of an organization's workforce rated in the top two performance levels. In a more formal investigation of leniency bias and centrality bias, [Moers \(2005\)](#) finds that, on average, performance ratings on the subjective dimension are higher and closer to the median rating than performance ratings on the objective dimension. This finding is consistent with the prediction that superiors provide more lenient and compressed ratings when subjectively assessing performance.

Determinants of Managers' Performance Evaluation Biases

Although there is abundant evidence that managers tend to apply biases in performance ratings, there is less evidence of what causes managers to display these biases. Most work that examines the

² Inaccurate performance ratings are clearly not the only type of potentially negative consequences of subjectivity in performance-based compensation contracts. Other concerns, such as uncertainty about measurement criteria, are also likely to influence the effectiveness of performance-based compensation contracts ([Bol 2008](#)).

determinants of performance evaluation biases focuses on managers' ability to rate accurately (Banks and Murphy 1985). For example, several research studies examine how rater training (Smith 1986; Woehr and Huffcutt 1994) and different types of rating formats (Landy and Farr 1980) influence rating accuracy. However, little attention has been given to examining how managers' personal incentives influence rating behavior. In particular, while several theoretical studies highlight the importance of rating incentives in performance evaluation (e.g., Holmstrom 1981; Murphy and Cleveland 1991; Harris 1994), virtually no empirical investigations explore this effect. The current study helps to fill this void by examining how managers' personal rating incentives affect their tendency to bias performance ratings.

Managers' Rating Incentives

Several analytical models in the agency literature analyze subjective performance evaluation performed by principals who are the residual claimants of the firm (e.g., Bull 1983; MacLeod and Malcomson 1989; Baker et al. 1994). These models show that residual claimants have incentives to under-report when subjectively assessing performance in order to keep compensation costs down. However, since most companies are multi-layered, the principals in most principal-agent relationships are not residual claimants (Prendergast 1999). As a result, their financial incentives to under-report are limited or nonexistent. However, this does not mean that principals have no incentive to use their discretion in the performance evaluation process to their own benefit. Principals (i.e., managers) are predicted to still take their own preferences into account when they appraise performance. That is, they have incentives to minimize the time and effort invested in the performance evaluation process, avoid confrontations, prevent damage to personal relationships, and limit criticism, and these preferences will influence the managers' ratings behavior (Harris 1994). Specifically, I hypothesize that they lead to centrality bias and leniency bias.

To make unbiased subjective assessments, managers need to invest time and effort in gathering information on employee performance. Monitoring employee behavior can be very costly to the manager, especially if the manager cannot observe the employee's actions (e.g., because of physical location) and thus must dedicate time exclusively to monitoring the employee. Since managers have a preference to limit their time and effort spent on performance evaluation, performance information will likely be incomplete when information-gathering costs are high (Higgins and Bargh 1987).

I argue that this lack of complete information leads to centrality bias. Since the probability that an employee is extremely good or bad is statistically low, managers are likely to compress ratings when they possess incomplete information, as compression increases the probability that their estimation based on imperfect information is close to the true performance level. This prediction is in line with Bernardin and Villanova (1986), who find that managers self-report that one of the reasons they rate inaccurately is lack of time spent on performance appraisal. It is also in line with Freeberg (1969), who demonstrates in a laboratory experiment that rating accuracy increases when raters are able to view more employee behavior that is *directly relevant* to the performance under evaluation. However, while these prior studies document the importance of time and effort spent on performance appraisal for rating accurately, they do not show that high information-gathering costs result in bias.

I also hypothesize that the lack of complete information leads to leniency bias. When employees are unsatisfied with their rating, they will likely ask their manager for a justification of the performance rating they have received. This process is time-consuming, and managers therefore have incentives to avoid these discussions (Levy and Williams 2004). A meeting to discuss performance ratings will be especially costly to the manager, when s/he does not have complete

information on the employees' performance. In such a situation, it will be even harder to explain and justify the provided performance rating, and the manager might even have to collect additional information on the employee's past performance.³ In order to avoid these discussions with employees who believe that their ratings are too low, managers will have a tendency to bias ratings upward, especially when they do not have complete information on the employee's performance (Friedrich 1993).⁴ This argument is consistent with Bernardin and Villanova (1986), who find that the majority of surveyed managers indicate that the desire to avoid discussions with subordinates is a significant source of rating inaccuracy. In a similar vein, Bernardin et al. (2000) find that individuals give more lenient ratings when they have a strong desire to avoid discussions. However, these studies have not linked higher information-gathering costs to increased cost of providing justifications and leniency bias.

Evaluating performance can also put a psychological burden on managers (Villanova et al. 1993). In particular, communicating harsh, but accurate performance ratings can damage personal relationships and lead to criticism, in addition to time-consuming and uncomfortable confrontations (Napier and Latham 1986). I predict that managers anticipate these negative reactions when rating performance and consequently offer compressed and lenient ratings. In doing so, managers are likely to feel that they are being kind to their employees and as a result expect to be sheltered from criticism (Barrett 1966).⁵ Thus, the tendency to give lenient and compressed ratings is a defensive mechanism applied to avoid the ramifications of deserved, but tough performance ratings.

Prior survey research finds evidence indicating that managers' rating behavior is influenced by personal relationships and fear of criticism. For example, Lawler (1990) finds that managers are concerned that making accurate performance ratings will hurt their relationships with subordinates. Similarly, Murphy and Cleveland (1991) suggest that the possibility of employee criticism is a frequently voiced concern of managers. The present study extends this research by using archival data to provide empirical evidence that stronger employee-manager relationships lead to increased centrality bias and leniency bias.

Hypotheses

Taken together, the above discussion predicts that managers bias performance ratings to reduce the personal costs of evaluating employee performance. Specifically, I hypothesize that the cost of gathering information and the psychological burden of confronting employees with whom the manager has a personal relationship, lead to centrality bias and leniency bias in performance ratings. More formally:

H1a: Information-gathering costs positively affect centrality bias.

H1b: Information-gathering costs positively affect leniency bias.

H1c: Strong employee-manager relationships positively affect centrality bias.

H1d: Strong employee-manager relationships positively affect leniency bias.

³ This argument is consistent with Mero et al. (2003). They show that, in a laboratory setting, participants who are asked to justify their ratings collect more complete employee performance information.

⁴ Consistent with the situation of the company used in this study, I assume there is no set limit to the amount of compensation that managers can award. Hence, there is no bonus pool system where increasing one employee's compensation means lowering another employee's compensation.

⁵ Although commonly assumed, it is not necessarily true that lenient and compressed ratings protect managers from criticism—e.g., exceptional high performers might complain about being buried among the less talented (Barrett 1966).

Employee Incentives and Managers' Performance Evaluation Biases

Although, to my knowledge, no prior research empirically investigates the performance implications of performance evaluation biases, numerous studies examine how bias can be reduced by training managers in performance evaluation (Smith 1986; Woehr and Huffcutt 1994) or improving rating scales (Landy and Farr 1980). These studies explicitly or implicitly assume that biases are detrimental *per se* for performance evaluation (Landy and Farr 1980; Rynes et al. 2005). To assess the validity of this assumption, I study whether managers' evaluation biases are negative for compensation contracting.

I start by using an agency perspective to analyze the effect of bias on employee incentives.⁶ Specifically, I examine how the influence of centrality bias and leniency bias on the link between pay and performance affects employee incentives. However, given the behavioral perspective that the effectiveness of an incentive system is not necessarily determined only by its pay-for-performance sensitivity, I also analyze the effect of bias on employee incentives from a behavioral perspective. Specifically, I analyze how centrality bias and leniency bias influence the perceived fairness of the compensation plan and, consequently, employees' future incentives.

Link between Pay and Performance

One of the main objectives of using a performance-based compensation system is to motivate employees to exert effort (Rees 1985).⁷ According to agency theory, linking pay to performance motivates employees to exert greater effort to improve performance because increased performance results in increased pay (Holmstrom 1979; Shavell 1979). However, employees will be motivated to increase effort only when they expect improved performance to translate into more compensation (Holmstrom and Milgrom 1991). When managers assess performance subjectively, this translation is likely impacted, as managers have both motive and opportunity to bias ratings and biases disrupt the link between pay and performance (Prendergast 1999). Therefore, I predict that performance evaluation biases influence the incentive effect of the performance-based compensation plan.

Centrality bias creates a disproportionate pay-to-performance ratio. Since compression is achieved by deflating the rating of above-average performers and inflating the rating of below-average performers, above-average performers have to provide a larger performance increase to get the same rating increase as below-average performers. This ratings compression negatively affects incentives because it reduces the likelihood that above-average performers will believe that the value of a marginal rating increase (i.e., additional compensation) outweighs the costs of improving performance sufficiently to receive this increase.

The inflated ratings of below-average performers also provide little incentive to improve performance. Although the performance rating difference between below- and above-average performers is relatively small (because of the compression), the real performance increase needed to move from below to above average is relatively large (Cichello et al. 2009). In order to move to an above-average rating, employees would have to exert sufficient effort to overcome the switch from inflated to deflated performance assessments. Hence, moving from below- to above-average ratings will take a considerable real performance increase, while the additional compensation resulting

⁶ In this study I assume that the levels of centrality bias and leniency bias are not optimal equilibrium choices. I work under the assumption that managers are on the path toward equilibrium and that it is therefore possible to observe cross-sectional differences in performance (Hogarth and Reder 1987; Camerer et al. 2004). This is not an unrealistic assumption in this setting as the performance-based compensation system was first introduced in 2003 (my first year of analysis).

⁷ Another important objective of performance measurement is to differentiate between highly skilled and less skilled employees. Although the effect of less-than-optimal personnel decisions might be severe, investigating the effect of bias on selection issues is beyond the scope of this study.

from the rating increase will be relatively small, making it less likely that the benefits will outweigh the costs. In sum, compressed ratings reduce incentives for both above- and below-average performers.

In the case of leniency bias, each level of performance receives a higher rating (and thus more compensation). This leniency negatively affects incentives, because the employee has to exert less effort to reach the same level of compensation. The effect of leniency bias is comparable to lowering the performance targets, which reduces employee incentives (Blanchard et al. 1986). Note that this argument is only valid if the total compensation level is not fixed (as at the company investigated in this study). When evaluations are all biased upward, but the total compensation level is fixed (e.g., through the use of a budget or bonus pool), leniency bias will have no effect on compensation and, consequently, employee effort will not be affected.

Employee Perception of Fairness

Behavioral research shows that the perceived fairness of performance-based compensation plans has a significant impact on employee incentives (Akerlof and Yellen 1988; Blinder and Choi 1990; Colquitt et al. 2001). Contrary to predictions based on the agency perspective, behavioral research argues that, in order to be motivated, employees care not only about how much compensation they receive, but also about how their performance rating compares to their expectations and to the ratings of others. If an employee perceives his/her rating to be unfair, then he/she will react negatively toward the compensation system and not be motivated by it (Greenberg 1990; Cohen-Charash and Spector 2001; Colquitt et al. 2001). Performance evaluation biases influence the perceived fairness of the compensation system because biases change the outcome distribution of the compensation plan.

The effect of centrality bias on perceived fairness likely depends on whether the employee's performance is above or below average. Above-average performers become less motivated when they receive roughly the same compensation as employees who do not perform as well. As predicted by equity theory (Adams 1963), individuals compare their own reward-to-input ratio to the corresponding ratios of their peers. If the ratios are unequal, then the party whose ratio is lower will feel upset and dissatisfied. This negatively influences incentives, as individuals will lower their performance in order to restore a feeling of equity (Garland 1973).

The situation is different for below-average performers, as compression positively influences their ratings. Since most employees believe that their own performance is above-average (Beer and Gery 1972; Meyer 1975), providing below-average employees with their true comparative rating could be counterproductive in terms of motivation and performance (Pearce and Porter 1986). Although compression does not result in below-average performers receiving the above-average rating they think they deserve, it likely has a more positive effect on perceived fairness than uncompressed ratings. Hence, the compressed ratings provide below-average performers with a rating that is more similar than an uncompressed rating to their own assessment of their relative performance. Consequently, below-average performers will consider compressed ratings to be less unfair and will react less negatively.

The behavioral perspective expects leniency bias to positively affect perceived fairness by increasing the congruence between the rating the employee thinks s/he deserves and the rating s/he actually receives. Individuals have a tendency to over-estimate their abilities relative to their supervisors (McFarlane Shore and Thornton 1986; Harris and Schaubroeck 1988).⁸ As a result,

⁸ Abundant evidence exists in the psychology literature that most people over-estimate their abilities and past achievements (Arkin et al. 1980), they tend to recall their successes more than their failures (Mischel et al. 1976), and they have the tendency to be unrealistically optimistic about their future (Weinstein 1980).

employees are likely to perceive a non-inflated rating as unfair. When an employee believes that his/her performance warrants a higher rating than the rating he/she received, the employee will have a negative reaction against the compensation system, which weakens incentives (Akerlof and Yellen 1988; Colquitt et al. 2001). Thus, since lenient ratings are more in line with the expectations of self-over-estimating employees, leniency bias can increase perceived fairness and strengthen incentives.

Hypotheses

Based on the above discussion, I distinguish two primary mechanisms through which managers' performance evaluation bias can influence employee incentives: (1) through its effect on the link between pay and performance, and (2) through its effect on perceived fairness.

For centrality bias, the agency perspective predicts a negative incentive effect, as it creates a disproportionate pay-to-performance ratio. Consistent with the agency perspective, the behavioral view predicts that centrality bias negatively affects the incentives of above-average performers because it creates unequal reward-to-input ratios between peers. However, for below-average performers, the behavioral view predicts that centrality bias has a positive effect on incentives because it enhances the perceived fairness of the compensation system. Hence, for below-average performers, the theoretical predictions for centrality bias lead to a composite hypothesis (which I state in null form).⁹ Separating the hypotheses for above- and below-average performers, I formally state:

H2a: Centrality bias negatively affects the performance incentives of above-average performers.

H2b: Centrality bias does not affect the performance incentives of below-average performers.

According to the agency perspective, leniency bias negatively influences employee incentives, as it weakens the link between pay and performance. The behavioral perspective, in contrast, predicts that lenient ratings improve perceived fairness and strengthen incentives. This again leads to a composite hypothesis, stated in null form:

H2c: Leniency bias does not affect employee performance incentives.

III. RESEARCH DESIGN

Research Setting

To test the hypotheses, I use data from one of the main financial service providers (FSP) in The Netherlands.¹⁰ FSP serves over nine million customers, with a total asset value of approximately €475 billion in 2005, the last year of the sample period.

As part of an initiative to move to more results-oriented management, FSP introduced a new incentive system for its branch offices in 2003. The new performance-based compensation system consists of a fixed salary and a bonus of up to 15 percent of the fixed salary. The size of the bonus

⁹ Although stating hypotheses in the null form is less informative, I use the null form because it was unclear *a priori* whether the agency or the behavioral perspective would be predictive of the examined relationships.

¹⁰ The Netherlands is, in many respects, very similar to the U.S. when it comes to incentive contracting (see Abernethy et al. 2004; Indjejikian and Matejka 2006; Bouwens and Van Lent 2007). One noted difference, however, is the less prominent use of performance-based compensation systems in The Netherlands, which, according to Jansen et al. (2009), is explained by the relatively lower score on Hofstede's masculinity dimension (Dutch people have a lower preference for competitiveness, achievement, and material success).

and a potential fixed salary increase are determined by the employee's end-of-year overall performance rating, which is the weighted average of an output-based score and a competence-based score.¹¹ The output score is based on three to six equally weighted measures that capture employee output. Examples are "consumption rate," "market share," and "cross-selling ratio" (see Figure 1). The competence score is also based on three to six equally weighted measures. The competencies of "cooperative behavior" and "customer focus" are standard for all, while the manager chooses the other measures to capture the behavioral attributes that are essential to the employee's tasks.

Sample

The analyses employ two years of proprietary archival data on the incentive system of five branch offices of FSP.¹² I hand-collect completed performance documents for all employees of the five branches. The analyses are limited to those individuals who are employed by FSP in both 2003 and 2004 and to departments that make significant use of objective measures (at least 25 percent), where the classification of objective versus subjective measures is discussed below.¹³ The managers in the sample range from top management to middle management (not including the general director of the branch), while the included subordinates vary from senior private banker to assistant account managers.¹⁴ Lower-level positions such as tellers and custodians are not included in the sample. The sample-selection procedure results in a total of 396 (198 per year) completed performance documents.

The performance documents provide information on the employee's position, department, and manager. This information allows me to identify reference groups—i.e., employees who work together, perform similar tasks, and are evaluated by the same manager. There are 69 reference groups, the average number of employees in a reference group is 6, and each manager oversees an average of 1.2 reference groups.

The performance documents also indicate the performance measures, performance ratings, and the old and new compensation. Besides determining the performance measures,¹⁵ managers must provide an extensive description on the performance document of how they are going to measure performance when performance is quantifiable. They are also required to indicate the target for each performance level for these quantitative measures. I classify output-based measures as objective or subjective, based on this information. An output-based measure is only classified as objective if the document states the formula that the manager uses to quantify output. It is important to note that the classification is based on how the manager chooses to evaluate performance (i.e., measure objectively or assess subjectively). The classification is not based on whether a dimension *could* be objectively measured with the right measurement system.¹⁶ Because the HR department collects the

¹¹ Figure 1 illustrates the relation between the overall performance rating and the bonus percentage.

¹² The five offices provide a representative mix of regions, office size, and area type (ranging from rural to urban).

¹³ Using employees who were employed by FSP in both 2003 and 2004 might lead to selection bias because employees who left the company in 2003 are excluded from the sample. To examine this issue, I asked the HR managers to identify cases where employees had left the company because of the new system. Only a few isolated cases at one branch were identified.

¹⁴ The average salary scale of the managers is 8.4 (range 5 to 12), while the average salary scale of the employees is 6.1 (range 3 to 11).

¹⁵ To facilitate consistent use of the new system, the designers provide managers with a set of sample performance measures for each position. Managers, however, determine their employees' performance measure and targets independently and are free to use non-sample performance measures.

¹⁶ In Figure 1, the first four output-based measures are classified as objective as the document clearly states how the output will be measured and what the specific targets are. The last measure, "networking," is classified as subjective because no details on measurement are provided.

FIGURE 1
The Performance Measurement Document

PERFORMANCE DOCUMENT					
Name: Henk ten Broek Supervisor: Els Jansen			Department: Banking Services Function: Private banker		
Output performance measures					
	Evaluation				
	Bad	Regular	Good	Very Good	Excellent
1) Consumption rate N of services sold /n of active clients	O 2.4	O 2.5	O 2.6	O 2.7	O 2.8
2) Market share - mortgage (N of active mortgage clients*100%)/n of potential clients	O 24%	O 28%	O 32%	O 36%	O 40%
3) Cross-selling ratio (N of sold home insurances*100%)/n of sold mortgages	O 15%	O 20%	O 25%	O 30%	O 35%
4) Renewal rate private loans (N of renewed loans*100%)/n of renewable loans	O 60%	O 62.5%	O 65%	O 67.5%	O 70%
5) Networking The individual's involvement in network activities	O	O	O	O	O
6)	O	O	O	O	O
Competence performance measures					
	Evaluation				
	Bad	Regular	Good	Very Good	Excellent
1) Cooperative behavior	O	O	O	O	O
2) Customer focus	O	O	O	O	O
3) Business development skills	O	O	O	O	O
4) Organization skills	O	O	O	O	O
5) Initiative	O	O	O	O	O
6)	O	O	O	O	O
Total rating output performance measures			50%		
Total rating competence performance measures			50%		
Total score					
Salary Determination					
Bad	Regular	Good	Very Good	Excellent	
0%	1-3%	4-8%	9-12%	13-15%	
Current Salary	Salary Scale	Growth %	Bonus		
New Salary	Salary Scale	Growth %	Bonus		

performance documents and assesses whether they are used in accordance with the guidelines, it is unlikely that the managers used formula-based criteria and did not report them.¹⁷ The competence measures are subjective by definition.

I also use information from a survey administered by FSP's compensation system designers. In August 2004, FSP surveyed a set of randomly selected employees from different branches to examine employee and manager reactions to the new system. In total, 161 employees returned and answered all questions of the survey (the performance information portion was optional). Of these respondents, only 55 qualify for my analysis because I require information on how performance is evaluated—objectively measured or subjectively assessed—in order to distinguish objective from subjective output-based measures. Unfortunately, FSP did not include this information in the survey. To avoid mistakenly labeling any measure as objective, I include only respondents who indicate in one of the survey questions that their output-based measures were very quantifiable. Because of anonymity, I am also unable to link the survey responses to the archival data.

Finally, I collect demographic information and interview several employees, managers, and compensation system designers to gain a better understanding of the company.

Variables

To examine the determinants of centrality bias and leniency bias, I measure the extent of bias applied to the ratings. Recall that centrality bias refers to the tendency of managers to provide ratings that fail to adequately discriminate between subordinates in terms of their respective performance level. Since I am interested in how much managers discriminate between *their* employees when subjectively evaluating performance (the amount of compression), I compare the compression in the subjective rating with the compression in the objective rating at the reference group level (i.e., subordinates who work together and have similar positions). I use reference groups instead of all employees evaluated by a specific manager, as employees are likely to compare themselves with colleagues in a similar position (Adams 1963; Meyer 1975).¹⁸ Specifically, I calculate the ratio between the standard deviation of the objective ratings and the standard deviation of the subjective ratings of all employees in the reference group (*CEN_BIAS*). I capture leniency bias, the tendency of managers to provide their subordinates with higher ratings than is warranted by their performance, by comparing the manager's subjective performance assessments with the formula-based objective performance score. Specifically, I subtract the objective rating from the subjective rating (*LEN_BIAS*).¹⁹

The primary independent variables of interest are information-gathering costs and the strength of the employee-manager relationship. At FSP, some managers share their direct workspaces with their subordinates, while others work in close proximity but are unable to observe their subordinates directly. Moreover, some managers' daily tasks overlap to a large extent with those of their subordinates, while other managers never perform tasks similar to their subordinates' tasks. These factors influence the costs of gathering employee performance information. Specifically, for managers who share a workspace and overlap in duties, it takes relatively little time and effort to monitor their subordinates' performance. To capture these relatively lower information-gathering

¹⁷ Interviews with managers and HR personnel confirm that managers put significant effort into completing the performance documents.

¹⁸ During my interviews, employees confirmed that they had some, though not full, knowledge of their peers' ratings. Previous survey research also shows that employee knowledge on the rating distribution is a common characteristic of incentive systems (Bernardin and Villanova 1986).

¹⁹ Results are robust to alternative specifications of leniency bias such as $\text{Max}(\text{subjective rating} - \text{objective rating}, 0)$, $(\text{subjective rating}/\text{objective rating})$, and using the values of a residual model that controls for contract design and centrality bias.

costs, I create a binary variable that codes managers as low in information-gathering costs if they share a workspace and overlap in duties with their subordinates, and high if they do not (*INFO_C*).²⁰ To make this classification, I examine the workspace layout at each branch office and compare the managers' job descriptions with their subordinates' job descriptions. The classification is double-checked by the HR manager of each branch and, in case of disagreement, we re-evaluate the case and reconcile the disagreement.

I use three proxies to capture the strength of the employee-manager relationship. First, the strength of the relationship likely depends on the amount of time that the employee and manager have worked together. When the employee-manager relationship is relatively new, a close relationship between the two parties is less likely. To capture new employee-manager relationships, I include a binary variable that equals 1 if the employee or manager recently joined the company (*NEW_R*). Second, based on research that shows that individuals with similar demographic characteristics are more likely to develop a close relationship (Tsui and O'Reilly 1989), I include the employee-manager age difference (*DIF_AGE*) and a binary variable that indicates a gender difference between the manager and employee (*DIF_GEN*).

Finally, I measure each branch office's profit growth (*GROWTH*). Since financial resources are limited, managers will feel some pressure to keep compensation costs down, especially during economic downturns (Longenecker et al. 1987). By capturing the branch-level financial performance, I control for variation in this pressure. When examining the determinants of centrality bias, I also control for the number of employees in the reference group (*NR_REF*).²¹

To examine the performance effects of performance evaluation biases, I perform two separate analyses: one based on the archival data and one based on the survey data. For the archival-based analysis, the dependent variable performance improvement ($\Delta PERF$) is measured as the difference between the employee's objective (subjective and total) performance rating in 2003 and 2004. For the survey-based analysis, the dependent variable is employees' self-reported change in effort. More specifically, I proxy for performance improvement by using the employee's own assessment of how his/her effort is affected by the new incentive system (measured on a five-point scale, *EFFORT*). In both analyses, the main independent variables, leniency bias (*LEN_BIAS*) and centrality bias (*CEN_BIAS*), are measured as discussed above.²²

As control variables, I include employee characteristics that are predicted to make an employee less sensitive to the incentives provided by the new system. I include a variable that indicates whether the employee has reached his/her salary scale limit (*MAX_SC*), as performance improvements beyond this point will only affect the bonus. Since part-time and older employees have differential career concerns, I include variables indicating part-time employees (*PART_T*) and employees older than 50 (*OLDER*), respectively. In the archival-based analysis, I also include a variable that indicates a job change (*DIF_JOB*) and variables that control for changes made to the design of the performance document (see Table 5).

²⁰ Note that *INFO_C* does not refer to the seniority of the manager. Even more strongly, when comparing the salary scale of the managers (a proxy for seniority) in the low and high information-gathering costs groups, I find no significant difference in seniority. The mean scale for the managers when *INFO_C* is 0 (1) is 8.7 (8.2). *INFO_C* is also not picking up the difference in seniority between the manager and the employee. The mean difference in salary scale is 2.69 (2.68) when *INFO_C* is 0 (1).

²¹ The introduction of additional variables that control for contract choices (e.g., the contractual weight placed on the subjective measures and the number of subjective measures used), seniority, or department leads to similar inferences (untabulated).

²² The survey-based centrality bias proxy must be interpreted in light of the following limitations. Since FSP surveyed a random sample of employees, the centrality bias proxy is based on a randomly selected subset of the reference group. Moreover, since the participants were not asked to name their manager in the survey, I have to assume that all employees within a certain department of a specific branch were evaluated by the same manager.

IV. RESULTS

Summary statistics for the main variables indicate that the variance in the objective ratings is significantly larger than the variance in the subjective ratings in both 2003 and 2004 (see Tables 1 and 2). This is true for the variance in the entire sample (SD is 0.76 versus 0.46 in 2003 and 0.74 versus 0.45 in 2004) as well as for the average variance per reference group (mean SD is 0.48 versus 0.31 in 2003 and 0.63 versus 0.37 in 2004).

The descriptive statistics also show higher ratings on the subjective versus objective performance measures in 2003. Both the mean (2.94 versus 2.80) and the median (3.00 versus 2.75) are higher. Pairwise observations show that the subjective rating exceeds the objective rating for 59 percent of the employees in 2003. However, in 2004, this relationship reverses, as the mean (3.14

TABLE 1
Descriptive Statistics

	2003					2004				
	Mean	SD	Median	Range	n	Mean	SD	Median	Range	n
Overall performance rating	2.87	0.52	2.86	1.6–4.7	198	3.20	0.50	3.23	1.6–4.4	198
Objective performance rating ^a	2.80	0.76	2.75	1.0–5.0	198	3.28	0.74	3.33	1.0–5.0	198
Subjective performance rating ^a	2.94	0.46	3.00	1.5–4.3	198	3.14	0.45	3.13	2.1–4.6	198
Total # of performance measures	9.70	1.77	10.00	5–12	198	10.68	1.49	11.00	5–12	198
# of objective performance measures	4.04	1.35	4.00	1–6	198	4.48	1.30	5.00	2–6	198
# of subjective performance measures	5.66	1.36	6.00	2–9	198	6.20	1.06	6.00	3–8	198
% of total rating objectively determined	44.33	8.72	50.00	25–50	198	44.37	7.57	50.00	25–50	198
% of total rating subjectively determined	55.67	8.72	50.00	50–75	198	55.63	7.57	50.00	50–75	198
Employee age	38.39	8.60	37	23–60	198	39.39	8.60	38	24–61	198
Contract hours	32.68	6.52	36	12–40	198	32.27	6.80	36	12–40	198
Tenure	11.40	9.50	7.00	1–38	198	12.40	9.50	8.00	2–39	198
Profit growth ^b	1.09	2.62	1.60	–6.2–11.9	198	5.68	18.21	1.27	–20–67	198
Size reference group	5.35	2.41	5.00	2–11	37	6.19	3.02	5.00	2–11	32
SD objective rating per reference group ^c	0.48	0.27	0.49	0.0–0.9	37	0.63	0.30	0.58	0.3–1.2	32
SD subjective rating per reference group ^c	0.31	0.17	0.29	0.1–0.7	37	0.37	0.20	0.31	0.1–1.0	32

^a t-tests for mean-comparison indicate significant differences between the mean objective and subjective performance ratings in both 2003 and 2004 ($p < 0.01$ in both cases).

^b The percentage difference between each branch office's profit in the current year and the branch office's profit in the previous year.

^c t-tests for mean-comparison indicate significant differences between the mean SD in the objective and subjective performance ratings per reference group in both 2003 and 2004 ($p < 0.01$ in both cases).

TABLE 2
Correlation between Variables

Panel A: 198 Observations for 2003

	1	2	3	4	5	6	7	8	9	10
1. Overall performance rating										
2. Objective performance rating	0.90***									
3. Subjective performance rating	0.80***	0.47***								
4. Total # of perf. measures	-0.14**	-0.11	-0.13*							
5. # of objective perf. measures	-0.12*	-0.17**	-0.02	0.65***						
6. # of subjective perf. measures	-0.06	0.03	-0.15**	0.66***	-0.15**					
7. % objectively determined	-0.10	-0.16**	0.02	0.10	0.52***	-0.39***				
8. Employee age	-0.02	-0.03	-0.03	-0.04	-0.02	-0.03	-0.03			
9. Work hours	0.18**	0.19***	0.12*	0.30***	0.13*	0.26***	0.04	-0.04		
10. Tenure	-0.03	-0.05	-0.03	-0.02	0.01	-0.03	0.09	0.73***	0.02	
11. Profit growth	-0.16**	-0.15**	-0.12	-0.12*	-0.05	-0.10	-0.01	0.04	-0.17**	0.11

Panel B: 198 Observations for 2004

	1	2	3	4	5	6	7	8	9	10
1. Overall performance rating										
2. Objective performance rating	0.88***									
3. Subjective performance rating	0.78***	0.41***								
4. Total # of perf. measures	-0.08	-0.08	-0.05							

(continued on next page)

TABLE 2 (continued)

	1	2	3	4	5	6	7	8	9	10
5. # of objective perf. measures	0.02	-0.03	0.07	0.72***						
6. # of subjective perf. measures	-0.14**	-0.08	-0.16**	0.53***	-0.21***					
7. % objectively determined	0.18***	0.10	0.19***	-0.05	0.57***	-0.77***				
8. Employee age	-0.22***	-0.25***	-0.09	0.04	-0.02	0.08	-0.06			
9. Work hours	0.14**	0.15**	0.11	0.10	0.08	0.04	-0.01	0.00		
10. Tenure	-0.13*	-0.18***	-0.02	0.06	0.01	0.07	-0.04	0.73***	0.06	
11. Profit growth	-0.08	-0.09	-0.04	-0.28***	-0.19***	-0.16**	-0.10	-0.01	0.08	-0.01

*, **, *** Denote statistical significance at the 10 percent, 5 percent, and 1 percent levels (two-tailed), respectively.

versus 3.28) and median (3.13 versus 3.33) subjective ratings are lower than the objective ratings, and in only 40 percent of the pairwise observations is the subjective rating higher.²³

Consistent with Moers (2005), I investigate whether the centrality bias and leniency bias patterns hold after controlling for contract design and the characteristics of employees and managers. To examine compression, I use the ratio between the employee's overall objective (subjective) rating and the mean objective (subjective) rating of the reference group ($R_RATING = \text{Max}(\text{rating}/\text{mean rating}, \text{mean rating}/\text{rating})$) as the dependent variable. The main independent variable is an indicator that equals 1 if the rating is subjective ($I_{Subjectivity}$). To control for differences in contract design that can cause different effort allocations, I include the number of performance measures that the rating comprises (NR_PM) and the total weight placed on these measures ($WEIGHT$) (see Moers 2005). To control for the fact that compression might be driven by a certain type of employee and/or manager, I include several employee characteristics (age [AGE], gender [GEN], and part-time employment [$PART_T$]) and manager indicator variables ($I_{Manager}$). Year (Y_{2003}) and office indicator variables (I_{Office}) are also included.

Column (1) of Table 3 presents the estimation results of the OLS regression analysis with robust standard errors clustered by employee. Consistent with prior studies (e.g., Moers 2005), the significant negative coefficient on $I_{Subjectivity}$ indicates that the subjective performance ratings are more compressed than the objective ratings.

To investigate the presence of leniency bias, I use the employees' overall objective and subjective ratings ($RATING$) as the dependent variable and $I_{Subjectivity}$ as the main independent

²³ The cumulative distribution functions of the subjective and objective ratings indicate that neither rating has a first-order stochastic dominance over the other rating.

variable. The control variables are as specified above (for more details on specification and variable definition see Table 3).

Column (2) of Table 3 presents the results of the OLS regression analysis with robust standard errors clustered by employee. The coefficient of $I_{Subjectivity}$ is positive and significant at the 5 percent level, indicating that the subjective ratings are higher than the objective ratings, after controlling for relevant influences. Since the descriptive statistics show different patterns across the two sample years, I also perform annual analyses and find that, on average, the subjective ratings are not higher in 2004 (see Column (4) of Table 3).²⁴

To investigate what might explain this apparent absence of leniency bias in 2004, I first identify that the 2004 ratings are higher, on average, than the 2003 ratings. These higher ratings are likely the result of substantially higher growth in the Dutch GDP in 2004 compared to 2003 (2.2 percent versus 0.3 percent). Consistent with this macroeconomic factor, actual profit growth exceeded budgeted growth in four of the five branches in 2004. This economic growth resulted in higher individual sales figures (frequently used objective measures) than anticipated. Consistent with this argument, I find a significant correlation between the extent to which branches surpass their budgeted growth and the objective performance scores of the employees of that particular branch ($\rho = 0.14$ and $p = 0.01$).²⁵

Since supervisors have incentives to avoid negative employee reactions that often accompany low performance ratings (Latham 1986; Shore and Tashchian 2002), I examine whether leniency bias is more prominent for those employees whose objective ratings are relatively low. I split the sample into two groups, one with objective ratings above the average of the rating scale (3), and one below. I find that the subjective ratings are significantly higher than the objective ratings in the below-average group in 2003 and 2004, for both the mean rating (2.76 versus 2.20 in 2003 and 2.97 versus 2.35 in 2004) and the median rating (2.67 versus 2.33 in 2003 and 2.88 versus 2.50 in 2004).²⁶ Pairwise observations show that the subjective rating exceeds the objective rating for 86 percent of the employees in 2003 and for 87 percent of the employees in 2004.

The opposite pattern emerges in the above-average group. The subjective ratings are significantly lower than the objective ratings for both the mean (3.17 versus 3.64 in 2003 and 3.25 versus 3.75 in 2004) and the median (3.14 versus 3.60 in 2003 and 3.21 versus 3.67 in 2004). Pairwise observations show that the subjective rating exceeds the objective rating for 12 percent of the employees in 2003 and for 17 percent of the employees in 2004. These patterns continue to hold after controlling for contract design and employee and manager characteristics, as discussed above (see Columns (5) and (6) of Table 3). Hence, the overall results indicate that managers tend to engage in leniency bias and, consistent with managers wanting to avoid confrontations with low performers, this tendency is stronger when the objective measure is below the average of the rating

²⁴ The observed pattern (59 percent higher subjective ratings in 2003 followed by 40 percent higher subjective ratings in 2004) could be the result of mean reversion. To examine this possibility, I regress the difference between the subjective and objective ratings in 2004 on the difference between the subjective and objective rating in 2003 ($SUBJ_R_{2004} - OBJ_R_{2004} = \alpha + \beta(SUBJ_R_{2003} - OBJ_R_{2003}) + \varepsilon$). A negative coefficient indicates mean reversion, while a positive coefficient indicates consistency in the sign of the difference between the subjective and objective rating over the years. The results show a positive significant coefficient of 0.31 ($p < 0.01$), indicating that mean reversion does not explain the pattern found.

²⁵ The survey data also indicate that the compensation system had an overall positive influence on effort and personal development. This positive incentive effect was corroborated in the interviews.

²⁶ t-tests for mean-comparison indicate significant differences between the mean objective and subjective performance rating for the below and above average of the rating scale group in both 2003 and 2004 ($p < 0.01$ in all cases).

TABLE 3
The Impact of Manager Subjectivity on Employee Performance Ratings

$$R_RATING_{it} = \alpha_0 + \alpha_1 I_{Sub} + \alpha_2 WEIGHT_{it} + \alpha_3 NR_PM_{it} + \alpha_4 AGE_{it} + \alpha_5 GEN_i + \alpha_6 PART_T_{it} + \sum_{j=7}^{46} \alpha_j I_{Manager_j} + \sum_{k=47}^{50} \alpha_k I_{Office_k} + \alpha_{51} Y_{2003t} + \varepsilon_{it} \quad (1)$$

$$RATING_{it} = \alpha_0 + \alpha_1 I_{Sub} + \alpha_2 WEIGHT_{it} + \alpha_3 NR_PM_{it} + \alpha_4 AGE_{it} + \alpha_5 GEN_i + \alpha_6 PART_T_{it} + \sum_{j=7}^{46} \alpha_j I_{Manager_j} + \sum_{k=47}^{50} \alpha_k I_{Office_k} + \alpha_{51} Y_{2003t} + \varepsilon_{it} \quad (2)$$

Independent Variables ^a	<i>R_RATING</i> (1)	<i>RATING</i> (2)	<i>RATING</i> 2003 (3)	<i>RATING</i> 2004 (4)	<i>RATING</i> Below-2003 (5)	<i>RATING</i> Below-2004 (6)
<i>I_Subjectivity</i>	-0.10*** (-7.29)	0.10** (2.01)	0.29*** (4.47)	-0.10 (-1.46)	0.66*** (8.88)	0.60*** (5.79)
<i>NR_PM</i>	-0.01 (-0.15)	-0.06** (-2.36)	-0.07** (-2.40)	-0.06 (-1.01)	-0.02 (-0.63)	0.07 (1.42)
<i>WEIGHT</i>	0.00 (1.21)	0.00 (0.28)	-0.00 (-0.72)	0.00 (0.74)	-0.01** (-1.98)	-0.00 (-1.06)
<i>AGE</i>	0.01** (1.97)	-0.01*** (-2.54)	-0.01 (-1.32)	-0.01*** (-2.83)	-0.01 (-0.10)	-0.01 (-0.76)
<i>GEN</i>	0.04* (1.63)	-0.12 (-1.30)	-0.09 (-1.21)	-0.16 (-1.23)	-0.18** (-1.97)	-0.33* (-1.69)
<i>PART_T</i>	0.03 (1.02)	-0.10 (-1.10)	-0.10 (-1.17)	-0.12 (-0.97)	-0.21*** (-2.52)	-0.17 (-0.80)
<i>Y₂₀₀₃</i>	0.01 (0.56)	-0.40*** (-9.50)				
Adj. R ²	0.18	0.29	0.40	0.24	0.59	0.56

*, **, *** Indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels (two-tailed), respectively. t-values are in parentheses, *i* = ratings, *t* = time, *j* = managers, *k* = offices, and *n* = 792 for Columns (1) and (2), *n* = 396 for Columns (3) and (4), *n* = 206 for Column (5), and *n* = 110 for Column (6).

^a An intercept, as well as manager and office indicator variables are included but not reported.

Variable Definitions:

R_RATING = performance rating variation, measured as the ratio between employees' individual objective (subjective) performance rating, and the mean objective (subjective) performance rating of the reference group: Max ([rating/mean rating], [mean rating/rating]);

RATING = subjective or objective performance rating;

I_Subjectivity = indicator variable that equals 1 if the observation refers to a subjective performance rating, and 0 otherwise;

NR_PM = number of objective (subjective) performance measures;

WEIGHT = contractual incentive weight on the objective (subjective) performance measures;

AGE = employee age;

PART_T = indicator variable that equals 1 if the employee works part-time, and 0 otherwise;

GEN = employee gender (0 for female and 1 for male); and

Y₂₀₀₃ = indicator variable that equals 1 if the observation relates to 2003, and 0 otherwise.

scale. Thus, the higher objective ratings in 2004 (only 28 percent received a rating under the average of the rating scale) explain why the ratings are less subject to leniency bias in 2004.²⁷

In sum, the results indicate that managerial discretion leads to centrality bias, because the subjective ratings are significantly more compressed than the objective ratings. The evidence also indicates a pattern consistent with leniency bias, especially for those employees who perform below the average of the rating scale on the objective dimension.

Note that I draw these inferences in light of the following assumptions. First, I assume that the true variance in performance (without any performance evaluation bias) is, on average, similar for the objectively measured and the subjectively assessed elements, and that the variance in the objective performance ratings is, on average, similar to the true performance variance in the sample. In a similar respect, I assume that the objective ratings are, on average, unbiased or at least significantly closer to the true performance value than the subjective ratings. I also assume that, on average, employees' abilities on the objective elements are similar to their abilities on the subjective elements. This assumption is appropriate in this setting because the subjectively assessed competencies are chosen to represent skills that are essential in performing the objectively measured tasks. Hence, the abilities needed to improve performance on the objectively measured tasks are, by design, related to the abilities needed to improve the subjectively assessed competencies.²⁸ Finally, I assume that performance standards are, on average, similar for the objective and subjective elements.

The Determinants of Centrality Bias and Leniency Bias

As discussed in Section II, I predict that information-gathering costs and strong employee-manager relationships will have positive effects on both centrality and leniency bias. To examine the influence of these factors on centrality bias and leniency bias, I estimate Equations (3) and (4), respectively (see Table 4) using OLS regression analysis with robust standard errors clustered by manager to control for lack of independence between manager evaluations.²⁹

The results in Table 4 show that the coefficients on *INFO_C*, the proxy for information-gathering costs, are positive and significant for both centrality bias and leniency bias ($p = 0.06$ for *CEN_BIAS* and $p < 0.01$ for *LEN_BIAS*). This indicates that, consistent with H1a and H1b, managers exhibit more centrality bias and leniency bias when it is more costly for the manager to acquire information on the employee's performance.

The results also show a relationship between performance evaluation biases and the strength of the employee-manager relationship. *NEW_R* is negative and significant for both centrality bias and leniency bias ($p = 0.06$ for *CEN_BIAS* and $p = 0.03$ for *LEN_BIAS*). The results also show a negative coefficient on *DIF_AGE* for leniency bias ($p < 0.01$). Higher values on *DIF_AGE* and *NEW_R* suggest a relatively weak manager-employee relationship. Thus, consistent with H1c and H1d, the evidence indicates that managers give less compressed and less lenient ratings when the employee-manager relationship is weaker. *DIF_AGE* does not show a significant association with

²⁷ Another factor that might explain the more lenient ratings in 2003 is that managers were provided more leeway in 2003 since the system was new. My interviews indicate that top management asked the managers to be more rigorous in their use of the compensation system in 2004. Note that managers were not exclusively required to be stricter in rating performance; for example, they also were asked to be more timely with their performance reviews.

²⁸ Unfortunately, this is not true for the subjective output-based measures; they are not, by design, related to the objective output-based measures. To examine if this drives the results, I re-analyze Equations (1) to (5) for only those observations where all output-based performance measures are objective. The results (untabulated) show that all inferences remain unchanged when this limited set of observations is used.

²⁹ Variance Inflation Factors (VIF) indicate the absence of multicollinearity problems.

TABLE 4
The Determinants of Managers' Centrality Bias and Leniency Bias

$$CEN_BIAS_{jt} = \beta_0 + \beta_1 INFO_C_{jt} + \beta_2 NEW_R_{jt} + \beta_3 DIF_AGE_{jt} + \beta_4 DIF_GEN_j + \beta_5 GROWTH_{jt} + \beta_6 Y_{2003t} + \beta_7 NR_REF_{jt} + \sum_{k=8}^{11} \beta_k I_{Office_k} + \varepsilon_{jt} \quad (3)$$

$$LEN_BIAS_{it} = \beta_0 + \beta_1 INFO_C_{it} + \beta_2 NEW_R_{it} + \beta_3 DIF_AGE_{it} + \beta_4 DIF_GEN_i + \beta_5 GROWTH_{it} + \beta_6 Y_{2003t} + \sum_{k=7}^{10} \beta_k I_{Office_k} + \varepsilon_{it} \quad (4)$$

Independent Variables ^a	<i>CEN_BIAS</i>	<i>LEN_BIAS</i>
<i>INFO_C</i>	0.50* (1.97)	0.40*** (5.30)
<i>NEW_R</i> ^b	-1.22* (-1.93)	-0.13** (-2.23)
<i>DIF_AGE</i> ^b	-0.01 (-0.17)	-0.01*** (-2.87)
<i>DIF_GEN</i> ^b	0.14 (0.18)	0.01 (0.15)
<i>GROWTH</i>	0.01 (1.69)	0.01 (1.27)
<i>Y₂₀₀₃</i>	0.30 (1.02)	0.32*** (4.73)
<i>NR_REF</i>	0.11** (2.20)	
Adj. R ²	0.21	0.17

*, **, *** Indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels (two-tailed), respectively. t-values are in parentheses. For the centrality bias analysis j = reference group, t = time, k = offices, and n = 69. For the leniency bias analysis i = employee evaluations, t = time, k = offices, and n = 396.

^a An intercept and office indicator variables are included but not reported.

^b In Equation (3), presented in Column (2) *DIF_AGE* is the age difference between the manager and the average age of the reference group; *DIF_GEN* equals 1 if there is a gender difference between the manager and any of the reference group members; and *NEW_R* indicates the percentage of new relationships that exist in the reference group.

Variable Definitions:

CEN_BIAS = centrality bias, measured as the ratio between the standard deviation of the objective ratings of all employees in the reference group, and the standard deviation of the subjective ratings of all employees in the reference group;

LEN_BIAS = leniency bias, measured as the difference between the subjective and the objective performance rating; *INFO_C* = indicator variable that equals 1 if the manager does not share a workspace and does not overlap in duties with his subordinates, and 0 otherwise;

NEW_R = indicator variable that equals 1 if the employee or manager joined the company in the last four years, and 0 otherwise;

DIF_AGE = age difference between manager and employee;

DIF_GEN = indicator variable that equals 1 if there is a gender difference between the manager and the employee, and 0 otherwise;

GROWTH = percentage difference between the profit of this year and the profit of last year, calculated per office per year;

Y₂₀₀₃ = indicator variable that equals 1 if the observation relates to 2003, and 0 otherwise; and

NR_REF = number of employees in the reference group.

centrality bias, possibly because it is unable to capture the group dynamics that influence the manager's decision to compress ratings.

These results, together with the descriptive statistics, indicate that managers take their personal incentives and preferences into account when rating employee performance. Specifically, managers are lenient when objective ratings are low, and managers provide compressed and lenient ratings when it is costly to collect performance information and when the relationship between the manager and employee is relatively strong.

The results also provide indirect support for the assumptions underlying the centrality bias and leniency bias proxies. Consistent with previous research on centrality bias and leniency bias (Moers 2005), I assume that, on average, ability levels on the objectively measured and subjectively assessed tasks are equal. If the ability levels on the subjective performance elements had been higher, then the positive values on the leniency bias proxy could simply have been a reflection of these higher ability levels. Similarly, centrality bias may be mistaken for truly divergent ability levels on the objectively measured and subjectively assessed elements. However, as there is no reason to expect a strong association between relatively higher (lower variance in) ability levels and, for example, the manager's information-gathering costs, it is unlikely that the leniency (centrality) bias proxy merely reflects differences in ability levels.

The Performance Effects of Centrality Bias and Leniency Bias

As discussed earlier, proxies for centrality bias and leniency bias depend on the validity of the assumption that the objective rating is a good benchmark. However, this assumption is less essential when examining employee reactions. Even if the objective rating is not a good benchmark of actual performance, employees will likely use the objective ratings as a benchmark in comparing their subjective ratings and deriving their perceptions. Thus, because employees react according to their own perceptions, the following analyses are less dependent on the validity of the earlier assumptions.

Recall from Section II that the agency perspective predicts that centrality bias negatively affects performance incentives. The behavioral explanation, on the other hand, predicts that centrality bias positively affects incentives for below-average performers by enhancing employees' perceptions of fairness, while the incentives of above-average performers are negatively affected.³⁰ Hence, an overall negative relationship between centrality bias and performance would be consistent with the agency perspective, while no effect or a positive effect for below-average performers would be consistent with a behavioral explanation. I estimate Equation (5) (see Table 5) using OLS regression analysis with robust standard errors clustered by manager to control for lack of independence between manager evaluations.

The results presented in Table 5 show, consistent with H2a, that the coefficient on *CEN_BIASA* is negative and significant at the 5 percent level for the change in objective performance ratings (ΔPER_O) and at the 1 percent level for the change in subjective ratings (ΔPER_S) and total performance ratings (ΔPER_T).³¹ The results also show that centrality bias (*CEN_BIASB*) negatively influences the performance improvement of below-average performers ($p = 0.06$ for ΔPER_O and $p = 0.01$ for ΔPER_S and ΔPER_T). Consistent with predictions based on the agency perspective, this result indicates that centrality bias has a negative effect on all employees' incentives.

³⁰ To analyze the effect of centrality bias for above- and below-average performers separately, I split *CEN_BIAS* into two variables: *CEN_BIASA* and *CEN_BIASB*, where *CEN_BIASA* (*CEN_BIASB*) takes on the value of the ratio when the employee's objective rating is above (below) the average of the reference group, and 0 otherwise.

³¹ VIF indicates the absence of multicollinearity problems for all analyses presented in Table 5.

TABLE 5
The Effects of Managers' Performance Evaluation Biases on Employee Performance Improvement

$$\Delta PERF_i = \beta_0 + \beta_1 CEN_BIASA_i + \beta_2 CEN_BIASB_i + \beta_3 LEN_BIAS_i + \beta_4 MAX_SC_i + \beta_5 PART_T_i + \beta_6 OLDER_i + \beta_7 DIF_JOB_i + \beta_8 DIF_WO_i + \beta_9 DIF_NR_i + \sum_{k=10}^{13} \beta_k D_{Office_k} + \varepsilon_i \quad (5)$$

Independent Variables ^a	$\Delta PERF_O$	$\Delta PERF_S$	$\Delta PERF_T$
<i>CEN_BIASA</i>	-0.12** (-2.12)	-0.09*** (-4.10)	-0.10*** (-3.35)
<i>CEN_BIASB</i>	-0.11* (-1.95)	-0.07*** (-2.62)	-0.09*** (-2.67)
<i>LEN_BIAS</i>	0.69*** (6.90)	-0.08 (-1.12)	0.26*** (3.57)
<i>MAX_SC</i>	-0.23** (-2.42)	-0.02 (-0.23)	-0.12* (-1.87)
<i>PART_T</i>	0.04 (0.43)	0.05 (0.69)	0.04 (0.48)
<i>OLDER</i>	-0.25 (-1.23)	-0.01 (-0.12)	-0.12 (-1.01)
<i>DIF_JOB</i>	0.16 (0.85)	0.08 (0.92)	0.13 (1.32)
<i>DIF_WO</i>	0.01 (1.38)	-0.01 (-0.22)	0.01 (1.63)
<i>DIF_NR^b</i>	-0.02 (-0.57)	-0.06 (-1.54)	-0.02 (-0.67)
Adj. R ²	0.46	0.28	0.42

*, **, *** Indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels (two-tailed), respectively. t-values are in parentheses. For all three regressions i = employee evaluations, k = offices, and n = 198.

^a An intercept, as well as office indicator variables are included but not reported.

^b For the regressions with the dependent variable $\Delta PERF_O$, $\Delta PERF_S$, and $\Delta PERF_T$, I use the number of objective, subjective, and total performance measures, respectively, to calculate DIF_NR .

Variable Definitions:

$\Delta PERF_O$ = difference between the objective performance rating of 2003 and the objective performance rating of 2004;
 $\Delta PERF_S$ = difference between the subjective performance rating of 2003 and the subjective performance rating of 2004;

$\Delta PERF_T$ = difference between the total performance rating of 2003 and the total performance rating of 2004;

CEN_BIASA = variable that equals *CEN_BIAS* if the observation relates to an above-average performer, and 0 otherwise;

CEN_BIASB = variable that equals *CEN_BIAS* if the observation relates to a below-average performer, and 0 otherwise;

LEN_BIAS = leniency bias, measured as the difference between the subjective and the objective performance rating;

MAX_SC = indicator variable that equals 1 if the employee has reached the maximum of his salary scale, and 0 otherwise;

PART_T = indicator variable that equals 1 if the employee works part-time, and 0 otherwise;

OLDER = indicator variable that equals 1 if the employee is older than 50, and 0 otherwise;

DIF_JOB = indicator variable that equals 1 if the employee's job type is different in 2004 compared to 2003, and 0 otherwise;

DIF_WO = difference between the contractual incentive weights on the objective performance measures in 2003 and 2004; and

DIF_NR = difference between the number of performance measures included in 2003 and 2004.

Recall that the predictions for H2c based on the agency perspective and the behavioral perspective are opposing. If I find that leniency bias negatively affects performance improvements, then the agency view is more consistent with the results. However, if I find evidence of no effect or a positive effect, then the results are more consistent with the explanations of behavioral theory. The results in Table 5 show a positive association between leniency bias (*LEN_BIAS*) and performance improvements in the objective and total performance ratings ($p < 0.01$ for ΔPER_O and ΔPER_T).³² Thus, consistent with the behavioral perspective, I find empirical evidence that employee performance increases with greater leniency bias.

The results, however, show no significant association between leniency bias and performance changes in the subjective performance ratings (ΔPER_S). The less lenient subjective performance ratings in 2004 could possibly explain this insignificant association. Since the subjective performance ratings are biased, a performance change in the subjective ratings could be the result of a change in employee performance and/or a variation in the degree of leniency bias applied by the managers. Accordingly, even if the leniency bias of 2003 had a positive performance effect on the subjective elements, the reduction in leniency bias in 2004 may have prevented me from finding evidence of this positive effect.³³

Considering the limitations of this analysis, I perform a second analysis based on survey responses that indicate how self-reported employee effort changes after the introduction of the new incentive system. I estimate Equation (6) (provided in Table 6) using an ordered probit regression model with robust standard errors clustered by manager to control for lack of independence between manager evaluations. The results in Table 6 confirm the findings of the previous analysis. That is, I find a negative association between centrality bias (*CEN_BIAS*) and reported effort (*EFFORT*), indicating that the effort improvements are more modest for employees who received compressed ratings. Alternatively, the results show a positive association between leniency bias (*LEN_BIAS*) and reported effort (*EFFORT*), which indicates that the effort improvement is greater for employees who received lenient ratings.

In summary, I find strong support for the prediction that centrality bias and leniency bias affect the capacity of the performance-based compensation system to provide incentives. The results indicate that both the weakening of the link between pay and performance, as predicted by the agency perspective, and the improvement in perceived fairness, as predicted by behavioral theory, play a role in explaining the effect of performance evaluation biases on incentives.

V. CONCLUSION

Rating inaccuracy caused by managers' performance evaluation biases is perceived to be one of the main problems of introducing subjectivity into compensation contracts. This perception is due to the presumed negative effect of such bias on the compensation system's ability to motivate employees and provide valuable information for personnel decisions. However, our knowledge of

³² To investigate whether the leniency bias results on ΔPER_O are not simply the result of the objective rating increasing from 2003 to 2004 and the subjective rating staying the same, I perform an additional test. I generate a subsample for which both the objective and the subjective ratings are approximately the same for 2003 and 2004 (I split the sample at the average objective rating in 2003 [2.80] and use the above-average performers). Following, I perform the same analysis (Equation (5)) but only use the subsample. The results (untabulated) show that leniency bias has a significantly positive effect on ΔPER_O , which indicates that the findings are not simply the result of the significantly larger improvement in the objective ratings.

³³ Another limitation of this analysis is that the relationship between leniency bias and performance improvements is not free from mechanical positive associations. The 2003 objective and subjective ratings are used to calculate leniency bias and to capture performance improvements. For example, ΔPER_O is defined as OBJ_R_{2004} minus OBJ_R_{2003} , whereas *LEN_BIAS* is measured as $SUBJ_R_{2003}$ minus OBJ_R_{2003} . The second analysis examining the performance effects of performance evaluation biases does not suffer from this limitation.

TABLE 6

The Effects of Managers' Performance Evaluation Biases on Employee Effort Improvement

$$EFFORT_i = \beta_0 + \beta_1 CEN_BIAS_i + \beta_2 LEN_BIAS_i + \beta_3 MAX_SC_i + \beta_4 PART_T_i + \beta_5 OLDER_i + \varepsilon_i \quad (6)$$

Independent Variables^a	<i>EFFORT</i>
<i>CEN_BIAS</i>	-0.49** (-2.28)
<i>LEN_BIAS</i>	0.44*** (2.63)
<i>MAX_SC</i>	-0.97** (-2.14)
<i>PART_T</i>	-0.79* (-1.73)
<i>OLDER</i>	-0.20 (-0.35)
McFadden's Pseudo R ²	0.14

*, **, *** Indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels (two-tailed), respectively. Z-values are in parentheses, i = employees, and $n = 55$.

^a The intercept is included, but not separately reported.

Variable Definitions:

EFFORT = change in employee effort, measured by using the employee's self-reported change in effort attributed to the new incentive system (measured on a five-point Likert scale);

CEN_BIAS = centrality bias, measured as the ratio between the standard deviation of the objective ratings of all employees in the reference group, and the standard deviation of the subjective ratings of all employees in the reference group;

LEN_BIAS = leniency bias, measured as the difference between the subjective and the objective performance rating;

MAX_SC = indicator variable that equals 1 if the employee has reached the maximum of his salary scale, and 0 otherwise;

PART_T = indicator variable that equals 1 if the employee works part-time, and 0 otherwise; and

OLDER = indicator variable that equals 1 if the employee is older than 50, and 0 otherwise.

the impact of bias on the effectiveness of compensation contracting is limited. Empirical studies examining the consequences of biased performance ratings are especially lacking.

This study contributes to the literature by investigating (1) what causes managers to bias performance ratings, and (2) how biased ratings affect employee incentives. As predicted, the results show that managers respond to their own incentives and preferences when subjectively evaluating performance. Specifically, I find that information gathering and strong employee-manager relationships positively affect centrality bias and leniency bias. The results also show that bias affects not only current performance ratings, but also future employee incentives. Contrary to the predictions based on the agency perspective, I find that managers' performance evaluation biases are not necessarily detrimental to compensation contracting. Although centrality bias negatively affects performance improvement, the evidence does not reveal a significant negative relation between leniency bias and performance. Rather, I find that leniency bias affects performance positively. This is consistent with the behavioral argument that by improving perceived fairness, bias can positively influence incentives.

The finding that managers' performance evaluation biases can have a positive effect on employee incentives provides an explanation for earlier empirical research that shows that managers generally do not receive rewards for rating accurately (e.g., [Napier and Latham 1986](#)). Companies appear to be interested in the effectiveness of performance-based compensation contracts in increasing employees' future performance, but not necessarily in the accuracy of the performance ratings (or at least not for incentive purposes). An interesting question for future research would be to examine whether managers foresee the positive incentive effect of leniency bias and act accordingly, or whether leniency bias is merely the result of rating costs. That is, it would be interesting to learn whether the tendency to provide lenient ratings is (partly) strategic or just fortuitous.

Limitations to this study offer additional opportunities for future research. First, although there are no theoretical reasons to expect that the results would not extend to other settings, the ability to generalize the findings is limited by the reliance on data from only one firm. Moreover, I was able to obtain information for only two consecutive years. Consequently, I am unable to examine whether the documented effects of centrality bias and leniency bias persist over time. More extensive time-series data from multiple organizations would address these limitations.

The organization examined in this study is located in The Netherlands. Performance-based compensation systems play a less prominent role in The Netherlands than they do in, for example, the U.S. ([Jansen et al. 2009](#)). It would be interesting to examine whether the effects found in this study would be even stronger in countries where performance-based compensation systems play a more dominant role.

Another opportunity for future research would be to examine the effect of managers' performance evaluation biases on other uses of performance ratings. While I limit this study to incentive provision, subjective performance ratings are also likely to impact training and promotion decisions.

Finally, this study focuses only on the effects of biased performance ratings. Subjectivity in compensation contracting is likely to give rise to additional concerns, such as uncertainty about performance criteria ([Bol 2008](#)). Investigating how these other concerns, in combination with biased ratings, influence the effectiveness of compensation contracts would make an important contribution to the literature.

REFERENCES

- Abernethy, M. A., J. Bouwens, and L. van Lent. 2004. Determinants of control system design in divisionalized firms. *The Accounting Review* 79 (3): 545–570.
- Adams, J. S. 1963. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology* 67: 422–436.
- Akerlof, G. A., and J. L. Yellen. 1988. Fairness and unemployment. *The American Economic Review* 78 (2): 44–49.
- Arkin, R., H. Cooper, and T. Kolditz. 1980. A statistical review of the literature concerning the self-serving attribution bias in interpersonal influence situations. *Journal of Personality* 48 (4): 435–448.
- Baiman, S., and M. V. Rajan. 1995. The informational advantages of discretionary bonus schemes. *The Accounting Review* 70 (4): 557–579.
- Baker, G. P., R. Gibbons, and K. J. Murphy. 1994. Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics* 109 (4): 1125–1156.
- Banks, C. G., and K. R. Murphy. 1985. Toward narrowing the research-practice gap in performance appraisal. *Personnel Psychology* 38: 335–345.
- Barrett, R. S. 1966. *Performance Rating*. Chicago, IL: Science Research Associates.

- Beer, M., and G. J. Gery. 1972. Individual and organizational correlates of pay system preferences. In *Managerial Motivation and Compensation*, edited by H. L. Tosi. East Lansing, MI: Graduate School of Business of Michigan State University.
- Bernardin, H. J., D. K. Cooke, and P. Villanova. 2000. Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology* 85 (2): 232–234.
- Bernardin, H. J., and P. Villanova. 1986. Performance appraisal. In *Generalizing from Laboratory to Field Settings*, edited by E. A. Locke. Lexington, MA: Lexington Books D.C. Heath and Company.
- Blanchard, G. A., C. W. Chow, and E. Noreen. 1986. Information asymmetry, incentive schemes, and information biasing: The case of hospital budgeting under rate regulation. *The Accounting Review* 61 (1): 1–15.
- Blinder, A. S., and D. H. Choi. 1990. A shred of evidence on theories of wage stickiness. *The Quarterly Journal of Economics* 105 (4): 1003–1015.
- Bol, J. C. 2008. Subjectivity in compensation contracting. *Journal of Accounting Literature* 27: 1–27.
- Bouwens, J., and L. van Lent. 2007. Assessing the performance of business unit managers. *Journal of Accounting Research* 45 (4): 667–697.
- Bretz, R. D., G. T. Milkovich, and W. Read. 1992. The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management* 18 (2): 321–352.
- Bull, C. 1983. Implicit contracts in the absence of enforcement and risk aversion. *The American Economic Review* 73 (4): 658–671.
- Camerer, C. F., G. Loewenstein, and M. Rabin. 2004. *Advances in Behavioral Economics*. New York, NY: Princeton University Press.
- Cichello, M. S., C. E. Fee, C. J. Hadlock, and R. Sonti. 2009. Promotions, turnover, and performance evaluation: Evidence from the careers division managers. *The Accounting Review* 84 (4): 1119–1143.
- Cohen-Charash, Y., and P. E. Spector. 2001. The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes* 86 (2): 278–321.
- Colquitt, J. A., D. E. Conlon, M. J. Wesson, C. O. L. H. Porter, and K. Yee Ng. 2001. Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology* 86 (3): 425–445.
- Feldman, J. M. 1981. Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology* 66 (2): 127–148.
- Freeberg, N. E. 1969. Relevance of rater-ratee acquaintance in the validity and reliability of ratings. *Journal of Applied Psychology* 53 (6): 518–524.
- Friedrich, J. 1993. Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review* 100 (2): 298–319.
- Garland, H. 1973. The effects of piece-rate underpayment and overpayment on job performance: A test of equity theory with a new induction procedure. *Journal of Applied Social Psychology* 3: 325–334.
- Gibbs, M., K. A. Merchant, W. A. Van der Stede, and M. E. Vargus. 2004. Determinants and effects of subjectivity in incentives. *The Accounting Review* 79 (2): 409–436.
- Greenberg, J. 1990. Organizational justice: Yesterday, today, and tomorrow. *Journal of Management* 16 (2): 399–432.
- Harris, M. M. 1994. Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management* 20 (4): 737–756.
- Harris, M. M., and J. Schaubroeck. 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology* 41 (1): 43–62.
- Hayes, R. M., and S. Schaefer. 2000. Implicit contracts and the explanatory power of top executive compensation for future performance. *RAND Journal of Economics* 31 (2): 273–293.
- Higgins, E. T., and J. A. Bargh. 1987. Social cognition and social perception. *Annual Review of Psychology* 38: 369–425.
- Hogarth, R. M., and M. W. Reder. 1987. *Rational Choice*. Chicago, IL: The University of Chicago Press.
- Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10 (1): 74–91.
- Holmstrom, B. 1981. Contractual models of the labor market. *The American Economic Review* 71 (2): 308–313.

- Holmstrom, B., and P. Milgrom. 1991. Multi-task principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization* 7: 24–52.
- Indjejikian, R. J. 1999. Performance evaluation and compensation research: An agency perspective. *Accounting Horizons* 13 (2): 147–157.
- Indjejikian, R. J., and M. Matejka. 2006. Organizational slack in decentralized firms: The role of business unit controllers. *The Accounting Review* 81 (4): 849–872.
- Jansen, P., K. A. Merchant, and W. A. Van der Stede. 2009. National differences in performance-dependent compensation practices: The differing roles of financial performance measurement in the United States and The Netherlands. *Accounting, Organizations and Society* 34: 58–84.
- Jawahar, I. M., and C. R. Williams. 1997. Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology* 50 (4): 905–926.
- Kingsbury, F. A. 1922. Analyzing ratings and training raters. *Journal of Personnel Research* 1: 377–383.
- Kneeland, N. 1929. That lenient tendency in rating. *Personnel Journal* 7: 356–366.
- Landy, F. J., and J. L. Farr. 1980. Performance rating. *Psychological Bulletin* 87 (1): 72–107.
- Latham, G. 1986. Job performance and appraisal. In *International Review of Industrial and Organizational Psychology*, edited by C. Cooper, and I. Robertson. Chichester, U.K.: Wiley.
- Lawler, E. E. 1990. *Strategic Pay: Aligning Organizational Strategies and Pay Systems*. San Francisco, CA: Jossey-Bass.
- Levy, P. E., and J. R. Williams. 2004. The social context of performance appraisal: A review and framework for the future. *Journal of Management* 30 (6): 881–905.
- Longenecker, C. O., H. P. Sims, and D. A. Gioia. 1987. Behind the mask: The politic of employee appraisal. *The Academy of Management Executive* 1 (3): 183–193.
- MacLeod, B. W., and J. M. Malcomson. 1989. Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica* 57 (2): 447–480.
- McFarlane Shore, L., and G. C. Thornton. 1986. Effects of gender on self- and supervisory ratings. *Academy of Management Journal* 29 (1): 115–129.
- Mero, N. P., S. J. Motowidlo, and A. L. Anna. 2003. Effects of accountability on rating behavior and rater accuracy. *Journal of Applied Social Psychology* 33 (12): 2493–2514.
- Meyer, H. H. 1975. The pay-for-performance dilemma. *Organizational Dynamics* 3 (3): 39–50.
- Mischel, W., E. B. Ebbesen, and A. M. Zeiss. 1976. Determinants of selective memory about the self. *Journal of Consulting and Clinical Psychology* 44 (1): 92–103.
- Moers, F. 2005. Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society* 30 (1): 67–80.
- Motowidlo, S. J., and W. C. Borman. 1977. Behaviorally anchored scales for measuring morale in military units. *Journal of Applied Psychology* 62 (2): 177–183.
- Murphy, K. R., and J. Cleveland. 1991. *Performance Appraisal: An Organizational Perspective*. Boston, MA: Allyn and Bacon.
- Napier, N. K., and G. P. Latham. 1986. Outcome expectancies of people who conduct performance appraisals. *Personnel Psychology* 39 (4): 827–837.
- Pearce, J. L., and L. W. Porter. 1986. Employee responses to formal performance appraisal feedback. *Journal of Applied Psychology* 71 (2): 211–218.
- Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1): 7–63.
- Rees, R. 1985. The theory of principal and agent: Part I. *Bulletin of Economic Research* 37 (1): 3–26.
- Rynes, S. L., B. Gerhart, and L. Parks. 2005. Personnel psychology: Performance evaluation and pay for performance. *Annual Review of Psychology* 56: 571–600.
- Saal, F. E., R. G. Downey, and M. A. Lahey. 1980. Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin* 88 (2): 413–428.
- Saal, F. E., and F. J. Landy. 1977. The mixed standard rating scale: An evaluation. *Organizational Behavior and Human Performance* 18: 19–35.
- Shavell, S. 1979. Risk sharing and incentives in the principal and agent relationship. *The Bell Journal of Economics* 10 (1): 55–73.

- Shore, T. H., and A. Tashchian. 2002. Accountability forces in performance appraisal: Effects of self-appraisal information, normative information, and task performance. *Journal of Business and Psychology* 17 (2): 261–274.
- Smith, D. E. 1986. Training programs for performance appraisal: A review. *Academy of Management Review* 11 (1): 22–40.
- Thorndike, R. L. 1949. *Personnel Selection: Test and Measurement Techniques*. Oxford, U.K.: Wiley.
- Tsui, A. S., and C. A. O'Reilly. 1989. Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *The Academy of Management Journal* 32 (2): 402–423.
- Villanova, P., H. J. Bernardin, S. A. Dahmus, and R. L. Sims. 1993. Rater leniency and performance appraisal discomfort. *Educational and Psychological Measurement* 53: 789–799.
- Weinstein, N. D. 1980. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology* 39 (5): 806–820.
- Woehr, D. J., and A. I. Huffcutt. 1994. Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology* 67: 189–205.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.