

Discretion and bias in performance evaluation: the impact of diversity and subjectivity

Citation for published version (APA):

Moers, F. (2001). *Discretion and bias in performance evaluation: the impact of diversity and subjectivity*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 038 <https://doi.org/10.26481/umamet.2001038>

Document status and date:

Published: 01/01/2001

DOI:

[10.26481/umamet.2001038](https://doi.org/10.26481/umamet.2001038)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

**Discretion and Bias in Performance Evaluation:
The Impact of Diversity and Subjectivity***

Frank Moers

Maastricht University
Faculty of Economics and Business Administration
MARC / Department of Accounting & Auditing
P.O. Box 616
6200 MD Maastricht
The Netherlands
f.moers@berfin.unimaas.nl

Working Paper

July 2001

*I gratefully appreciate the support of the firm that provided me with the data for this study, especially the human resource manager of the firm. I am further grateful to Ken Merchant and the participants at the accounting seminar of Maastricht University for their comments and suggestions.

Discretion and Bias in Performance Evaluation:

The Impact of Diversity and Subjectivity

ABSTRACT: In this paper, I use a proprietary archival data set of one firm to examine the impact of performance measure diversity and the use of subjective performance measures on performance evaluation bias. The data set used is unique in the sense that it provides detailed information about a number of components of the subordinate-specific incentive contract, such as, the ‘incentive weights’ for both objective and subjective performance measures and the number of objective and subjective performance measures. The empirical results indicate that discretion, due to performance measure diversity and subjectivity, is related to performance evaluation bias. More specifically, I find that increased discretion is related to a compression of performance ratings and more lenient performance ratings. These results suggest that increasing the number of performance measures and using subjectivity in performance evaluation make it more difficult to differentiate among subordinates, which can lead to problems in personnel decisions and future incentives.

Key Words: *Discretion, diversity, subjectivity, incentives, bias*

Data Availability: *The confidentiality agreement with the firm that provided the data for this study precludes revealing its identity and disseminating data without its written consent.*

I. INTRODUCTION

The design of incentive contracts has attracted a lot of attention in accounting research. This research focuses on the factors that determine the use of performance measures in incentive contracts (e.g., Lambert and Larcker 1987) and the effects that these measures have on managerial behavior (e.g., Wallace 1997). Agency theory states that any (costless) performance measure that is informative about the agent's effort should be used for incentive purposes (Holmström 1979). Because no single performance measure is likely to be complete, the informativeness principle states that incentive contracts should include multiple performance measures. Especially the discussion about the incompleteness of financial performance measures has led to the call for the use of multiple performance measures (e.g., Kaplan and Norton 1996).

Analytical studies indicate that incentive contracting can be improved by incorporating more diverse performance measures, including the use of subjective performance measures. For example, Feltham and Xie (1994) show that the use of alternative performance measures can improve the effort allocation of the agent and reduce the risk imposed on the agent. Baker et al. (1994) indicate that the combined use of objective performance measures (explicit contracts) and subjective performance measures (implicit contracts) is, in some circumstances, optimal. Similarly, Baiman and Rajan (1995) show that the use of contractible (objective) and non-contractible (subjective) information in bonus pool arrangements leads to Pareto improvements compared to a situation where only contractible information is used.

Although increasing the number of performance measures and using subjective performance measures can provide more efficient incentives, it also provides the principal with more discretion in performance evaluation. Most analytical studies assume that the principal will not renege on promised incentive payments when having discretion either because of reputation concerns (Baker et al. 1994) or because he commits to a fixed bonus

pool (Baiman and Rajan 1995). However, superiors in organizations often are not the residual claimants of subordinates' outputs and therefore have no incentives to renege but rather have incentives to bias the performance evaluation (Prendergast and Topel 1993). Research in psychology indicates that superiors compress performance ratings and give more lenient performance ratings when these are used for incentive purposes (Milkovich and Newman 1993; Jawahar and Williams 1997). This bias in performance evaluation is problematic because it becomes more difficult to make the right personnel decisions, such as promotions. The problem of bias has only relatively recently been addressed in the economics literature (e.g., Prendergast and Topel 1993; 1996) and there is only limited empirical evidence of how performance measurement affects bias. Given the increased call for more diverse performance measures, including the use of subjective performance measures, it seems warranted to examine how performance measure diversity and subjectivity affects performance evaluation bias.

In this paper, I use a proprietary archival data set of one firm to examine the impact of performance measure diversity and the use of subjective performance measures on performance evaluation bias. Performance measure diversity is defined in this paper as the number of performance measures used for incentive purposes. Subjective performance measures, on the other hand, are defined as superior's subjective judgements about qualitative performance indicators. The data set used is unique in the sense that it provides detailed information about a number of components of the subordinate-specific incentive contract, such as, the 'incentive weights' for both objective and subjective performance measures and the number of objective and subjective performance measures. The empirical results indicate that discretion, due to performance measure diversity and subjectivity, is related to performance evaluation bias. More specifically, I find that increased discretion is related to a compression of performance ratings and more lenient performance ratings. These results

suggest that increasing the number of performance measures and using subjectivity in performance evaluation make it more difficult to differentiate among subordinates, which can lead to problems in personnel decisions and future incentives.

This paper contributes to the literature in several ways. First, it is one of the few studies that examines the effects performance measure diversity. Ittner and Larcker (1999) examine the effects of performance measure diversity on incentive plan outcomes, such as perceived financial benefits and plan terminations. They find little support for the benefits of using multiple (non-financial) performance measures. This paper extends their study by examining *why* this might be the case. Second, this paper addresses a topic that has recently come to the attention of the economics literature on incentive systems, i.e., discretion and bias, and it extends the theoretical literature by empirically examining how discretion, due to performance measurement, affects bias in performance evaluation.

The remainder of this paper is organized as follows. In section II, I describe the theoretical background and develop hypotheses. In section III, I discuss the research site, incentive plan, and the data collection. In section IV, I present the empirical results, and finally in section V, I discuss the implications of the empirical results for scorecard-type of performance measurement systems and provide a summary.

II. THEORY AND HYPOTHESIS DEVELOPMENT

Diversity, subjectivity and incentives

The dominant goal of incentive contracts is to give employees incentives to provide effort. Agency theory predicts that by relating pay to performance, employees are motivated to exert more effort in order to increase pay through improved performance (Holmström 1979). The incentive effects of incentive contracts are determined by the performance measures used because employees direct their attention to those aspects of the job that are being measured. As a result, the choice of performance measures is crucial in providing the

correct incentives. The informativeness principle indicates that any performance measure that provides (incremental) information about the employee's actions should be used for incentive purposes. Since no single performance measure is likely to be complete (Baker et al. 1994; Kaplan and Norton 1996), the informativeness principle predicts that incentive contracting is improved by incorporating a more diverse set of performance measures. For example, objective performance measures, like financial performance measures, are only informative about the measurable aspects of an employee's job and provide no incentives for the more qualitative aspects, like cooperation and innovation. Subjective performance measures, on the other hand, are informative about the qualitative job aspects and are therefore of value to incentive contracts since they provide incentives not provided by objective performance measures (Baker et al. 1994; Baiman and Rajan 1995). If objective and subjective performance measures are informative about different aspects of the agent's job, then the use of both measures in incentive contracts leads to a better allocation of effort.

Furthermore, performance measure diversity can also improve incentive contracting by reducing the risk imposed on the employee (Feltham and Xie 1994). Incentive contracts not only provide employees with incentives to provide effort but it also imposes risk on these employees. By relating pay to noisy measures of performance, employees' performance is affected by factors beyond their control and they need to be compensated for the risk imposed on them. Feltham and Xie (1994) state that the risk imposed on employees can be reduced by incorporating additional performance measures. They show analytically that the effect of uncontrollable factors can be (partly) adjusted for if multiple performance measures are used that are correlated. Thus, diversity in performance measurement not only leads to a better allocation of effort but it also reduces the risk imposed on employees, which leads to an improvement in the trade-off between incentives and risk sharing.

Diversity, subjectivity and performance evaluation bias

The previous discussion of the value of diversity and subjectivity in incentive contracts is based on agency theory, which traditionally assumes that an honest principal contracts with an agent who can't be trusted. The assumption of an honest principle becomes crucial if the principal has discretion in performance evaluation and incentive contracts are implicit (Prendergast and Topel 1993). Both subjective performance measures and diversity in performance measurement provide superiors with discretion in performance evaluation. Subjective performance measures provide the superior with discretion because no clear performance standards exist for these measures and assessed performance is solely determined by subjective judgements. Similarly, more diversity in performance measurement gives the superior a portfolio of performance measures that is likely to consist of partly conflicting outcomes. As a result, the superior has the opportunity to *ex-post* attach different weights of importance to each measure and give a performance rating that he sees fit. If the superior honestly reveals his subjectively assessed performance of the agent, then any implicit contract is in fact explicit. However, this assumption appears to be inconsistent with empirical evidence, which indicates that discretion in performance evaluation gives rise to a number of problems (Prendergast and Topel 1993).

Although the most obvious problem seems to be renegeing, which means that contracted performance is not rewarded, the incentives to renege are often non-existent for superiors because they are not the residual claimants of subordinates' output (Prendergast and Topel 1993). A more important problem is the issue of performance evaluation bias. The mere fact that superiors are not the residual claimants implies that superiors have an opportunity to let their preferences determine the allocation of rewards. Previous research in psychology suggests that performance ratings are lenient when these ratings are used for administrative purposes such as incentive pay and promotion decisions (Jawahar and Williams 1997).

Furthermore, superiors often insufficiently differentiate among subordinates, leading to a compression of performance ratings (Milkovich and Newman 1993). The incentives of superiors to bias the performance evaluation of subordinates relate to the psychological cost of communicating poor performance, favoritism, and preferences for equity in rewards (Prendergast and Topel 1993).

Bias in performance evaluation is problematic because there are not only direct costs associated with bias but also indirect costs. The direct costs relate to higher compensation costs than those warranted by the 'true' performance of the subordinates. The indirect costs relate to the difficulty of making important personnel decisions based on the performance ratings and the impact of incentives on motivation. If the performance ratings are biased, then 'all' employees seem to be above average performers and it becomes difficult to select the 'right' subordinate for the 'right' job. Further, if subordinates become aware of the bias, they might become less motivated and therefore provide less effort in the future. Since personnel decisions and incentives are important determinants of firm performance, the indirect costs of bias can be substantial and are likely to be much higher than the direct costs. Bias in performance evaluation is therefore an important aspect to consider when designing incentive contracts.

Although previous evidence suggests that bias in performance ratings exists when these ratings are used for administrative purposes, there is limited evidence on how discretion, due to differences in performance measurement, affects bias. Since diversity in performance measurement and subjective performance measures give the superior discretion in performance evaluation, it seems warranted to examine if performance measure diversity and subjectivity increase performance evaluation bias. Based on the previous discussion, I state the following hypotheses.

H1: Increased discretion, due to performance measure diversity, increases performance evaluation bias.

H2: Increased discretion, due to subjectivity in performance measurement, increases performance evaluation bias.

III. RESEARCH METHOD

Research site

The research site used in this study is a privately held Dutch industrial company focused on maritime activities. The company, hereafter called MARITCORP, was founded in 1875 and has since then always been active in shipbuilding and ship conversions. MARITCORP is primarily located in the Netherlands, although it has also subsidiaries in Belgium and the United Kingdom. MARITCORP has two major divisions, labeled Division A and Division B. Division A has three sub-holdings that are focused on shipbuilding, maintenance and repair of ships, and construction of oil and gas extraction installations for offshore and onshore fields. Division B is primarily focused on machining of intermediate and finished parts for ships, design and manufacturing of high-grade gear transmissions for ships, and technical services in the field of materials and welding technology. The peripheral activities of MARITCORP include the design, engineering and manufacturing of energy systems and high-grade, fiber-reinforced composite structures for aerospace, shipbuilding, wind energy, and other applications. MARITCORP currently employs approximately 1,300 employees and has sales of approximately \$200 million.

Incentive plan

In 1997, MARITCORP implemented an incentive plan for their higher-level subordinates. Previous to this implementation, subordinates' pay was fixed and was predominately determined by seniority. The firm stated that the incentive plan served the following four related purposes: (1) create clear responsibilities and performance-oriented

behavior, (2) promotion based on performance, skills and competencies, (3) employee differentiation, and (4) performance-based compensation. That is, the firm's goal is to give employees incentives to provide (additional) effort and to be able to promote the 'above average' performers.

The firm's incentive plan includes approximately 160 higher-level subordinates and consists of an annual bonus plan. The annual bonus is determined by two performance ratings, one based on objective performance measures and one based on subjective performance measures. For each employee, the specifics of the annual bonus plan are determined on an annual basis and involve the following steps. First, the size of the total annual bonus as a percentage of annual salary is determined by the human resource department and is based on the individual's job, education, and experience. Second, the employee's direct superior allocates the size of the total annual bonus to the two types of performance measures. Third, the superior chooses the number of measures in each performance dimension. Finally, each performance dimension is evaluated ex-post and the actual annual bonus is split into two different payments; one based on 'objective' performance and one based on 'subjective' performance. Performance ratings for these dimensions are made on a scale from zero to 100%, where 100% means full bonus payment.

Data

The firm provided me with proprietary archival data related to the plan's second year, i.e., 1998. Data are available with respect to the (1) compensation risk (total, objective, and subjective), (2) number of performance measures (total, objective, and subjective), and (3) bonus-related performance ratings (objective and subjective).¹ Data on the size of the compensation risk is available for 141 subordinates. The mean total compensation risk,

¹ The data provided do not relate to all employees included in the incentive plan because some compensation documents received by the human resource department were incomplete.

measured by the target annual bonus divided by annual salary, is 8.1%, i.e., approximately one month's salary. Although this might seem low compared to US standards, it is a significant incentive for MARITCORP's subordinates given that the firm moved from no performance evaluation and fixed pay to performance-based compensation. The mean compensation risk based on objective performance measures equals 4.3% of salary, while the mean compensation risk based on subjective performance measures equals 3.8%. The mean percentage point difference between the subjective and objective compensation risk, which reflects the relative use of subjectivity for incentive purposes, equals -0.6% and is significantly different from zero (two-tail $p < 0.01$). This means that, on average, superiors rely more on objective performance measures relative to subjective performance measures for incentive purposes. This result is consistent with previous research, which finds that superiors have a general tendency to rely more on objective performance measures (e.g., Ittner et al. 1999).

Data on the number of performance measures are available for 128 of the 141 employees. The mean total number of performance measures used is 4.7, which consists of 2.3 objective performance measures and 2.5 subjective performance measures. The difference between the number of objective and subjective performance measures is statistically significant (two-tail $p < 0.05$), which indicates that, on average, superiors use more subjective performance measures than objective performance measures.²

Bonus-related performance ratings are available for 141 employees. The mean overall performance rating is 59.8%, which means that, on average, approximately 60% of the maximum annual bonus is actually paid. The mean objective (subjective) performance rating equals 60.3% (58.9%).

² These results suggest that the number of performance measures used in each performance dimension does *not* reflect its relative importance for incentive purposes.

Table 1 provides the summary statistics for all variables used in this paper.

Insert TABLE 1 about here

Proxies for discretion

The prediction that performance measure diversity and subjectivity leads to performance evaluation bias is based on the assumption that both diversity and subjectivity increases discretion. I therefore use the following two proxies for discretion. First, the use of subjective performance measures allows more discretion in performance evaluation than the use of objective performance measures. This implies that the higher the relative importance of subjective performance measures for incentive purposes the higher the discretion. As a result, I split the sample into two groups based on the mean relative compensation risk based on subjectivity. The subgroup of low (high) relative compensation risk based on subjectivity proxies for low (high) discretion.

Second, although the use of objective performance measures allows less discretion than the use of subjective performance measures, increasing the number of objective performance measures adds a subjective component. The use of multiple objective performance measures provides the superior with an opportunity to attach subjective weights to the different performance measures, which increases discretion. I therefore split the sample into two subgroups of low and high discretion. The low discretion subgroup consists of subordinates who are evaluated on a single objective performance measure, while the high discretion subgroup consists of subordinates who are evaluated on multiple objective performance measures.

Measurement of bias

I measure performance evaluation bias by analyzing and comparing the moments of the distribution of performance ratings for the different subgroups identified above. More specifically, I examine the mean, standard deviation and skewness of the distributions. Evidence of bias, due to increased discretion, exists when the mean is higher, the standard deviation is lower, and there is more negative skewness in the high discretion subgroup compared to the low discretion subgroup, *ceteris paribus*. A higher mean and more negative skewness implies more lenient performance ratings, while a lower standard deviation implies greater compression of performance ratings.

IV. EMPIRICAL RESULTS

Diversity, discretion and performance evaluation bias

In order to examine whether diversity in performance measurement affects performance evaluation bias, I first regress the objective and subjective performance rating on respectively the number of objective and subjective performance measures. I include the compensation risk based on respectively the objective and subjective performance measures in the regression to control for the differences in incentives provided. The results, presented in Table 2, indicate that the objective performance rating is positively related to both the number of objective performance measures (two-tail $p < 0.10$) and the compensation risk based on objective performance measures (two-tail $p < 0.05$). The subjective performance rating, on the other hand, is not affected by either the number of subjective performance measures or compensation risk. If the performance ratings reflect the ‘true’ performance of subordinates, then these results suggest that annual bonus plans and performance measure diversity are effective when performance is measured objectively but not effective when performance is subjectively assessed. However, a more plausible explanation for the lack of significant

results with respect to the subjective performance ratings is that superiors have significant discretion in this situation and that the performance ratings are biased.

Insert TABLE 2 about here

Given that bias is likely to be present in the performance ratings observed, this raises the question whether the significant positive effect of the number of objective performance measures on the objective performance rating is due to the ‘economic value’ of performance measure diversity or due to discretion and bias. That is, although agency theory predicts that performance measure diversity positively affects performance, diversity in performance measurement also provides the superior with discretion, which might lead to bias and thus higher performance ratings. If the economic value of performance measure diversity drives the results in Table 2, then the same results should hold when superiors have discretion in performance evaluation. I therefore regress the objective performance rating on the number of objective performance measures and compensation risk for the subgroup of subordinates who are evaluated on multiple objective performance measures. Table 3 shows that the objective performance rating and the number of objective performance measures are not related when superiors have discretion. This implies that the results in Table 2 are not due to a gradual increase in the number of performance measures but due to a move from a single performance measure to multiple performance measures, i.e., discretion.

Insert TABLE 3 about here

Although the previous results indicate that the objective performance rating is positively related to discretion, this does not necessarily mean that the performance ratings are

biased. To examine whether the performance ratings are biased, I analyze the distribution of the objective performance ratings for subgroups of low discretion and high discretion, where these subgroups are based on the number of objective performance measures. It is important to note that the distribution of the compensation risk for these two sub-samples shows no significant differences. Therefore, any observed differences in the distribution of objective performance ratings cannot be explained by differences in incentives provided.

Panel A of Table 4 presents the distribution of the objective performance ratings for the subgroup with no discretion. The mean objective performance rating is 47%, which implies that, on average, almost half of the annual bonus based on objective performance measures is paid. The standard deviation of the distribution is 37%, while the distribution is symmetric in the sense that the number of observations below or above the mean is equal. Further, comparing the mean to scale midpoint, which is often used as a benchmark to evaluate leniency, indicates no significant difference. Since the distribution of this subgroup is most likely to approximate the distribution of ‘true’ performance, it provides evidence that, contrary to recent criticism (e.g., Kane et al. 1995), the scale midpoint is in fact a valid indicator of ‘average performance’ and therefore useful as a benchmark to evaluate leniency. Overall, the results suggest that the performance ratings in the low discretion subgroup are not lenient.

The distribution of the objective performance ratings for the subgroup with high discretion is shown in Panel B of Table 4. The mean objective performance rating equals 65% and is significantly different from the scale midpoint (two-tail $p < 0.01$). The standard deviation equals 27%, the distribution is mildly negatively skewed, and more than two-thirds of the observations are above the scale midpoint. These results indicate that the performance ratings are lenient when the superior has discretion.

Comparing the different moments of the distributions between the two subgroups shows first of all that the mean objective performance rating is higher for the subgroup with high discretion (two-tail $p < 0.01$). On average, the subordinates in the high discretion subgroup get paid an additional 18% of their maximum annual bonus based on objective performance measures. Second, the variance in performance ratings is significantly lower for the subgroup with high discretion (two-tail $p < 0.01$). This implies that there is significantly more compression of performance ratings and therefore less differentiation among subordinates in the high discretion subgroup. Finally, the distribution of objective performance ratings is more negatively skewed for the subgroup with high discretion.

Taken together, the empirical results lend support for hypothesis 1, which suggest that discretion, due to performance measure diversity, leads to both leniency and a compression of performance ratings.

Insert TABLE 4 about here

Subjectivity, discretion and performance evaluation bias

The results in Table 2 and, more specifically, the lack of significant results regarding the subjective performance ratings provide a first indication that subjectivity leads to bias. In order to examine to what extent discretion, due to an increased use of subjectivity, affects performance evaluation bias, I analyze the distribution of the overall performance rating for subgroups of low and high relative compensation risk based on subjectivity. Once again, a low (high) relative compensation risk based on subjectivity indicates low (high) discretion in performance evaluation.

The distribution for the low discretion subgroup, shown in Panel A of Table 5, indicates that the mean overall performance rating is 56%, which is not significantly different

from the scale midpoint. The standard deviation of the distribution equals 28% and the distribution is somewhat negatively skewed with approximately 59% of the observations above the scale midpoint. Panel B of Table 5 presents the distribution of the overall performance rating for the high discretion subgroup. The results show that the mean overall performance rating is 63%, which is significantly different from the scale midpoint (two-tail $p < 0.01$), while the standard deviation equals 24%. The distribution is mildly negatively skewed and approximately 75% of the observations above the scale midpoint.

Comparing the distributions for the low and high discretion subgroups indicates the following. First, the mean overall performance rating for the high discretion subgroup is higher than the mean rating for the low discretion subgroup, although the difference is only marginally significant (two-tail $p = 0.11$). Second, there is greater compression of performance ratings and less differentiation among the subordinates when discretion is high compared to when discretion is low, reflected by the lower variance of the overall performance rating for the high discretion subgroup (two-tail $p < 0.10$). Finally, the distribution for the high discretion subgroup is more negatively skewed than the distribution for the low discretion subgroup, which indicates higher leniency in the high discretion subgroup.

Insert TABLE 5 about here

The previous analysis is based on the assumption that the higher the relative compensation risk based on subjectivity the higher the discretion. However, the empirical results with respect to diversity indicate that using multiple objective performance measures also provides superiors with discretion. This implies that the differences in the distribution of the overall performance rating for low and high relative compensation risk based on subjectivity should be less pronounced for those subordinates who are evaluated on multiple

objective performance measures. I therefore re-run the analysis for this particular sub-set. Table 6 shows the results of this additional analysis and indicates that the difference in means for the two subgroups becomes less significant (two-tail $p=0.15$). Further, the difference in the variances of the overall performance ratings for the two subgroups becomes not significant, while there is a small decrease in the difference in negative skewness between the two distributions. These results are consistent with the expectation that the impact of subjectivity on performance evaluation bias becomes less pronounced when superiors use multiple objective performance measures to evaluate subordinates.

Insert TABLE 6 about here

In sum, the empirical results suggest that greater subjectivity in performance evaluation leads to more lenient performance ratings and a compression of performance ratings, which provides support for hypothesis 2.

V. DISCUSSION AND SUMMARY

The empirical results presented in the previous section have some important implications for scorecard-type of performance measurement and reward systems. These systems are characterized by multiple performance measures (diversity) and an increased use of subjectivity (e.g., Ittner et al. 1999). Furthermore, the purpose of these systems is to motivate employees to improve performance and to differentiate among employees based on their ability and skills so as to make better promotion decisions. Previous empirical research indicates that when multiple performance measures are available and superiors have discretion in ‘weighting’ the different performance measures, there is a general tendency to put more weight on objective and common measures of performance, which results in less ‘balance’ (e.g., Ittner et al. 1999; Lipe and Salterio 2000). The empirical results in this paper

indicate an additional problem with the use of multiple performance measures and subjectivity. First, the results show that performance measure diversity leads to more lenient performance ratings and less differentiation among employees. Kaplan and Norton's (2001) claim that, although the Balanced Scorecard should be used as a 'strategic management system', it can be linked to incentive compensation. They state that 'compensation can be based on 25 strategic measures' without causing problems (2001,152). However, this is more than five times the mean number of performance measures and almost three times the maximum number of performance measures used in the research site examined in this paper. It is therefore questionable whether a Balanced Scorecard that includes a large number of performance measures is effective as a performance measurement and reward system.

Second, Kaplan and Norton (1996) state that a balanced scorecard with multiple performance measures makes the use of subjectivity for incentive purposes easier. However, the empirical results indicate that subjectivity also leads to bias in performance evaluation. That is, if more subjectivity is used in evaluating and rewarding employees, superiors give higher performance ratings and compress these ratings. As a result, the firm is unable to separate the highly skilled employees from the less skilled employees. If skills and competencies are important determinants of promotions, then the use of subjective performance measures makes these promotion decisions more difficult.

As with any empirical study, this study has its limitations. First, the data do not allow me to examine the behavior of individual superiors and the analysis therefore assumes that all superiors behave in an identical way. Although research in psychology indicates that superiors have a general tendency to bias the performance ratings, it might be that superior-specific characteristics influence bias and future research can address these issues. Second, because the data are cross-sectional data of a single year, I am unable to examine to what extent performance evaluation bias persists and whether the firm actually incurs the assumed indirect

costs of bias. An opportunity for future research is therefore to gather time-series data and to examine to what extent performance evaluation bias affects, for example, motivation. Finally, the data relate to a single firm. Using data from a single firm controls for 'other' factors that can affect performance evaluation bias but it reduces the generalizability of the results. Future research can examine to what extent the findings in this paper are generalizable by gathering data from multiple firms.

REFERENCES

- Baiman, S., and M.V. Rajan. 1995. The Informational Advantage of Discretionary Bonus Schemes. *The Accounting Review* 70: 557-579.
- Baker, G., R. Gibbons, and K.J. Murphy. 1994. Subjective Performance Measures in Optimal Incentive Contracts. *Quarterly Journal of Economics* 109: 1125-1156.
- Feltham, G.A., and J. Xie. 1994. Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *The Accounting Review* 69: 429-453.
- Holmström, B., 1979. Moral Hazard and Observability. *Bell Journal of Economics* 10: 74-91.
- Ittner, C.D., and D.F. Larcker. 1999. The Effects of Performance Measure Diversity on Incentive Plan Outcomes. Working Paper, University of Pennsylvania.
- , ———, and M.W. Meyer. 1999. The Use of Subjectivity in Multi-Criteria Reward Systems. Working Paper, University of Pennsylvania.
- Jawahar, I.M., and C.R. Williams. 1997. Where All the Children Are Above Average: The Performance Appraisal Purpose Effect. *Personnel Psychology* 50: 905-925.
- Kane, J.S., H.J. Bernardin, P. Villanova, and J. Peyrefitte. 1995. Stability of Rater Leniency: Three Studies. *Academy of Management Journal* 38: 1036-1051.
- Kaplan, R.S. and D.P. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press, Boston.
- , and ———. 2001. Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part II. *Accounting Horizons* 15: 147-160.
- Lambert, R.A., and D.F. Larcker. 1987. An Analysis of the Use of Accounting and Market Measures of Performance in Executive Compensation Contracts. *Journal of Accounting Research* 25: 85-125.
- Lipe, M.G., and S.E. Salterio. 2000. The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures. *The Accounting Review* 75: 283-298.

- Milkovich, G.T., and J.M. Newman. 1993. *Compensation*. Irwin, Homewood, IL.
- Prendergast, C., and R. Topel. 1993. Discretion and Bias in Performance Evaluation. *European Economic Review* 37: 355-365.
- , and ———. 1996. Favoritism in Organizations. *Journal of Political Economy* 104: 958-978.
- Wallace, J.S. 1997. Adopting Residual Income-Based Compensation Plans: Do You Get what You Pay for?. *Journal of Accounting and Economics* 24: 275-300.

TABLE 1
Descriptive Statistics

<i>Variable</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Min-max</i>	<i>n</i>
Total compensation risk	8.1%	2.0%	3.00 – 15.00	141
Compensation risk based on objective performance measures ^a	4.3%	1.5%	0.00 – 9.38	141
Compensation risk based on subjective performance measures ^a	3.8%	1.2%	0.00 – 7.00	141
Total # of performance measures	4.7	1.5	2.00 – 9.00	128
# of objective performance measures	2.3	0.8	1.00 – 5.00	127
# of subjective performance measures	2.5	0.9	1.00 – 6.00	125
Overall performance rating	59.8%	25.6%	0.00 – 100	141
Objective performance rating	60.3%	30.3%	0.00 – 100	140
Subjective performance rating	58.9%	26.8%	0.00 – 100	138

^aOne subordinate is not evaluated on objective performance measures, while three subordinates are not evaluated on subjective performance measures.

TABLE 2
The Effect of Diversity and Compensation Risk on Performance Ratings
(t-statistics are in parentheses)

<i>Independent Variable</i>	<i>Dependent Variable</i>	
	<i>Objective Performance Rating</i>	<i>Subjective Performance Rating</i>
Intercept	30.70*** (3.09)	52.60*** (5.91)
# of OPM	6.20* (1.94)	
Compensation risk based on OPM	3.90** (2.20)	
# of SPM		-0.08 (-0.03)
Compensation risk based on SPM		2.05 (0.92)
Adjusted R-square	0.06	-0.01
F-statistic	5.27***	0.50
Sample size	127	125

***, **, * is statistically significant at respectively the 1%, 5%, and 10% level (two-tailed).

OPM = objective performance measures

SPM = subjective performance measures

TABLE 3
The Effect of Diversity and Compensation Risk on Objective Performance Ratings
when Superiors Have Discretion in Performance Evaluation
(t-statistics are in parentheses)

<i>Independent Variable</i>	<i>Objective Performance Rating</i>
Intercept	44.30*** (3.47)
# of OPM	2.29 (0.55)
Compensation risk based on OPM	3.31* (1.72)
Adjusted R-square	0.02
F-statistic	1.84
Sample size ^a	106

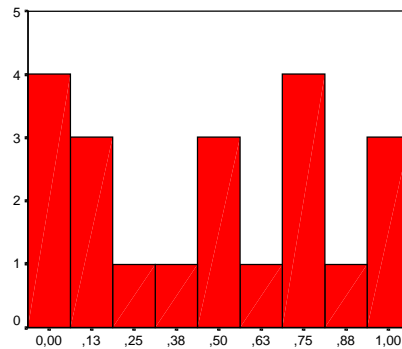
***, **, * is statistically significant at respectively the 1%, 5%, and 10% level (two-tailed).

OPM = objective performance measures

^aThe sample consists of those subordinates who are evaluated on multiple objective performance measures.

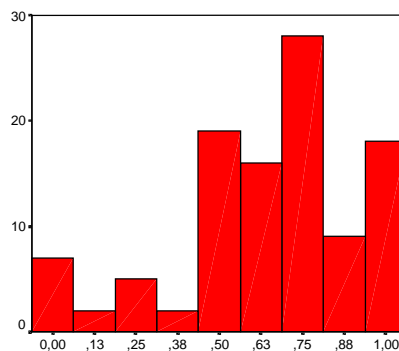
TABLE 4
Moments of the Distribution of Objective Performance Ratings
for Subgroups of Low and High Discretion in Performance Evaluation

Panel A: Low discretion (evaluation based on single objective performance measure)



<i>n</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Skewness</i>	<i>% of observations below scale midpoint</i>	<i>% of observations at scale midpoint</i>	<i>% of observations above scale midpoint</i>
21	47% ^{a,b}	37% ^d	0.036	43%	14%	43%

Panel B: High discretion (evaluation based on multiple objective performance measures)



<i>n</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Skewness</i>	<i>% of observations below scale midpoint</i>	<i>% of observations at scale midpoint</i>	<i>% of observations above scale midpoint</i>
106	65% ^{b,c}	27% ^d	-0.830	17%	15%	68%

^aNot significantly different from the scale midpoint.

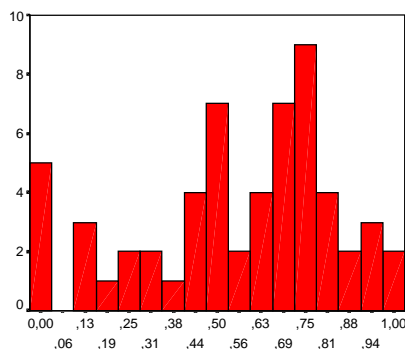
^bMean for the low discretion subgroup is significantly lower than the mean for the high discretion subgroup (two-tail p<0.01).

^cSignificantly different from the scale midpoint (two-tail p<0.01).

^dStandard deviation for the low discretion subgroup is significantly higher than the standard deviation for the high discretion subgroup (two-tail p<0.01).

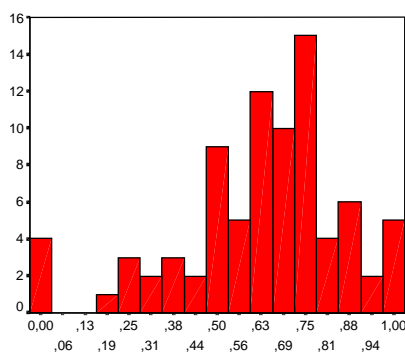
TABLE 5
Moments of the Distribution of Overall Performance Ratings
for Subgroups of Low and High Discretion in Performance Evaluation

Panel A: Low discretion (low relative compensation risk based on subjectivity)



<i>n</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Skewness</i>	<i>% of observations below scale midpoint</i>	<i>% of observations at scale midpoint</i>	<i>% of observations above scale midpoint</i>
58	56% ^{a,b}	28% ^d	-0.579	33%	9%	59%

Panel B: High discretion (high relative compensation risk based on subjectivity)



<i>n</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Skewness</i>	<i>% of observations below scale midpoint</i>	<i>% of observations at scale midpoint</i>	<i>% of observations above scale midpoint</i>
83	63% ^{b,c}	24% ^d	-0.862	18%	7%	75%

^aNot significantly different from the scale midpoint.

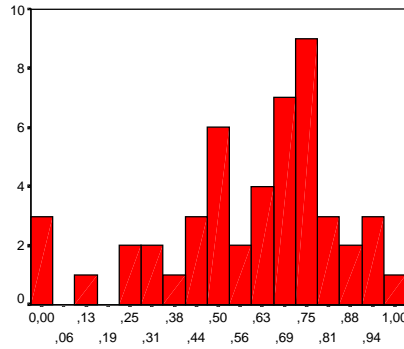
^bThe difference in the mean for the low discretion subgroup and high discretion subgroup is marginally significant (two-tail p=0.11).

^cSignificantly different from the scale midpoint (two-tail p<0.01).

^dStandard deviation for the low discretion subgroup is significantly higher than the standard deviation for the high discretion subgroup (two-tail p<0.10).

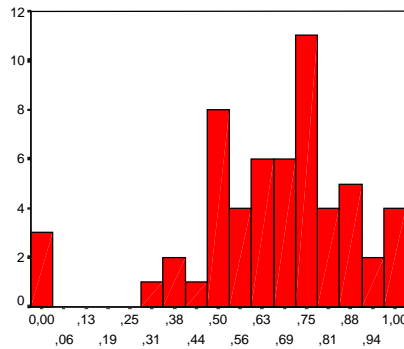
TABLE 6
Moments of the Distribution of Overall Performance Ratings
for Subgroups of Low and High Relative Compensation Risk Based on Subjectivity
Given the Use of Multiple Objective Performance Measures

Panel A: Low relative compensation risk based on subjectivity



<i>n</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Skewness</i>	<i>% of observations below scale midpoint</i>	<i>% of observations at scale midpoint</i>	<i>% of observations above scale midpoint</i>
49	59% ^{a,b}	25% ^d	-0.801	27%	10%	63%

Panel B: High relative compensation risk based on subjectivity



<i>n</i>	<i>Mean</i>	<i>St.Dev.</i>	<i>Skewness</i>	<i>% of observations below scale midpoint</i>	<i>% of observations at scale midpoint</i>	<i>% of observations above scale midpoint</i>
57	66% ^{b,c}	23% ^d	-1.106	12%	9%	79%

^aSignificantly different from the scale midpoint (two-tail $p < 0.05$).

^bThe difference in the mean for the 'low' subgroup and 'high' subgroup is marginally significant (two-tail $p = 0.15$).

^cSignificantly different from the scale midpoint (two-tail $p < 0.01$).

^dStandard deviation for the 'low' subgroup is not significantly different from the standard deviation for the 'high' subgroup.

