The British Accounting Review 45 (2013) 149–166

Contents lists available at SciVerse ScienceDirect

The British Accounting Review

journal homepage: www.elsevier.com/locate/bar

Procedural justice in managerial performance evaluation: Effects of subjectivity, relationship quality, and voice opportunity

François Bellavance [1], Suzanne Landry*, Eduardo Schiehll [2]

HEC Montreal, 3000 chemin de la Côte-Sainte-Catherine, Montreal, Quebec H3T 2A7, Canada

Keywords:
Procedural justice
Subjectivity
Performance evaluation
Voice
Relationship quality

A B S T R A C T

This study investigates the effect of subjectivity in performance evaluation on managerial perceptions of procedural justice. Using survey data from a sample of 317 managers, we examine two forms of subjectivity: use and weight of subjective performance measures and *ex post* flexibility in the weighting of multiple performance measures. We also examine the interaction effects of two contextual factors, superior–manager relationship quality and voice opportunity, on the association between subjectivity and perceived procedural justice. The results suggest that only the superior's use of *ex post* flexibility in weighting multiple performance measures adversely affects managers' perceptions of procedural justice. Moreover, superior–manager relationship quality reduces the negative effects of *ex post* flexibility in weighting multiple performance measures on procedural justice, whereas voice opportunity amplifies this negative effect. These findings have practical and theoretical implications, as they shed new light on the trade-off between the informative benefits and perceived unfairness of incorporating subjectivity into performance evaluation.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Performance evaluation is central to management control systems, and has been shown to influence managerial behaviour and performance (Feltham & Xie, 1994; Murphy & Cleveland, 1995; Otley, 1999). This influence depends on the properties of the performance measures and the fairness of the procedures used for the performance evaluation (Bol & Smith, 2011; Giraud, Langevin, & Mendoza, 2008; Greenberg, 1986; Hartman & Slapnicar, 2012).

Although performance evaluation systems can differ in many respects, one main distinction concerns the use of objective and subjective assessment (Baker, Gibbons, & Murphy, 1994). An objective performance evaluation is based on quantitative, verifiable measures and targets of organizational outcomes (e.g., productivity, profitability, and sales growth). Subjectivity in performance evaluation means that the superior uses discretion and judgement to assess non-quantifiable aspects of managerial performance, such as cooperation and knowledge-sharing within the organization and leadership and communication skills. Subjectivity can be introduced into performance evaluation in three different ways (Bol, 2008, p. 2; Bol

* Corresponding author. Tel.: +1 514 340 6564.
E-mail addresses: francois.bellavance@hec.ca (F. Bellavance), suzanne.landry@hec.ca (S. Landry), eduardo.schiehll@hec.ca (E. Schiehll).
[1] Tel.: +1 514 340 6485.
[2] Tel.: +1 514 340 6516.

0890-8389/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bar.2013.06.013

& Smith, 2011, p. 1214; Gibbs, Merchant, Van der Stede, & Vargus, 2004): 1) using subjective performance measures, 2) allowing for *ex post* flexibility in the weighting of multiple performance measures, and 3) allowing for *ex post* discretional adjustments based on factors other than performance measures specified *ex ante*.

In this study, we investigate the effects of the first two forms of subjectivity in performance evaluation—use of subjective performance measures and *ex-post* flexibility in the weighting of multiple performance measures—on managerial perceptions of procedural justice.[3] It is worth noting that our first form of subjectivity refers to the superior's use of subjective, non-quantifiable, non-financial performance measures only, and not the use of subjectivity in the assessment of quantifiable (financial and non-financial) measures. We also investigate whether two contextual factors, superior–manager relationship quality and voice opportunity, interact with these two forms of subjectivity to affect managerial perceptions of the procedural justice of the performance evaluation. We use primary data from a large cross-sectional sample of managers in Canadian firms to examine these issues.

Subjectivity improves performance evaluation insofar as it allows superiors to use private, non-quantifiable information to assess managerial efforts that are not captured by incomplete and noisy objective performance measures (Baker et al., 1994; Bol & Smith, 2011; Gibbs et al., 2004; Ittner, Larcker, & Meyer, 2003; Prendergast & Topel, 1993). However, given the cognitive limitations of evaluators and the lack of outcome verifiability, subjectivity is not always used effectively to evaluate performance. The most widely recognized problem is that subjectivity may give rise to *ex-post* rationalizations and untruthful, biased performance assessments (Krishnan, Luft, & Shields, 2005; Moers, 2005; Otley, 1999) as well as opportunism on the part of both evaluees and evaluators (Woods, 2009, p. 45). Consequently, the benefits of incorporating subjectivity into performance evaluation have been brought into question.

This study contributes to the literature in several ways. First, it extends the growing empirical research on the use of subjectivity in incentive mechanisms (e.g., Bol & Smith, 2011; Gibbs et al., 2004; Maas, Rinsum, & Towry, 2012; Moers, 2005). Whereas previous research has focused mainly on the determinants and/or performance effects of incorporating subjectivity into performance evaluation (e.g., Gibbs et al., 2004; Ittner et al., 2003; Nisar, 2007; Van der Stede, Chow, & Lin, 2006), our study investigates the effects of subjectivity in performance evaluation on perceptions of procedural justice. Our investigation contributes to this research stream by providing significant insights into the interplay between subjectivity in performance evaluation, contextual factors, and evaluees' fairness concerns. Gibbs et al. (2004), for example, report that the positive effects of subjective bonus on pay satisfaction depends on trust between subordinate and supervisor. Moers (2005) finds that the use of multiple objective and subjective performance measures leads to more lenient and compressed performance ratings, which could adversely affect incentive system effectiveness. Bol and Smith (2011) show that the level and controllability of an objective performance measure affect the supervisor's subjective evaluation of distinct, non-quantifiable aspects of an employee's performance. Whereas these studies enhance our understanding of how and why subjectivity is used in performance evaluation, they only indirectly assess the effects of subjectivity in performance evaluation on evaluees' perceptions of procedural justice, which is the main motivation for our study.

Second, our study departs from the previous literature by distinguishing between two forms of subjectivity in performance evaluation: 1) the use and weight of subjective performance measures to assess managerial tasks with non-quantifiable outcomes (Baker et al., 1994; Bol, 2008; Moers, 2005); and 2) *ex-post* flexibility in the weighting of multiple performance measures (Bol, 2008; Krishnan et al., 2005). It is important to examine different forms of subjectivity because they have different incentive purposes, and may consequently have different effects on evaluees' perceptions of procedural justice. On the one hand, when incentive contracts are incomplete, supervisors can use subjective measures to consider all dimensions of a manager's job, and not just the quantified measures in the formula-based bonus.[4] Thus, subjective measures reduce managers' short-term focus, prevent manipulations of quantitative measures, and induce managers to direct their attention towards all value-enhancing tasks (Baker et al., 1994; Gibbs et al., 2004; Ittner et al., 2003; Prendergast & Topel, 1996). Further, allowing for *ex post* changes in the performance weighting can reduce managers' compensation risk in the presence of organizational interdependencies and/or when uncontrollable factors may affect objective performance measures. At the same time, flexible weighting preserves the obligation for managers to react appropriately to events under their control (Bol & Smith, 2011; Gibbs et al., 2004). On the other hand, the literature also suggest that incorporating subjectivity in performance evaluation is beneficial only when the superior makes fair, unbiased judgements (Moers, 2005), if the subordinate does not try to inappropriately influence the superior's assessment (Woods, 2009, p. 45), and if the superior's subjective evaluation is not affected by the subordinate's performance on objective measures (Bol & Smith, 2011).

Furthermore, previous research suggests that the effect of performance measure properties on perceptions of procedural justice tends to be influenced by contextual factors such as target difficulty and participation (Libby, 1999), interpersonal trust (Gibbs et al., 2004; Hartman & Slapnicar, 2009; Maas et al., 2012), and task uncertainty (Hartman & Slapnicar, 2012). We extend this research stream by examining whether the effects of two forms of subjectivity on managerial perceptions of performance evaluation justice interact with two contextual factors: the quality of the superior–manager relationship and voice opportunity. Whereas there is substantial evidence on the direct positive effects of a high-quality superior–manager

[3] By multiple performance measures, we mean combining objective financial with both objective and subjective nonfinancial performance measures (Kaplan & Norton, 1996; Van der Stede et al., 2006).

[4] Incentive contracts are incomplete when they are based on objective measures that do not capture all dimensions of managerial performance (Baker et al., 1994).

relationship (Hartman & Slapnicar, 2009; Kennedy, Kohlmeyer, & Parker, 2009; Rupp & Cropanzano, 2002) and voice opportunity (Hartman & Slapnicar, 2012; Shaw, Wild, & Colquitt, 2003) on perceptions of procedural justice, we focus on their interaction effect with subjectivity on performance evaluation.

Our results suggest that the use and weight of subjective performance measures does not have an overall significant effect on perceived procedural justice. We find strong evidence that only *ex post* flexibility in weighting multiple performance measures negatively affects managers' perceptions of procedural justice. Moreover, *ex post* flexibility interacts with each of the two contextual factors (quality of the superior–manager relationship and voice opportunity). A high-quality superior–manager relationship reduces the negative effects of *ex post* flexibility in weighting multiple performance measures on procedural justice. However, voice opportunity strengthens the negative effect of flexibility in weighting on perceived procedural justice.

In the next section we provide the theoretical background for our study and develop our hypotheses, which relate the two forms of subjectivity in performance evaluation to managerial perceptions of procedural justice. In the third section we describe the research methods. In the fourth section we present the empirical results, and in the fifth section we discuss the findings and present our conclusions.

2. Literature review and hypothesis development

Organizational justice is an area of psychological inquiry that concerns perceptions of fairness in the workplace, including how individuals experience different aspects of their employment (Fortin, 2008). Initial studies focused on the perceived fairness of outcomes (distributive justice). Subsequently, new dimensions were gradually added, including respectful and sensitive treatment (interpersonal justice), adequate and sufficient communication (informational justice), and the fairness of the procedures used to determine outcomes (procedural justice) (Folger & Cropanzano, 1998). We focus on this last dimension, which we have termed managerial perceptions of procedural justice.[5]

2.1. Procedural justice in performance evaluation

The notion of procedural justice was introduced by Thibaut and Walker (1975). They examined courtroom settings and distinguished between the fairness of the verdict and the fairness of the process leading to the verdict. Procedural justice was transferred to business organizations by Leventhal (1980), who identified six criteria of perceived procedural justice.[6] According to these criteria, subsequent research has shown that performance evaluation is most likely to be perceived as fair when evaluees have access to detailed information about the performance measures used and perceive them as highly relevant, and when the performance evaluation is conducted uniformly and without bias among subordinates (Greenberg, 1986; McFarlin & Sweeny, 1992; Prendergast & Topel, 1996). Despite the influence of procedural justice on the effectiveness of control mechanisms, the management accounting research, with a few exceptions (Hartmann & Slapnicar, 2012; Mass et al., 2012; Lau & Moser, 2008), has largely ignored the conditions that bring performance evaluation procedures more or less in line with the principles of procedural justice.

Baker et al.'s (1994) agency-based model proposes that an employee's contribution to firm value cannot be completely measured by objective measures, and that superiors who are well placed to observe subtleties of employee behaviour can perform a subjective evaluation to complement these measures. Moreover, at managerial levels, where jobs are complex, subjectivity in performance evaluation can therefore complement the available objective measures and improve incentive contracting (Bol, 2008; Gibbs et al., 2004). Thus, incorporating subjectivity into performance evaluation can benefit both firms and managers. Subjectivity enables superiors to use non-quantifiable and other relevant information that emerges during the evaluation period, reducing incentive costs for firms and incentive risks for managers (Gibbs et al., 2004; Ittner et al., 2003; Prendergast & Topel, 1996).

However, studies also suggest that subjectivity in performance evaluation is effective only if the superior makes fair, unbiased judgements, and if the subordinate does not try to inappropriately influence the superior's assessment. For example, Moers (2005) shows that the use of subjective performance measures leads to performance evaluation bias. Woods (2009, p. 45) demonstrates that superiors subjectively adjust objective measures according to managerial influence activities as well as their own personal preferences. Bol and Smith (2011) show that the level and controllability of an objective performance measure affect the supervisor's subjective evaluation of distinct, non-quantifiable aspects of employee performance. Our study extends this research stream by examining how subjectivity in performance evaluation affects managerial perceptions of procedural justice.

We take the perspective that subjectivity in performance evaluation may affect evaluees' perceptions of procedural justice (Hartmann & Slapnicar, 2009; Maas et al., 2012; Murphy & Cleveland, 1995; Lau & Moser, 2008; Prendergast & Topel, 1996). We therefore focus on the performance evaluation of multi-task managers who deal with uncontrollable events and outcomes that are not always easily quantifiable. In such situations, subjectivity is a necessary condition for effective incentive

[5] The terms "justice" and "fairness" are used interchangeably in this field.

[6] Organizational procedures are perceived as fair when they 1) are consistent across people and over time, 2) are free of bias, 3) are accurate, 4) include mechanisms for correcting wrong decisions, 5) adhere to prevalent conceptions of morality, and 6) are task representative, which implies process and decision control (Leventhal, 1980).

contracting, but it may adversely impact perceptions of procedural justice. Consistent with Bol (2008), Bol and Smith (2011), and Gibbs et al. (2004), we consider that subjectivity in performance evaluation can be incorporated by 1) the use and weight of subjective performance measures and 2) allowing for *ex post* flexibility in the weighting of multiple performance measures.[7] We also examine the interaction effects of two contextual factors (quality of the superior–manager relationship and voice opportunity) on the association between subjectivity in performance evaluation (use and weight of subjective measures and *ex post* flexible weighting) and managerial perceptions of procedural justice.

In the next sections, we distinguish the two forms of subjectivity and discuss their expected association with perceptions of the procedural justice of the performance evaluation. This is followed by a discussion of the potential interaction effects of the two contextual factors: superior–manager relationship quality and voice opportunity.

2.2. Use and weight of subjective performance measures

The management accounting literature suggests that an optimal performance measure is accurate, informative, and timely, and that it takes into account the effects of current actions on the firm's future profitability, all without imposing undue risk on the employee (Baker et al., 1994; Bushman, Indjejikian, & Smith, 1996; Holmstrom, 1979). In practice, however, objective (financial and nonfinancial) performance measures rarely meet these attributes.[8] They are considered too aggregate, narrow in focus, historical in nature, and inadequate to measure a manager's complex, multi-task efforts (Baker et al., 1994; Bol, 2008; Holmstrom, 1979; Ittner et al., 2003; Prendergast & Topel, 1993). Consequently, objective performance measures are often complemented by subjective performance measures in an attempt to mitigate the incentive distortions caused by imperfect objective performance measures (Moers, 2005; Van der Stede et al., 2006).[9]

Accordingly, evaluators can use subjective performance measures such as cooperation and knowledge-sharing within the organization, leadership skills, and loyalty to the organization in order to assess managerial tasks with non-quantifiable outcomes. Emphasis can also be placed on subjective measures of a manager's contribution in order to offset uncontrollable events that affect managerial performance, thereby reducing the manager's incentive risk and the organization's contracting costs (Baker et al., 1994; Bol & Smith, 2011; Gibbs et al., 2004; Prendergast & Topel, 1996). More weight given to subjective measures also provides evaluators with more discretion to rate and rank managerial performance (Ittner et al., 2003). However, because there are no clear performance standards for subjective measures, they can be problematic for decision making about personnel and future incentives (Moers, 2005). More importantly, the uncertainty in the absence of explicit performance standards prevents evaluees from verifying whether or not their performance was assessed fairly, and could lead them to believe that bias and favouritism played a major role (Bol, 2008).[10]

In fact, experimental research shows that shortcomings of human judgement limit incentives that are based on subjective performance measures (Bol & Smith, 2011; Fisher, Maines, Peffer, & Sprinkle, 2005; Moers, 2005). For example, Moers (2005) provides empirical evidence that both performance measure diversity and subjectivity are positively related to the evaluator's performance evaluation bias, while Bol and Smith (2011) demonstrate cognitive distortion in subjective evaluations when evaluators are informed about the evaluees' level and controllability of an objective, a quantifiable performance outcome. Therefore, we expect that the more weight placed on subjective performance measures, the lower the manager's perceptions of the procedural justice of the performance evaluation, regardless of the benefits for incentive contracting. This leads to the following hypothesis:

H1a. The use and weight of subjective measures is negatively associated with managerial perceptions of the procedural justice of the performance evaluation.

2.3. Flexibility in the weighting of multiple performance measures

As discussed above, in order to offset the incompleteness and short-term perspective of financial, accounting-based performance measures, many organizations combine objective financial measures with both objective and subjective nonfinancial performance measures (Kaplan & Norton, 1996; Van der Stede et al., 2006). A critical issue in implementing multiple performance measures is how to determine their relative weight in order to assess overall managerial performance (Ittner et al., 2003). One option is to use a formula-based approach that explicitly determines the *ex ante* weight for each preselected performance measure. Nevertheless, the appropriate weight can still be difficult to determine. Moreover, empirical research suggests that the formula-based approach may lead to an unbalanced performance evaluation, where not

[7] According to Bol (2008), another form of subjectivity in performance evaluation can be introduced by allowing the superior (evaluator) to make discretionary adjustments and assessments of quantifiable (financial and non-financial) measures, based on factors other than the performance criteria specified. This form of subjectivity is not investigated in this study.

[8] In the agency-based analytical literature, performance measures are considered imperfect if they are 1) insensitive to subordinate actions, 2) incongruent with organizational objectives, 3) noisy, 4) incomplete, 5) unverifiable, and 6) manipulable (Banker & Datar, 1989; Feltham & Xie, 1994).

[9] At Lincoln Electric, for example, half a manager's incentive pay comes from the superior's assessment of the manager's cooperation, innovation, dependability, and other subjective aspects of performance (Baker et al., 1994).

[10] The incentives for superiors to bias the performance evaluation of subordinates often relate to the psychological cost of communicating poor performance, favouritism, and preferences for equity in rewards (Prendergast & Topel, 1993).

all the relevant dimensions of managerial effort are fairly captured by *ex ante* performance measures (Baker et al., 1994; Gibbs et al., 2004; Ittner et al., 2003; Krishnan et al., 2005).

Alternatively, organizations can introduce *ex post* flexibility in the weighting of multiple performance measures (Bol, 2008).[11] Flexible weighting implies that superiors either do not establish explicit *ex ante* weights for the multiple performance measures or they have *ex post* (after the evaluation period) discretion to weight each performance measure (Fisher et al., 2005; Ittner et al., 2003). Hence, superiors can use *ex post* flexibility to weight multiple performance measures in order to adjust for uncontrollable factors and improve the sensitivity of incentives to managerial effort and decision-making rights (Banker & Datar, 1989; Gibbs et al., 2004; Holmstrom, 1979; Ittner et al., 2003).

However, Prendergast and Topel (1996) demonstrate that flexible weighting of multiple performance measures opens the door to favouritism: superiors may treat employees according to their personal preferences. In a field study, Ittner et al. (2003) show that flexible weighting allows superiors to tip the balance by placing more weight on outcome measures, changing the importance of evaluation criteria across periods, and overweighting measures that do not necessarily predict desired outcomes. They also report that flexible weighting gives rise to complaints about favouritism and generates uncertainty about the reward criteria. In addition, Lipe and Salterio's (2000) experimental study suggests that, due to cognitive limitations, performance measures commonly used across individuals tend to be more heavily weighted than measures that are unique to a particular job, resulting in misallocations of agent effort.

The above discussion suggests that an evaluator's cognitive limitations and inconsistent judgements may offset the benefits of *ex post* flexibility in the weighting of multiple performance measures. In view of these downside effects, we predict a negative association between the use of *ex post* flexible weighting and managers' perceptions of procedural justice in performance evaluation. This leads to the following hypothesis:

H1b. *Ex post* flexibility in the weighting of multiple performance measures is negatively associated with managerial perceptions of the procedural justice of the performance evaluation.

2.4. Interaction effects of superior–manager relationship quality and voice opportunity

Studies have demonstrated that superiors play a central role in managers' perceptions of the performance evaluation process: they are responsible for setting performance targets, providing formal and informal feedback, and determining overall managerial performance (e.g., Hartmann & Slapnicar, 2009, 2012; Libby, 1999; Prendergast & Topel, 1993). Moreover, experimental studies suggest that managers are active agents who attempt to influence the performance evaluation process and its outcomes (Maas et al., 2012; Woods, 2009, p. 45). Consequently, the quality of the superior–manager relationship and voice opportunity may be viewed as key contextual factors that influence perceptions of the performance evaluation process (Dulebohn & Ferris, 1999; Lindquist, 1995). Given the abundant evidence of the direct positive effects of a high-quality superior–manager relationship (Rupp & Cropanzano, 2002) and voice opportunity (Shaw et al., 2003) on perceptions of procedural justice, we focus on their interaction effect with subjectivity on performance evaluation.[12] To our knowledge, no previous study has investigated these effects in the presence of subjectivity in the performance evaluation process.

A good superior–manager relationship is assumed to improve communication and openness within the organization, and consequently to enhance perceptions of procedural justice (Lind & Tyler, 1988; Rupp & Cropanzano, 2002). Accordingly, the management accounting literature has demonstrated an association between managers' perceptions of their superior and their perceptions of incentive systems. Gibbs et al. (2004) show that a positive effect of subjective bonuses on managers' pay satisfaction depends on trust in the superior, measured by the managers' tenure in the organization. Similarly, Kennedy, Kohlmeyer, and Parker (2009) show that managers' perceptions of the procedural and distributive justice of a gain-sharing plan are positively influenced by their perceptions of their superior's character and trustworthiness. More importantly, Hartman and Slapnicar (2009) demonstrate that managers' trust in their superior influences the effects of the formality of the performance evaluation system on perceptions of procedural justice. Taken together, these findings suggest that a high-quality superior–manager relationship leads to perceptions that the superior will evaluate performance fairly and consistently, which would minimize fairness concerns. We therefore argue that a high-quality superior–manager relationship positively influences how managers perceive their superior's use of subjectivity in the performance evaluation. We therefore hypothesize that:

H2. The higher the quality of the superior–manager relationship, the weaker the negative association **a)** between the use and weight of subjective measures and managerial perceptions of procedural justice in the performance evaluation; and **b)** between *ex-post* flexibility in the weighting of multiple performance measures and managerial perceptions of procedural justice in the performance evaluation.

[11] Subjectivity in performance evaluation can also take the form of allowing *ex post* discretionary adjustments or subjective assessments of quantifiable measures, based on factors other than objective (quantifiable) financial and nonfinancial outcomes specified *ex ante*. However, this is not the focus of our study.

[12] We do not formulate hypotheses for the direct effects of superior–manager relationship quality or voice opportunity on perceived procedural justice, as they are well documented in the literature.

Since the seminal work on procedural justice by Thibaut and Walker (1975), researchers consistently find that subordinates' perceptions of justice are also influenced by the extent to which they are allowed to express their opinions, known as voice opportunity.[13] The justice-enhancing effect of voice opportunity results from providing subordinates with the opportunity to have input into specific evaluation procedures and from promoting positive relationships within social groups (Avery & Quinones, 2004; Fulk, Brief, & Steve, 1985; Greenberg, 1986; Shaw et al., 2003). The management accounting literature also shows that employees perceive performance evaluations as fairer when they have opportunities to provide their input on performance criteria and targets or to justify their performance outcomes. Libby (1999) shows that subordinates who have a voice in budget setting show higher performance, explained by the positive effect of voice on the perceived fairness of budgeting systems. Herath, Bremser, and Brinberg (2010) analytically demonstrate the influence of voice opportunity on weights and metrics in a balanced scorecard setting, while Hartman and Slapnicar (2012) find a stronger effect of voice opportunity on perceptions of performance evaluation justice in managers who face higher task uncertainty. We assume that voice opportunity comprises a variety of practices that allow managers to have input on specific evaluation procedures throughout the performance evaluation process. We expect voice opportunity to positively interact with managers' perceptions of their superior's use of subjectivity in the performance evaluation. We therefore hypothesize that:

H3. The greater the voice opportunity, the weaker the negative association **a)** between the use and weight of subjective measures and managerial perceptions of procedural justice in the performance evaluation; and **b)** between *ex-post* flexibility in the weighting of multiple performance measures and managerial perceptions of procedural justice in the performance evaluation.

3. Research methods

3.1. Sample and data collection

Given the private nature of the data required to test our research hypotheses, we used a cross-sectional mail survey. The sample is drawn from the Quebec Certified Management Accountant (CMA) membership database. Participants had to reside in Canada, have held professional designation for at least three years, and have more than five years' relevant experience. Participants also had to work for organizations with more than 50 employees, as larger organizations would increase the likelihood of a performance evaluation system and an associated reward system. In addition, participants had to have worked for the same employer for longer than one year. These criteria were designed to obtain a sample of respondents holding middle- and upper-level management positions, having considerable work experience, and having undergone at least one performance appraisal in the current organization. The survey package, including the questionnaire and a self-addressed, pre-paid return envelope, was mailed in mid-October 2008. Three weeks after the initial mailing, a first email reminder was sent, followed by a second reminder six weeks later. Survey administration therefore covered a total of nine weeks, with the last questionnaire received on December 17, 2008 (cut-off date). In all, 3280 questionnaires were sent to CMAs in the database who met the above inclusion criteria and 339 completed questionnaires were returned, for a response rate of 10%.

We used Dillman's total survey design method (Dillman, Smyth, & Christian, 2009) and followed the recommendations by Van der Stede, Young, and Chen (2005). Whenever possible, existing survey instruments were used or refined to capture the variables of interest. All measurement instruments included in the final survey were evaluated for content validity. First, we conducted in-depth interviews with two human resource consultants to validate the scales. We then asked four management accounting faculty and four practitioners to pre-test the survey and confirm their understanding of all the measurement instruments. The instruments were refined until consensus was reached. As suggested by Dillman et al. (2009), we also tested for non-response bias. We initially compared the descriptive statistics of the three demographic variables available—age, years in current position, and firm size (measured by the number of employees)—between all eligible participants in the CMA database ($N = 3280$, including respondents) and the respondents ($N = 339$). The results revealed no significant differences. The second step was to identify a subsample of late respondents (5.9% of all respondents) based on the date that the completed survey was received. We compared early and late respondents on the three above demographic variables and the five variables of interest: perceptions of procedural justice, use and weight of subjective performance measures, *ex post* flexibility in the weighting of multiple performance measures, superior–manager relationship quality, and voice opportunity. Comparisons between early and late respondents showed no systematic differences for any of the variables. We excluded 22 observations due to missing data for at least one measurement instrument ($n = 19$) or for years of experience ($n = 3$). For the categorical control variables, missing data were considered as an additional, separate category to limit the number of excluded observations due to missing data (Cohen, Cohen, West, & Aiken, 2003). All statistical analyses were consequently performed on a sample of 317 subjects.

[13] In fact, voice opportunity is one of the most frequently studied topics in the procedural justice research. See for example the meta-analytic review by Shaw et al. (2003).

3.2. Measurement instruments

3.2.1. Use and weight of subjective performance measures (Wsubj)

We adapted Van der Stede et al.'s (2006) survey instrument to capture the extent of use of subjective measures in performance evaluation systems.[14] Given our cross-sectional target sample, we also borrowed performance criteria from performance measurement systems currently used by financial institutions (Nisar, 2007; Ittner et al., 2003). As presented in Appendix A, the final list comprises 40 performance measures divided into five categories: financial (7 items), nonfinancial customer-oriented (8 items), nonfinancial employee-oriented (7 items), nonfinancial internal operating (8 items), and subjective (10 items). Respondents were asked to check the performance measures that their superior used to evaluate their performance, write in other measures that were not listed, and indicate the relative weights attributed to each performance measure category.

The collected data reveal a wide range of performance measures used across the sample, with 11.2 measures on average. Moreover, 36 respondents wrote in 14 additional performance measures under the five categories.[15] Compared to the study by Van der Stede et al. (2006), our sample reports a similar average number of subjective measures but a lower average number of financial and objective nonfinancial measures. We believe this difference can be explained by the upper-level managerial positions and the highly cross-sectional nature of our sample.[16]

Performance measure weights are balanced within the three main categories: a mean of 30% for financial, 38% for objective nonfinancial (12% for customer-oriented, 11% for internal operating, and 15% for employee-oriented), and 32% for subjective measures. In order to capture the use and importance of subjective performance measures in the performance evaluation system, we use the weight—in other words, the percentage—attributed to the subjective performance measure category (Wsubj). Hence, in contrast to previous studies that used a dichotomous variable to indicate the presence (or not) of subjective performance measures, our variable captures not only the use of subjective measures, but also their weight relative to objective financial and nonfinancial performance measures.

3.2.2. Ex post flexibility in the weighting of multiple performance measures (Flexw)

Based on the literature, we developed a three-item Likert scale to measure *ex post* flexibility in the weighting of multiple performance measures (Ittner et al., 2003; Krishnan et al., 2005). Participants were asked to rate their agreement with statements about the weighting system their superior used to evaluate their performance, whether the weights were clearly communicated by their superior, and whether their superior could change the weights without their knowledge. Responses were rated on a five-point Likert scale ranging from "strongly disagree" (coded 1) to "strongly agree" (coded 5). The two first items were reverse coded such that higher scores indicate that superiors have more *ex post* flexibility in the weighting of multiple performance measures.

3.2.3. Perceptions of the procedural justice of performance evaluation systems (Procj)

The six-item procedural justice instrument is based on the scales developed by Leventhal (1980) and Thibaut and Walker (1975). This scale addresses employees' perceptions of fairness in the performance evaluation process and whether the superior applied this process consistently across employees and evaluation periods. Respondents were asked to rate their agreement with six statements on a scale ranging from "strongly disagree" (coded 1) to "strongly agree" (coded 5). Higher scores indicate that managers perceived the performance evaluation process as fairer.

3.2.4. Superior–manager relationship quality (Quality)

We assessed managers' perceptions of the quality of their relationship with their superior using an established three-item scale (Dulebohn & Ferris, 1999; Wayne & Ferris, 1990). Managers reported whether they enjoyed working with the superior who evaluated them, whether they had a good working relationship with their superior, and whether they trusted their superior. Responses were rated on a five-point Likert scale ranging from "strongly disagree" (coded 1) to "strongly agree" (coded 5), with higher scores indicating a better superior–manager relationship.

3.2.5. Voice opportunity (Voice)

A three-item Likert scale developed and validated by Dulebohn and Ferris (1999) was used to capture managers' voice opportunity provided by their superiors during the performance evaluation process. Participants were asked whether their

[14] The survey instrument developed by Van der Stede et al. (2006) targets managers and directors in manufacturing firms. Therefore, we adapted the original instrument to our research setting by including other objective financial and nonfinancial performance criteria currently used in other industries. We also expanded the set of subjective measures. We conducted in-depth interviews with two incentive compensation specialists from Deloitte Human Capital (Consulting) in order to validate the new criteria included and the terminology used.

[15] Our interpretation is that respondents made an effort to correctly report the diversity of performance measures used to evaluate their performance. These 14 additional criteria were classified and coded as follows: financial (cash flow, tax savings, EVA), nonfinancial customer-oriented (new product development, new market development, risk management), nonfinancial employee-oriented (outreach, fostering workgroups, career development), and nonfinancial internal operating (sustainability, new alliance/partnership).

[16] Van der Stede et al. (2006) report means for the use of financial and objective nonfinancial measures of 6 and 14, respectively. Their sample is composed of managers and directors of manufacturing firms only. Therefore, their industry-specific sample and the respondents' level of responsibility may justify the greater use of aggregate financial-, cost-, and budget-based performance measures. Moreover, their study focuses on quality-based manufacturing strategies, which may also justify the greater reported use of nonfinancial measures.

superior listened to the reasons they gave to justify their performance, whether their superior gave them opportunities to ask questions about their performance assessment, and whether their views were considered when discussing their performance rating. Responses were rated on a five-point Likert scale ranging from "strongly disagree" (coded 1) to "strongly agree" (coded 5), with higher scores indicating higher managerial perceptions of voice opportunity.

Table 1 presents the items used to measure each of the above-mentioned constructs. An exploratory factor analysis was performed separately for the items in each measurement instrument. The loadings of the scale items for each factor (Table 1, column 4) vary from 0.68 to 0.87 for perceived procedural justice, 0.76 to 0.95 for flexibility in the weighting of multiple performance measures, 0.92 to 0.94 for superior–manager relationship quality, and 0.88 to 0.90 for voice opportunity. Cronbach's coefficients are respectively 0.89, 0.85, 0.92, and 0.87. The last four columns of Table 1 present the results of a factor analysis using varimax rotation performed on all 15 items, revealing four distinct factors (i.e., procedural justice, *ex post* flexibility in the weighting of multiple performance measures, superior–manager relationship quality, and voice opportunity). This supports the convergent and discriminant validity of the measurement instruments. For subsequent analyses, we used the average scores of the measurement instrument items for each construct.

4. Results and discussion

The descriptive statistics and the Pearson's correlation coefficients between average scores of the measurement instruments items are reported in Table 2. Procedural justice (Procj) is negatively and significantly correlated with *ex post* flexibility in weighting multiple measures (Flexw), which supports H1b. However, the correlation between Procj and the use and weight of subjective measures (Wsubj) is not significant, so H1a is not supported. As expected, Table 2 also shows that

Table 1
Descriptive statistics and results of the factor analyses performed separately for each measurement instrument and for all items.[a]

Items	Mean	SD	Factor loadings	Factor loadings after varimax rotation			
				Factor 1	Factor 2	Factor 3	Factor 4
Perceptions of the procedural justice of performance evaluation systems (Procj)							
My supervisor administers the performance measurement process fairly.	3.6	0.9	0.808	**0.664**	0.232	−0.267	0.222
The performance measurement process is enforced equally among all employees.	3.4	1.1	0.786	**0.690**	0.054	−0.180	0.153
My supervisor follows different rules when dealing with different employees. (*Coding was reversed*)	3.4	1.1	0.683	**0.613**	0.066	−0.077	0.076
My supervisor applies the performance measurement process consistently to all employees.	3.6	1.0	0.849	**0.787**	0.150	−0.139	0.171
My supervisor follows fair procedures in decision making regarding performance evaluation.	3.6	0.9	0.871	**0.760**	0.231	−0.195	0.276
All employees are treated equally by my supervisor.	3.5	1.0	0.865	**0.793**	0.253	−0.119	0.148
Cronbach's coefficient alpha	0.89						
Ex post flexibility in the weighting of multiple performance measures (Flexw)							
I know *ex ante* the exact weightings of the measures that my supervisor uses to evaluate my performance. (*Coding was reversed*)	2.6	1.3	0.917	−0.140	−0.069	**0.861**	−0.077
My supervisor clearly communicates the weightings of the measures used to evaluate my performance. (*Coding was reversed*)	2.7	1.3	0.946	−0.174	−0.107	**0.968**	−0.098
The weightings of the measures used to evaluate my performance can be changed without my awareness/knowledge.	2.6	1.3	0.768	−0.315	0.065	**0.544**	−0.115
Cronbach's coefficient alpha	0.85						
Superior–manager relationship quality (Quality)							
The supervisor who evaluated me and I enjoy working together.	4.0	0.8	0.935	0.174	**0.837**	−0.038	0.280
The supervisor who evaluated me and I have a good working relationship.	4.1	0.7	0.929	0.147	**0.846**	−0.043	0.243
I trust the supervisor who evaluated me.	3.9	0.9	0.925	0.313	**0.779**	−0.045	0.323
Cronbach's coefficient alpha	0.92						
Voice opportunity (Voice)							
The supervisor that evaluated your performance listened to reasons you gave to justify your performance.	3.9	0.8	0.900	0.210	0.331	−0.079	**0.745**
The supervisor gave you an opportunity to ask questions about your performance assessment.	3.9	0.8	0.881	0.230	0.248	−0.158	**0.717**
The supervisor considered your views when discussing your performance rating.	3.8	0.9	0.900	0.248	0.286	−0.102	**0.802**
Cronbach's coefficient alpha	0.87						

[a] Numbers in bold indicate the four distinct factors.

Table 2
Descriptive statistics and correlation coefficients between measurement instrument average scores ($N = 317$).

	Mean	SD	1	2	3	4
1. Procedural justice	3.5	0.8				
2. Wsubj	32.0	26.7	−0.045			
3. Flexw	2.6	1.1	−0.402[a]	0.246[a]		
4. Quality	4.0	0.8	0.474[a]	0.057	−0.162[a]	
5. Voice	3.9	0.8	0.497[a]	0.071	−0.276[a]	0.602[a]

[a] Statistically significant at a 1% level (2-sided test).

Table 3
Descriptive statistics of demographic data (control variables) and correlations with the measurement instrument scores ($N = 317$).

	n	% or mean (±SD)	*R*-square[a] (in bold) and Pearson correlation coefficients (in italic)				
			Procedural justice	Wsubj	Flexw	Quality	Voice
Panel A – Managers' characteristics							
Gender			**0.000015**	**0.0144[c]**	**0.0079**	**0.00017**	**0.0005**
Male	219	69.1%	*−0.004*	*−0.120[c]*	*−0.089*	*−0.013*	*−0.022*
Female	98	30.9%	*0.004*	*0.120[c]*	*0.089*	*0.013*	*0.022*
Task (job position)			**0.0066**	**0.0020**	**0.0092**	**0.0229**	**0.0087**
CEO	3	1.0%	*0.015*	*−0.032*	*0.082*	*0.001*	*0.018*
CFO	6	1.9%	*−0.016*	*−0.001*	*−0.047*	*−0.069*	*−0.014*
Upper management	145	45.7%	*0.069*	*0.030*	*−0.006*	*−0.036*	*0.075*
Middle management	151	47.6%	*−0.049*	*−0.016*	*0.010*	*0.098*	*−0.051*
Missing	12	3.8%	*−0.049*	*−0.021*	*−0.019*	*−0.113[c]*	*−0.062*
Age			**0.0188**	**0.0257[c]**	**0.0202**	**0.0076**	**0.0226**
30–39	136	42.9%	*0.066*	*0.149[b]*	*0.009*	*0.075*	*0.139[c]*
40–49	100	31.5%	*−0.125[c]*	*−0.082*	*0.109*	*−0.023*	*−0.111[c]*
50–59	75	23.7%	*0.073*	*−0.060*	*−0.121[c]*	*−0.047*	*−0.023*
≥60	6	1.9%	*−0.039*	*−0.075*	*−0.027*	*−0.049*	*−0.054*
Tenure (years of experience)							
In current organization	317	11.6 (9.1)	*−0.030*	*−0.130[c]*	*−0.139[c]*	*−0.034*	*−0.096*
In other organizations	317	8.2 (7.6)	*0.079*	*0.013*	*−0.008*	*−0.019*	*0.083*
Panel B – Firm characteristics							
Sales revenues or operating budget			**0.0054**	**0.0129**	**0.0621[b]**	**0.0020**	**0.0181**
<$100 million	95	30.0%	*−0.044*	*0.088*	*0.164[b]*	*0.013*	*−0.031*
$100 million–$1 billion	77	24.3%	*−0.024*	*0.032*	*0.049*	*−0.021*	*0.024*
>$1 billion	131	41.3%	*0.072*	*−0.106*	*−0.235[b]*	*0.021*	*0.061*
Missing	14	4.4%	*−0.024*	*−0.007*	*0.097*	*−0.037*	*−0.124[c]*
Industry classification			**0.0304**	**0.0549[c]**	**0.0487[c]**	**0.0263**	**0.0357**
Manufacturing	107	33.8%	*0.092*	*0.029*	*−0.039*	*0.046*	*0.097*
Public sector	61	19.2%	*−0.108*	*0.114[c]*	*0.154[b]*	*−0.077*	*−0.149[b]*
Banking	46	14.5%	*0.066*	*−0.139[c]*	*−0.168[b]*	*−0.029*	*−0.019*
High tech/Biotech	34	10.7%	*−0.080*	*0.019*	*0.048*	*0.032*	*0.025*
Natural resources	28	8.8%	*−0.033*	*−0.004*	*−0.010*	*−0.088*	*−0.032*
Consulting	16	5.1%	*0.042*	*−0.139[c]*	*0.016*	*0.086*	*0.096*
Consumer	16	5.1%	*−0.034*	*0.002*	*0.029*	*0.067*	*−0.019*
Health/Pharmaceutical	9	2.8%	*0.046*	*0.094*	*−0.030*	*0.003*	*0.024*

[a] *R*-square was obtained using one-way ANOVA.
[b] Statistically significant at a 1% level (2-sided test).
[c] Statistically significant at a 5% level (2-sided test).

Procj is positively and significantly correlated with superior–manager relationship quality (Quality) and voice opportunity (Voice). Moreover, Wsubj and Flexw are marginally positively correlated ($r = 0.246$), and Quality and Voice are positively and moderately correlated ($r = 0.602$).

Table 3 presents the descriptive statistics for the demographic data on the final sample ($N = 317$) and the correlations with the average scores on the measurement instruments.[17] In our final sample, 69.1% of respondents are male and 30.9% female. Of these, 48.6% held upper-management positions such as chief executive officer (CEO) or chief financial officer (CFO), senior manager, or managing director, and 47.6% held middle-management positions such as chief accountant, department head, management advisor, corporate controller, or unit or division manager. Information on job position was missing for 3.8% of respondents. Respondents' age groups were from 30 to 39 years (42.9%), 40 to 49 years (31.5%), and 50 to 59 years (23.7%).

[17] The Pearson correlation coefficient between a dummy variable and a quantitative score variable is equivalent to the point-biserial correlation (Cohen et al., 2003, chap. 8, p. 308). A positive (negative) sign indicate that the mean of the instrument score in that group category represented by the dummy variable is higher (lower) than the mean of all the other groups combined. One way analysis of variance was used to obtain the *R*-square value indicating the overall amount of variation accounted for by the set of dummy variables that composed the categorical variable (Cohen et al., 2003, chap. 8, pp. 311–312).

Table 4
Results of robust regression models.[a]

Variables	Model 1		Model 2		Model 3		Model 4		Model 5	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	3.704[b]	0.169	3.651[b]	0.150	3.604[b]	0.134	3.603[b]	0.135	3.600[b]	0.131
Gender										
Male	−0.109	0.109	−0.101	0.097	−0.112	0.086	−0.076	0.088	−0.089	0.085
Female	Reference		Reference		Reference		Reference		Reference	
Task (job position)										
CEO	0.312	0.504	0.550	0.448	0.585	0.398	0.412	0.403	0.484	0.390
CFO	−0.048	0.368	−0.250	0.327	0.177	0.293	−0.003	0.294	0.163	0.287
Upper management	0.106	0.102	0.074	0.090	0.114	0.080	0.002	0.081	0.064	0.079
Middle management	Reference		Reference		Reference		Reference		Reference	
Missing	−0.316	0.257	−0.403	0.228	−0.121	0.204	−0.194	0.205	−0.057	0.200
Tenure current organization (years of experience)	0.003	0.006	−0.003	0.006	−0.002	0.005	−0.001	0.005	−0.001	0.005
Tenure other organizations	0.012	0.007	0.007	0.007	0.008	0.006	0.005	0.006	0.006	0.006
Sales										
<$100 million	−0.208	0.123	0.012	0.111	−0.112	0.099	−0.069	0.100	−0.111	0.097
$100 million–$1 billion	−0.124	0.125	−0.018	0.112	−0.054	0.099	−0.059	0.100	−0.052	0.097
>$1 billion	Reference		Reference		Reference		Reference		Reference	
Missing	−0.155	0.248	0.038	0.222	0.022	0.197	0.124	0.200	0.117	0.193
Industry										
Manufacturing	Reference		Reference		Reference		Reference		Reference	
Public sector	−0.308[c]	0.147	−0.163	0.132	−0.069	0.117	−0.023	0.119	−0.028	0.115
Banking	0.007	0.156	−0.005	0.139	0.049	0.124	0.038	0.125	0.043	0.121
High tech/Biotech	−0.301	0.170	−0.160	0.151	−0.209	0.134	−0.123	0.136	−0.200	0.131
Natural resources	−0.224	0.181	−0.160	0.160	0.033	0.143	−0.007	0.144	0.047	0.140
Consulting	0.179	0.230	0.185	0.207	0.030	0.184	0.014	0.186	−0.031	0.180
Consumer	−0.266	0.230	−0.143	0.204	−0.234	0.181	−0.013	0.183	−0.159	0.178
Health/Pharmaceutical	0.095	0.295	0.037	0.262	0.133	0.233	0.133	0.236	0.151	0.228
Wsubj			0.0024	0.0017	0.0004	0.0015	−0.0002	0.0015	−0.0004	0.0015
Flexw			−0.341[b]	0.041	−0.230[b]	0.037	−0.218[b]	0.038	−0.192[b]	0.037
Quality					0.519[b]	0.052			0.365[b]	0.063
Voice							0.526[b]	0.053	0.282[b]	0.063
N	317		317		317		317		317	
R-square	0.054		0.199		0.339		0.315		0.368	
Maximum VIF	1.50		1.53		1.54		1.54		1.80	

[a] The independent variables Wsubj, Flexw, Quality, and Voice are centred in the regression models.
[b] Statistically significant at a 1% level (2-sided test).
[c] Statistically significant at a 5% level (2-sided test).

They had an average of 11.6 years of experience in their current organization and 8.2 years of experience in other organizations. About 75% had more than two years of experience in their current position, 72% had held at least one other position within the current organization, and 88% had held a previous position at another organization. These statistics suggest that the average respondent in our final sample ($N = 317$) is an experienced senior manager.

In terms of organization size, 65.6% of respondents worked for firms with sales revenues (or operating budgets) greater than $100 million. Respondents were distributed among the following economic sectors: 33.8% in manufacturing, 19.2% in the public sector, 14.5% in banking, 10.7% in high technology or biotechnology, 8.8% in natural resources, 5.1% in consulting, 5.1% in the consumer sector, and 2.8% in the health or pharmaceutical industry.

Of the demographic variables, only the 40–49 age group is significantly correlated to Procj at the 5% level; managers aged 40–49 have on average significantly lower Procj scores than managers in the other age groups (the correlation coefficient of −0.125 is negative and significantly different than zero), which accounts for $(-0.125)^2 = 0.0156$ of the variance in Procj. A few other correlation coefficients in Table 3 differ significantly from zero (between the demographic variables and Wsubj, Flexw, Quality, and Voice), but their magnitude is relatively small, at from −0.235 to 0.164. Gender, age, and industry account respectively for 0.0144, 0.0257, and 0.0549 of the variance in Wsubj; sales revenues or operating budget accounts for 0.0621 of the Flexw variance; and industry accounts for 0.0487. Although certain demographic variables show a statistically significant but low correlation coefficient with Quality or Voice, none of the global *R*-squared values between these variables and Quality or Voice are significantly greater than zero. Overall, our dataset provides a rich research setting, with detailed information on a large cross-sectional sample of middle- and upper-level managers.

To test our research hypotheses, we considered several linear regression models built in a stepwise fashion, and results are presented in Tables 4 and 5. Residuals analyses and diagnostic statistics across the different models revealed five outlying observations, with one having a stronger influence on the regression coefficient estimates obtained with ordinary least squares (OLS). We therefore performed additional analyses using robust regression[18] and OLS without the five outlying

[18] We used the ROBUSTREG procedure available in SAS for Windows version 9.2 with the MM estimation method.

Table 5
Results of the robust regression models[a] with interaction terms.

Parameters	Model 6		Model 7		Model 8		Model 9	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	3.643[b]	0.133	3.592[b]	0.131	3.611[b]	0.128	3.643[b]	0.128
Gender								
Male	−0.105	0.086	−0.087	0.085	−0.097	0.083	−0.128	0.083
Female	Reference		Reference		Reference		Reference	
Task (job position)								
CEO	0.566	0.394	0.436	0.390	0.511	0.379	0.552	0.379
CFO	0.208	0.293	0.152	0.285	0.239	0.282	0.354	0.289
Upper management	0.092	0.079	0.016	0.078	0.052	0.077	0.034	0.077
Middle management	Reference		Reference		Reference		Reference	
Missing	−0.157	0.203	−0.168	0.199	−0.107	0.195	−0.141	0.201
Tenure current organization (years of experience)	−0.002	0.005	−0.001	0.005	−0.001	0.005	0.001	0.005
Tenure other organizations	0.009	0.006	0.007	0.006	0.008	0.006	0.008	0.005
Sales								
<$100 million	−0.107	0.098	−0.098	0.097	−0.116	0.094	−0.140	0.095
$100 million–$1 billion	−0.070	0.098	−0.061	0.097	−0.097	0.095	−0.099	0.095
>$1 billion	Reference		Reference		Reference		Reference	
Missing	0.003	0.196	0.060	0.194	0.104	0.192	0.096	0.195
Industry								
Manufacturing	Reference		Reference		Reference		Reference	
Public sector	−0.073	0.116	−0.042	0.115	−0.041	0.112	−0.084	0.115
Banking	0.011	0.123	0.024	0.121	0.020	0.118	0.017	0.116
High tech/Biotech	−0.229	0.133	−0.140	0.131	−0.188	0.128	−0.171	0.130
Natural resources	0.007	0.142	0.004	0.140	0.009	0.137	0.059	0.140
Consulting	0.009	0.182	0.015	0.181	0.001	0.176	0.014	0.172
Consumer	−0.200	0.180	−0.012	0.177	−0.145	0.173	−0.201	0.171
Health/Pharmaceutical	0.069	0.231	0.085	0.229	0.086	0.222	0.058	0.213
Wsubj	0.0002	0.0015	−0.0014	0.0015	−0.0008	0.0014	−0.0014	0.0017
Flexw	−0.249[b]	0.037	−0.213[b]	0.037	−0.196[b]	0.036	−0.215[b]	0.040
Wsubj*Flexw							0.002	0.001
Quality	0.482[b]	0.052			0.338[b]	0.062	0.310[b]	0.068
Wsubj*Quality	0.002	0.002			0.001	0.002	−0.002	0.003
Flexw*Quality	0.115[b]	0.042			0.163[b]	0.048	0.180[b]	0.054
Wsubj*Flexw*Quality							0.005[c]	0.002
Voice			0.590[b]	0.052	0.334[b]	0.062	0.396[b]	0.068
Wsubj*Voice			0.006[b]	0.002	0.0038	0.002	0.008[b]	0.003
Flexw*Voice			0.021	0.042	−0.129[b]	0.049	−0.182[b]	0.052
Wsubj*Flexw*Voice							−0.004[d]	0.002
Quality*Voice							−0.151[c]	0.060
Wsubj*Quality*Voice							−0.0006	0.003
Flexw*Quality*Voice							0.094[d]	0.053
WsubJ*Flexw*Quality*Voice							−0.001	0.002
N	317		317		317		317	
R-square	0.352		0.324		0.397		0.404	
Maximum VIF	1.55		1.55		1.88		2.34	

[a] The independent variables Wsubj, Flexw, Quality, and Voice are centred in the regression models.
[b] Statistically significant at a 1% level (2-sided test).
[c] Statistically significant at a 5% level (2-sided test).
[d] Statistically significant at a 10% level (2-sided test).

observations. The main findings were very similar for the three estimation methods, but slightly more conservative with robust regression. We consequently report the robust regression results only (Cohen et al. 2003, p. 417–419).

Model 1 in Table 4 includes control variables only. Respondents' age was excluded due to high correlation with years of experience.[19] On average, managers working in the public sector have significantly lower perceptions of procedural justice than managers working in the manufacturing sector. However, this difference is not statistically significant in all other regression models that include the variables of interest: Wsubj, Flexw, Quality, Voice, and their interactions (see Tables 4 and 5, Models 2–9). None of the other control variables is significantly associated with our dependent variable Procj in all regression models.[20]

[19] Due to the high correlation between age and years of experience, we did not include age in the regression models. In the regression model including only the control variables, the variance inflation factors (VIF) for the age categories were between 11 and 20, and 3.8 for years of experience in current organization and 3.0 for years of experience in other organisations. These VIF values above 10 for the age dummy variables are indicating strong multi-collinearity with the other control variables, namely with years of experience.

[20] Results of the regression models for the variables of interest (i.e., Wsubj, Flexw, Quality, Voice, and their interactions) are very similar when we removed all control variables in models 2–9.

Model 2 in Table 4 tests hypotheses H1a and H1b. Recall that H1a predicts a negative association between Wsubj and Procj, whereas H1b predicts a negative relationship between Flexw and Procj. Only Flexw is statistically significant and negatively associated with Procj ($p < 0.01$). We note also that the addition of Flexw and Wsubj to the regression model increases the *R*-squared value by 15%, from 5% to 20%, and this increase is solely due to Flexw. Hence, the second regression model supports Hypothesis H1b but not H1a. Thus, the non-significant coefficient of Wsubj suggests that the use of subjective measures in performance evaluation *per se* does not affect perceptions of procedural justice (H1a). The significant negative coefficient of Flexw suggests that greater *ex post* flexibility in the weighting of multiple performance measures decreases perceptions of procedural justice (H1b).

Before testing H2 and H3, we first added quality of the superior–manager relationship (Quality) and voice opportunity (Voice) separately and together to the regression model (Table 4, Models 3, 4, and 5). As expected from the correlation coefficients in Table 3, the regression coefficients for both Quality and Voice are positive and statistically significant. Quality and Voice increase the *R*-squared value of the regression model with Wsubj and Flexw from 0.199 to 0.339 and 0.315, respectively (see Models 3 and 4). Adding Quality and Voice simultaneously to the regression model increases the *R*-square from 0.199 to 0.368 (see Model 5). Although Quality and Voice are correlated, we observe only a slight decrease in the magnitude of their regression coefficients in Model 5 compared to Models 3 and 4, but both remain positive and statistically significant. Furthermore, the variance inflation factors are 1.69 for Quality and 1.80 for Voice in Model 5. These results indicate that there is no multicollinearity problem when all variables are included in the same regression model. It is worth noting that the coefficient of Flexw on Procj remains negative and significant even after including Quality and Voice separately (Models 3 and 4) and simultaneously (Model 5), which provides strong support for H1b. To test H2 and H3, we therefore added the two-way interactions Wsubj by Quality and Flexw by Quality to Model 3 (Model 6 in Table 5), Wsubj by Voice and Flexw by Voice to Model 4 (Model 7 in Table 5), and these four two-way interactions to Model 5 (Model 8 in Table 5). In order to remove nonessential multicollinearity from the regression models, we centred these four independent variables in all regression models (Cohen et al., 2003; pp. 261–268).

The regression coefficients for Flexw, Quality, and the interaction Flexw by Quality are statistically significant in Models 6 and 8. The regression coefficients for Flexw, Voice, and the interaction Wsubj by Voice are statistically significant in Model 7, but not the interaction Flexw by Voice. However, Flexw, Voice, and the interaction Flexw by Voice are statistically significant in Model 8, but not the interaction Wsubj by Voice. Model 8 includes our two measures of subjectivity in performance evaluation (Wsubj and Flexw) and the two contextual factors with the four two-way interaction terms to test hypotheses 2 and 3. The *R*-square for Model 8 is the highest, at 39.7%, and the maximum inflation factor value of 1.88 is very low, which suggests that there is no multicollinearity problem. Except for the two-way interactions between Wsubj by Voice and Flexw by Voice in Model 7, the regression coefficients in Model 8 are consistent with all previous models, and it provides the best explanation for the variance in Procj. We therefore retain Model 8 to test hypotheses H2 and H3. As mentioned above, the two-way interactions Flexw by Quality and Flexw by Voice are statistically significant in Model 8. To visualize these significant interactions, Figs. 1 and 2 show the relationship between Procj and Flexw at three different levels of Quality and Voice, respectively (i.e., at one SD below the mean, at the mean, and at one SD above the mean). All other variables are set to the mean or the reference category. As shown in Table 5, Model 8, the regression coefficient for the interaction between Flexw and Quality is positive. The positive effect of Quality is corroborated by Fig. 1: at higher quality of the superior–manager relationship we observe a weaker negative association between *ex post* flexibility in weighting and perceptions of procedural justice, supporting Hypothesis H2b. However, the regression coefficient for the interaction between Flexw and Voice is negative. Therefore, the greater the voice opportunity, the

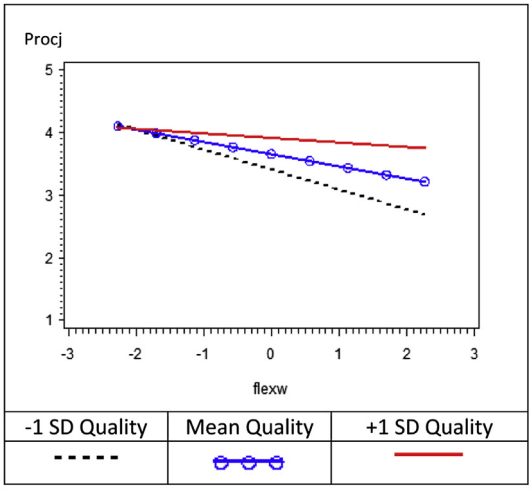


Fig. 1. Two-way interaction between Flexw and Quality (Model 8).

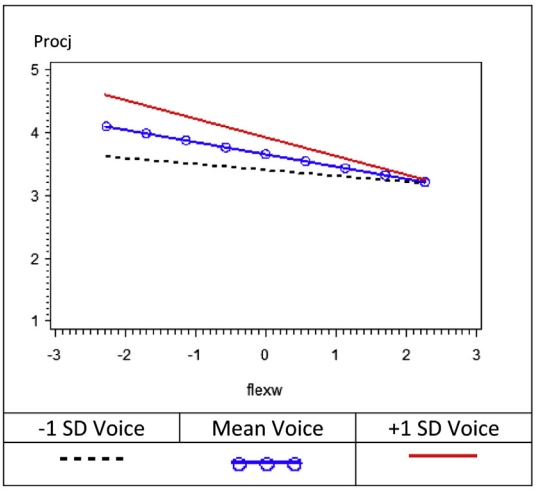


Fig. 2. Two-way interaction between Flexw and Voice (Model 8).

stronger the negative association between *ex post* flexibility in the weighting of multiple performance measures and perceptions of procedural justice in the performance evaluation (Fig. 2). This finding suggests that voice opportunity amplifies the adverse effect of *ex post* flexibility on perceived procedural justice, which goes against Hypothesis H3b. Moreover, the regression results for Wsubj and for the interactions Wsubj by Quality and Wsubj by Voice are not statistically significant, which provides no support for H2a or H3a. Consequently, we do not present the corresponding figures.

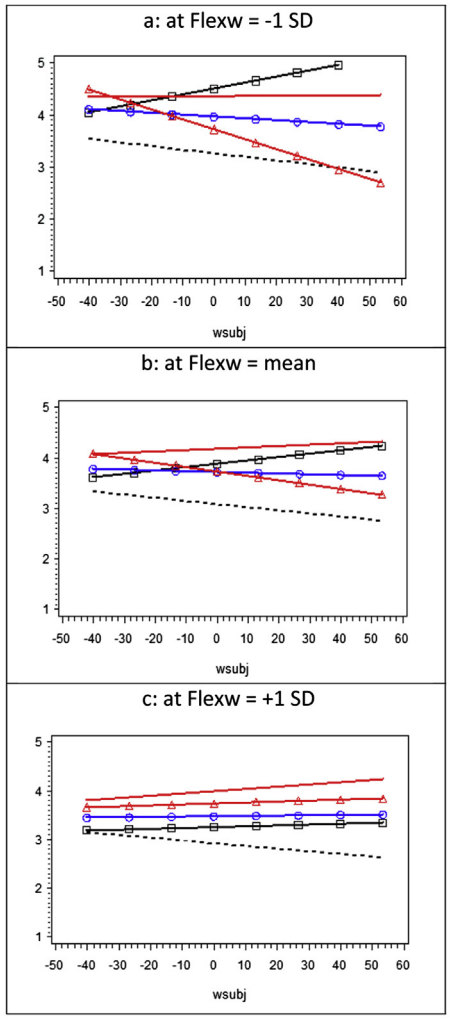
Model 8 in Table 5 assumes that all other two-way and higher-order interactions between Wsubj, Flexw, Quality, and Voice are negligible. Although we did not formulate hypotheses about these high-order interactions, this assumption should be verified to ensure that potentially significant findings on the examined variables are not overlooked. We therefore added these additional interaction effects to regression Model 8. In Table 5, Model 9, in addition to the previously reported statistically significant regression coefficients (Model 8), we find the following significant interactions: Wsubj by Flexw by Quality, Wsubj by Voice, and Quality by Voice. We also note that Wsubj by Flexw by Voice and Flexw by Quality by Voice are statistically significant at the 10% level. These further significant two- and three-way interactions imply that the association of Wsubj and Flexw with procedural justice is conditional on not only Quality and Voice, but also on Wsubj and Flexw. Hence, our regression analysis suggests a slightly more complex relationship with perceptions of procedural justice than the two-way interactions proposed in H2 and H3. However, although some of the added two- and three-way interaction terms are statistically significant, they account for a very small additional proportion of the variance in Procj, as the *R*-squared value for Model 9 is 40.4%, which is less than 1% greater than the *R*-square for Model 8.

To describe these more complex interactions between Wsubj, Flexw, Quality, and Voice in Model 9, we followed the recommendation of Cohen et al. (2003, p. 269) and plotted the predicted values of perceived procedural justice with Wsubj at different selected values of Quality, Voice, and Flexw (Fig. 3a–c) and with Flexw at different values of Quality, Voice, and Wsubj (Fig. 4a–c). In all cases, the categorical control variables are set at the reference value and years of experience in current and other organizations are set at the mean value. In the following paragraphs, we briefly explain the evidence provided by these figures and we interpret them in light of our research hypotheses.

Taken together, Fig. 3a–c show that the relationship between Wsubj and Procj is negative mainly at lower Voice and Flexw. This can be observed in the slope, where Voice is one SD below the mean in Fig. 3a–c. With increasing Voice, the relationship between Wsubj and Procj becomes less negative, then changes to become more positive. Comparing these figures, we also observe a positive change in the slope for Voice at one SD below the mean versus Voice at one SD above the mean for the same level of Quality. This supports Hypothesis H3a, with the additional finding that the changes in the relationship between Wsubj and Procj with higher Voice are greater when Flexw is lower.

Fig. 3a–c also suggest that as Quality increases, the relationship between Wsubj and Procj becomes more negative (at lower Voice) or less positive (at higher Voice) when Flexw is low (Fig. 3a). In contrast, when Flexw is high (Fig. 3c), as Quality increases, the relationship between Wsubj and Procj becomes less negative (at lower Voice) or more positive (at higher Voice). In other words, Quality appears to mitigate the negative effects of Wsubj on Procj only when Wsubj is combined with greater Flexw. We interpret this evidence as partial support for H2a. These disordinal interactions (Luft & Shields, 2003) explain why the main effect of Wsubj is not statistically significant in the regression models (Tables 4 and 5).

Overall, Fig. 4a–c show that the association between Flexw and Procj is negative except when Voice is low and Quality and Wsubj are high. More specifically, we observe that the relationship between Flexw and Procj becomes positive when Voice is one SD below the mean and Quality is one SD above the mean in Fig. 4b, where Wsubj is at the mean, and in Fig. 4c, where Wsubj is one SD above the mean. As Quality increases, the relationship between Flexw and Procj becomes less negative.



The independent variables WSUBJ, FLEXW, Quality, and Voice are centred in the regression models.

-1 SD Quality -1 SD Voice	-1 SD Quality +1 SD Voice	Mean Quality Mean Voice	+1 SD Quality -1 SD Voice	+1 SD Quality +1 SD Voice
- - - - -	▭▭▭	○○○	△△△	———

Fig. 3. Relationship between procedural justice and Wsubj according to Model 9.

Moreover, comparing slopes across Fig. 4a–c, as Wsubj increases, we observe a clear change in slope when Quality is one SD below the mean versus Quality at one SD above the mean for the same level of Voice. We interpret these results as support for H2b, with the additional finding that the effect of Quality on the association between Flexw and Procj is greater when Wsubj is higher. Although this finding indicates a disordinal interaction (Luft & Shields, 2003), they corroborate the results of Models 2 to 5, indicating an overall negative effect of Flexw on Procj.

In sum, Fig. 4a–c suggest that when the contextual factors Quality and Voice are at the same magnitude (high–high, low–low, mean–mean), they play a limited interaction role in the association between Flexw and Procj at all levels of Wsubj. However, an interaction effect is seen when Quality and Voice are at opposite magnitude (high–low, low–high), suggesting that Quality may compensate for lack of Voice in mitigating the negative effect of Flexw on Procj, but only when combined with greater Wsubj.[21]

Our interpretation of these results is that, combined with greater use and weight of subjective measures, superior–manager relationship quality can mitigate the negative effect of Flexw on Procj, even when voice opportunity is low. In our sample, superior–manager relationship quality appears to compensate for the adverse effect of Flexw on Procj, whereas

[21] In order to avoid repetition, this discussion focuses on the association between Flexw and Procj (Fig. 4) but, similar effects are observed through the association between Wsubj and Procj, presented in Fig. 3.

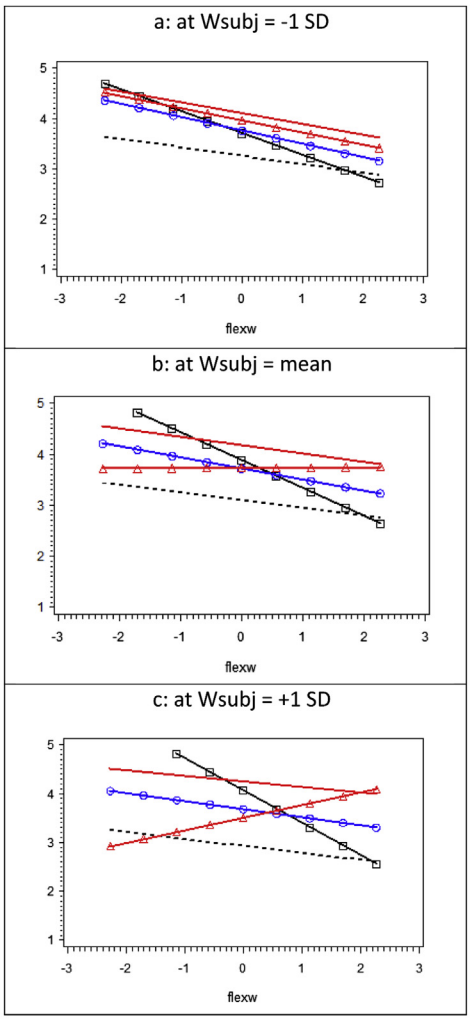


Fig. 4. Relationship between procedural justice and Flexw according to Model 9.

Voice opportunity appears to strengthen this effect. This finding extends the previous literature on fairness perceptions by corroborating the argument that, at higher hierarchy levels, a good superior–manager relationship and trust are more important than formal mechanisms such as voice opportunity "justifications" (Hartman & Slapnicar, 2012; Kennedy et al., 2009). Moreover, as Shaw et al. (2003) contend, voice opportunity—which can take the form of explanations or justifications—can adversely affect perceptions of justice when individuals feel that their opinions are ignored. We believe that this occurs when poor relationship quality is combined with superiors' greater use of subjectivity, as captured by the greater use and weight of subjective measures and *ex post* flexibility in weighting multiple performance measures (Fig. 4c).

In summary, Model 9 partially supports Hypothesis H2a, supports Hypotheses H2b and H3a but with additional findings, and does not support Hypothesis H3b. However it is important to note that the results of Model 8 are simpler and more robust, and they support Hypothesis H2b only.

5. Conclusions

In this study, we investigate the associations between two forms of subjectivity in performance evaluation—the use and weight of subjective performance measures and *ex post* flexibility in weighting multiple performance measures—and their effects on managerial perceptions of procedural justice. We also examine whether these associations are affected by two contextual factors: quality of the superior–manager relationship and voice opportunity.

Our findings suggest that, in our sample, although the use and weight of subjective measures does not show an overall significant effect on managers' perceptions of procedural justice, the superior's use of *ex post* flexibility in weighting multiple performance measures adversely affects managers' perceptions of procedural justice. More importantly, the two forms of subjectivity interact differentially with each of the two contextual factors (quality of the superior–manager relationship and voice opportunity), suggesting that the effects on procedural justice cannot be examined in isolation. More specifically, our main results indicate that superior–manager relationship quality mitigates the negative effects of *ex post* flexibility in weighting multiple performance measures on procedural justice, whereas voice opportunity amplifies this negative effect.

Our findings have important implications for both the accounting literature and the research on organizational justice. First, our results are consistent with analytical models that incorporate the use of subjectivity in managerial contracting (e.g., Baker et al., 1994; Prendergast & Topel, 1993), and they extend the management accounting research on the effects of performance measure properties on perceived performance evaluation fairness (Hartmann & Slapnicar, 2012; Kennedy et al., 2009; Lau & Moser, 2008). More importantly, our study contributes to the accounting literature on incentives and contract design, because it jointly investigates two forms of subjectivity in performance evaluation and two contextual factors. The primary contribution is therefore a deeper understanding of how different forms of subjectivity in performance evaluation are used and how they interact with contextual factors to impact evaluees' perceptions of fairness.

Second, our findings are meaningful for organizations because they emphasize how subjectivity can be incorporated into the design of performance evaluation systems. Adding to previous empirical evidence on the cognitive limitations of evaluators (Bol & Smith, 2011; Lipe & Salterio, 2000) and the fact that subjectivity leads to bias in performance evaluations (Moers, 2005), our study sheds new light on the trade-off between the incentive benefits and the perceived unfairness of subjectivity in performance evaluations. Hence, with an increase in *ex post* flexibility, the quality of the relationship appears to be more effective than voice opportunity in enhancing perceptions of procedural justice. At first glance, a superior's *ex post* flexibility may lead managers to believe that they can better influence their performance evaluation through voice opportunity. However, this may not be the case when superiors have high *ex post* flexibility and the quality of the superior–manager relationship is low. These results are in line with the findings on influence tactics in perceptions of performance evaluation and incentive systems (Dulebohn & Ferris, 1999; Maas et al., 2012; Woods, 2009, p. 45).

In addition, our results underscore the limitations of scorecard-type performance evaluation systems, which feature multiple performance measures, greater use of subjective evaluations, and flexible weighting (Bol & Smith, 2011; Moers, 2005; Ittner et al., 2003). We demonstrate the potential for a trade-off between contracting benefits and perceptions of procedural justice when subjectivity is introduced into the performance evaluation process. Although a positive association between perceptions of procedural justice and job satisfaction or motivation has been established in the literature, our study demonstrates that subjectivity may adversely affect the effectiveness of performance evaluation through lower perceptions of procedural justice.

Third, our findings also corroborate recurrent themes in the management accounting literature. Whereas the contracting literature proposes that subjectivity in performance evaluation can be effective in correcting incentive distortions caused by noisy or incomplete objective performance measures, our results suggest that perceptions of unfairness may offset these benefits. Thus, our results advance the research on the proactive dimension of procedural justice, which aims to identify key antecedents of perceived fairness of the performance evaluation (Lau & Moser, 2008; Lau, Wong, & Eggleton, 2008). Finally, in line with Giraud et al.'s (2008) field evidence, we demonstrate that the procedures used to solve managerial incentive problems, such as outcome measurability and uncontrollable factors, may give rise to other undesirable effects, such as employee perceptions of unfairness.

As with any empirical investigation, ours has some limitations. First, the survey questionnaire may have resulted in an unknown sample bias that limits the generalizability of our findings (Dillman et al., 2009). Nevertheless, we sought to minimize these effects by carefully designing our survey instrument and its administration and by assessing potential bias in the collected data. Second, although our sampling strategy contributes to the literature because it comprises experienced, multi-task managers occupying decision-making positions, their expertise and experience in performance evaluation could have influenced their perceptions of procedural justice. Third, because we wanted to assure confidentiality and reliable responses regarding justice perceptions, our survey instrument did not ask the respondents to identify their organization. Therefore, the sample may include more than one respondent from the same organization, which could have introduced some noise into the analysis. Furthermore, two of our key variables were purpose-developed, and could benefit from further validation in future studies. Future studies could also more deeply explore the complementarities between other forms of subjectivity in performance evaluation and determine their effects on other dimensions of organizational justice, such as distributive justice. Interactions between different forms of subjectivity and other features of performance evaluation systems, such as goal participation and perceived system knowledge, are also promising avenues for future research.

Acknowledgements

The authors are grateful to the editor and the two anonymous referees for their valuable comments and insightful suggestions. We would also like to thank Donna Bobek, Anne Fortin, Tom Groot, Frank Hartmann, Michal Matejka, Sergeja Slapnicar, and two anonymous referees at the 2010 MAS Annual Meeting as well as participants at the 2010 European Accounting Association Meeting and the 5th Conference on Performance Measurement and Management Control for their insightful comments. The authors gratefully acknowledge funding provided by the Roland-Chagnon Professorship of Accounting and Taxation, the Fonds Québécois de Recherche sur la Société et la Culture (FQRSC), and the CMA Canada/CAAA Research Grant Program.

Appendix A

Survey instrument used to capture organizational performance measurement practices

Please **check all measures** that are regularly used to evaluate your performance, and **indicate the relative weight** that you believe your supervisor gives to each performance measure category. The relative **weights** for all 5 categories should add up to **100%**.

1.1.1 Financial performance measures — **Weight (%)**

- Stock price-related measures
- Asset deployment (e.g., return on capital invested)
- Total sales or sales growth
- Bottom-line financial measures of organization's actual performance (e.g., accounting profits, net income)
- Gross margin or contribution margin
- Operating costs
- Budget-based measures
- Other measures (please specify)

1.1.2 Nonfinancial performance measures

a. Nonfinancial / Customer-oriented — **Weight (%)**

- Market share
- Time to fill customer orders
- Delivery performance (e.g., on-time delivery, percent of correct deliveries)
- Time to respond to customer problems
- Flexibility/responsiveness (e.g., products features/services offered)
- Customer satisfaction (e.g., customer survey results, % of customer complaints)
- Customer acquisition (e.g., number of new customers, % of new customer sales)
- Customer retention/loyalty (e.g., number of repeat customers)
- Other measures (please specify)

b. Nonfinancial / Employee-oriented — **Weight (%)**

- Employee satisfaction (e.g., employee survey results, number of grievances filed)
- Employee initiative (e.g., suggestions submitted)
- Employee productivity (e.g., number of hours worked or billed)
- Health and safety (e.g., number of incidents, number of injuries)
- Employee training (e.g., hours in training, employees' time allocated for training)
- Employee loyalty/turnover (e.g., years in the job, years with the organization)
- Absenteeism
- Other measures (please specify)

c. Nonfinancial / Internal operating — **Weight (%)**

- Production capacity
- Labour productivity (e.g., hours used/hours available, overtime hours)
- Machine productivity (e.g., running hours/available hours)
- Material usage (e.g., material usage inefficiency)
- Process cycle time (e.g., total process time)
- Inventory management (e.g., turnover)
- Quality control (e.g., defects, % of rework)
- Innovation (e.g., number of new products/projects)
- Other measures (please specify)

The four categories listed above deal with objective performance measures that are usually quantified and reported on a regular basis. However, performance evaluation may include **subjective assessments** of various, not always clearly specified, aspects of your performance. Which of the following subjective criteria do you believe your supervisor makes an assessment of when evaluating your performance?

1.1.3 Subjective performance criteria — **Weight (%)**

- My long-term perspective on the business
- My ability to effectively acquire new skills/knowledge
- My ability to effectively develop core competencies
- My communication skills
- My willingness to share knowledge within the organization
- My cooperation with other departments within the organization
- Employee spirit/ morale in my department
- My management style/leadership skills
- My loyalty to the organization
- The extent to which I embody and support the organization's values
- Other factors (please specify)

References

Avery, D. R., & Quinones, M. A. (2004). Individual differences and the voice effect: the moderating role of value of voice. *Group & Organization Management, 29*(1), 106–124.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics, 109*(4), 1125–1156.

Banker, R., & Datar, S. (1989). Sensitivity, precision and linear aggregation of signals for performance evaluation. *Journal of Accounting Research, 27*, 21–39.

Bol, J. C. (2008). Subjectivity in compensation contracting. *Journal of Accounting Literature, 27*, 1–24.

Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: bias and the asymmetric influence of controllability. *The Accounting Review, 86*(4), 1213–1230.

Bushman, R. M., Indjejikian, R. J., & Smith, A. (1996). CEO Compensation: the role of individual performance evaluation. *Journal of Accounting and Economics, 21*, 161–193.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. New Jersey: John Wiley & Son, Inc.

Dulebohn, J. H., & Ferris, G. R. (1999). The role of influence tactics in perceptions of performance evaluations' fairness. *Academy of Management Journal, 42*(3), 288–303.

Feltham, G., & Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review, 69*, 429–453.

Fisher, J. G., Maines, L. A., Peffer, S. A., & Sprinkle, G. B. (2005). An experimental investigation of employer discretion in employee performance evaluation and compensation. *The Accounting Review, 80*(2), 563–583.

Folger, R., & Cropanzano, R. (1998). *Organizational justice and human resource management*. Beverly Hills, CA: Sage.

Fortin, M. (2008). Perspectives on organizational justice: concept clarification, social context integration, time and links with morality. *International Journal of Management Reviews, 10*(2), 93–126.

Fulk, J., Brief, A. P., & Steve, H. B. (1985). Trust-in-supervisor and perceived fairness and accuracy of performance evaluations. *Journal of Business Research, 13*, 299–313.

Gibbs, M., Merchant, K. A., Van der Stede, W. A., & Vargus, M. E. (2004). Determinants and effects of subjectivity in incentives. *The Accounting Review, 79*(2), 409–436.

Giraud, F., Langevin, P., & Mendoza, C. (2008). Justice as a rationale for the controllability principle: a study of managers' opinions. *Management Accounting Research, 19*, 32–44.

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

Hartmann, F., & Slapnicar, S. (2009). How formal performance evaluation affects trust between superior and subordinate managers. *Accounting, Organizations and Society, 34*, 722–737.

Hartmann, F., & Slapnicar, S. (2012). The perceived fairness of performance evaluation: the role of uncertainty. *Management Accounting Research, 23*, 17–33.

Herath, H. S. B., Bremser, W. G., & Brinberg, J. G. (2010). Joint selection of balanced scorecard targets and weights in a collaborative setting. *Journal of Accounting and Public Policy, 29*, 45–59.

Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics, 10*, 74–91.

Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: evidence from a balanced scorecard. *The Accounting Review, 78*(3), 725–758.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Boston, MA.

Kennedy, F. A., Kohlmeyer, J. M., III, & Parker, R. J. (2009). The roles of organizational justice and trust in a gain-sharing control system. *Advances in Accounting Behavioral Research, 12*, 1–23.

Krishnan, R., Luft, J. L., & Shields, M. D. (2005). Effects of accounting-method choices on subjective performance-measure weighting decisions: experimental evidence on precision and error covariance. *The Accounting Review, 80*(4), 1163–1192.

Lau, C. M., & Moser, A. (2008). Behavioral effects of nonfinancial performance measures: the role of procedural fairness. *Behavioral Research in Accounting, 20*(2), 55–71.

Lau, C. M., Wong, K. M., & Eggleton, I. R. C. (2008). Fairness of performance evaluation procedures and job satisfaction: the role of outcome-based and non-outcome-based effects. *Accounting and Business Research, 38*(2), 121–135.

Leventhal, G. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationships. In P. Press, K. Gergen, M. Greenberg, & R. Willis (Eds.), *Social exchanges: Advances in theory and research* (pp. 27–55).

Libby, T. (1999). The influence of voice and explanation on performance in a participative budgeting setting. *Accounting, Organizations and Society, 24*, 125–137.

Lindquist, T. M. (1995). Fairness as antecedent to anticipative budgeting: examining the effects of distributive justice, procedural justice and referent cognitions on satisfaction and performance. *Journal of Management Accounting Research, 7*, 122–147.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York: Plenum.

Lipe, M. L., & Salterio, S. E. (2000). The balanced scorecard: judgmental effects of common and unique performance measures. *The Accounting Review, 75*(3), 283–298.

Luft, J., & Shields, M. D. (2003). Mapping management accounting: graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations and Society, 28*, 169–249.

Maas, V. S., van Rinsum, M., & Towry, K. L. (2012). In search of informed discretion: an experimental investigation of fairness and trust reciprocity. *The Accounting Review, 87*(2), 617–644.

McFarlin, D. B., & Sweeney, P. D. (1992). Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes. *Academy of Management Journal, 35*(3), 626–637.

Moers, F. (2005). Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society, 30*, 65–80.

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

Nisar, T. M. (2007). Evaluation of subjectivity in incentive pay. *Journal of Financial Services Research, 31*(1), 53–73.

Otley, D. (1999). Performance management: a framework for management control systems research. *Management Accounting Research, 10*(4), 363–382.

Prendergast, C., & Topel, R. H. (1993). Discretion and bias in performance evaluation. *European Economic Review, 37*, 355–365.

Prendergast, C., & Topel, R. H. (1996). Favoritism in organizations. *Journal of Political Economy, 104*(5), 958–978.

Rupp, D. E., & Cropanzano, R. (2002). The mediating effects of social exchange relationships in predicting workplace outcomes from multifoci organizational justice. *Organizational Behavior and Human Decision Processes, 89*, 925–946.

Shaw, J. C., Wild, E., & Colquitt, J. A. (2003). To justify or excuse?: a meta-analytic review of the effects of explanations. *Journal of Applied Psychology, 88*(3), 444–458.

Thibaut, J., & Walker, L. (1975). *Procedural justice*. Hillsdale, NJ: Lawrence Erlbaum.

Van der Stede, W. A., Chow, C. W., & Lin, T. W. (2006). Strategy, choice of performance measures, and performance. *Behavioral Research in Accounting, 18*, 185–205.

Van der Stede, W. A., Young, S. M., & Chen, X. C. (2005). Assessing the quality of evidence in empirical management accounting research: the case of survey studies. *Accounting, Organizations and Society, 30*(7/8), 655–684.

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

Woods, A. (2009). *Subjective adjustments to objective performance measures: An empirical examination in complex work settings*. Working paper 2009 MAS Mid-year Meeting.