



VU Research Portal

When one size does not fit all

Anderson, Shannon W.; Dekker, Henri C.; Sedatole, Karen L.; Wiersma, Eelke

published in

Management Accounting Research
2020

DOI (link to publisher)

[10.1016/j.mar.2020.100706](https://doi.org/10.1016/j.mar.2020.100706)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Anderson, S. W., Dekker, H. C., Sedatole, K. L., & Wiersma, E. (2020). When one size does not fit all: Using *ex post* subjective ratings to provide parity in risk-adjusted compensation. *Management Accounting Research*, 49, 1-12. [100706]. <https://doi.org/10.1016/j.mar.2020.100706>

General rights

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

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

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl



When one size does not fit all: Using *ex post* subjective ratings to provide parity in risk-adjusted compensation



Shannon W. Anderson^a, Henri C. Dekker^{b,d}, Karen L. Sedatole^c, Eelke Wiersma^{b,*}

^a University of California, Davis, Graduate School of Management, 540 Alumni Ln, Davis, CA 95616, United States

^b Vrije Universiteit Amsterdam, School of Business and Economics, De Boelelaan 1105, 1081HV, Amsterdam, the Netherlands

^c Emory University, Goizueta Business School, 1300 Clifton Rd, Atlanta, GA, 30322, United States

^d Copenhagen Business School, Department of Accounting, Solbjerg Plads 3, C4 2000, Frederiksberg Denmark

ARTICLE INFO

JEL classification:

M40
M52

Keywords:

Subjective evaluation
Performance measurement
Compensation risk
Subjectivity

ABSTRACT

Firms typically use a ‘one-size-fits-all’ (OSFA) compensation contract that specifies a common formulaic relation between performance and compensation (i.e., a performance bonus) for non-executive managers in similar jobs. However, a contract that is appropriate on average, may be suboptimal for individual managers if heterogeneity in the operating environment creates varying compensation risk. We use field data from a retail firm that introduced an OSFA bonus compensation plan for its store managers. The common bonus formula is based on a weighted sum of objective measures of performance and a subjective rating made by supervisors. The firm intended the supervisors’ discretionary subjective rating to evaluate performance on dimensions that are difficult to measure (e.g., store appearance). We test and find that supervisors give uniformly higher subjective ratings to managers whose objective measure of sales performance is measured with greater noise, and to managers who face higher performance target difficulty, the latter assessed both prior to (*ex ante*) and subsequent to (*ex post*) the evaluation period. These results obtain after controlling for manager ability and performance, and for alternative mechanisms to mitigate differences in compensation risk (e.g., salary changes, sales target changes, and bonus adjustments). The evidence suggests that supervisors use discretion in subjective ratings to provide manager-specific risk premiums for non-executive managers who are subject to an OSFA contract.

1. Introduction

Incentive contracts that tie compensation to performance outcomes as a partial solution to misaligned incentives between agents and a principal have been studied extensively. The preponderance of empirical compensation research focuses on the design of executive contracts (e.g., CEO, CFO), where contracts are tailored to the individual and the job. In theory, these contracts optimally balance the incentive-alignment benefits of incentive pay with the compensation risk premium necessary to induce the executive to accept the contract (Holmstrom 1979). The required risk premium varies with unique aspects of the executive’s job and uncertainty in the operating environment (cf., Prendergast 2002, 2011).

Non-executive (i.e., lower-level) managers are also frequently eligible for incentive compensation. However, it is rare to see individually-tailored contracts for these managers; likely because doing

so would be prohibitively costly. Instead, firms take a “one-size-fits-all” (hereafter “OSFA”) approach, offering a common incentive plan for non-executive managers who have a similar job scope. These OSFA plans typically define common performance measures, incentive weights, and a common pay formula for the group of managers subject to them.¹ The common incentive contract that is most efficient on average for all managers, may however be inefficient as compared with the most efficient contract for an individual manager because heterogeneity in tasks or operating environments produces variation in the compensation risk faced by managers subject to the contract.

Economic theory predicts that compensation risk will be associated with higher salary and lower reliance on incentive pay; however, this neglects the structural reality that salaries are adjusted infrequently and rarely in the negative direction. We posit that using a subjective measure in the OSFA incentive plan provides a more flexible and less costly way to provide parity in risk-adjusted compensation than writing

* Corresponding author.

E-mail addresses: swanderson@ucdavis.edu (S.W. Anderson), h.c.dekker@vu.nl (H.C. Dekker), ksedatole@emory.edu (K.L. Sedatole), e.wiersma@vu.nl (E. Wiersma).

¹ OSFA incentive contracts may also define the same performance targets for all managers and/or specify a common pay scale that indicates a narrow range of salaries for managers in a given job classification.

individual contracts that vary in the mix of salary and bonus. In this study, we investigate whether subjective performance ratings are set so as to provide agent-specific risk premia for non-executive managers. Specifically, we test whether heterogeneity in compensation risk explains cross-sectional differences in the level of subjective ratings awarded by supervisors; specifically, whether higher ratings are given to managers who are exposed to greater compensation risk arising (i) from noise in objective performance measures, and (ii) from variation in the difficulty of targets set for objective performance measures. We control for alternative mechanisms for achieving risk-adjusted compensation parity including: salary changes, performance target changes, and other discretionary bonus pay adjustments (e.g., post hoc adjustments to the payouts determined by the incentive pay formula).

We use archival and survey data from a large retail firm that introduced an OSFA incentive compensation plan with no contemporaneous changes to base salary. The incentive plan formula includes both objective measures of sales performance (i.e., sales per hour, sales-to-goal) and a subjective performance rating that is determined by the district manager (i.e., supervisor). For all store managers, the weights on the objective and subjective measures in the bonus payout formula are the same and cannot be influenced by the supervisor. Consistent with theory about the benefits of subjectivity in compensation (Baker et al. 1994; Baiman and Rajan 1995; Fischer 2000; Murphy and Oyer 2003), the firm intended the quarterly subjective rating to be an assessment of several difficult-to-measure but specified nonfinancial dimensions of performance (e.g., store appearance, employee engagement) that are not adequately captured in a timely manner by the objective performance measures. The new plan was thus viewed as an improvement in the firm's ability to reward effort directed toward desirable tasks. However, the new incentive plan also introduced heterogeneity in compensation risk because, while all store managers are subject to the same OSFA incentive plan, they face heterogeneous operating environments. We posit that heterogeneity in the operating environment is associated with varying compensation risk among the managers that must be met with an increased compensation risk premium, and that subjective ratings provide a mechanism for providing the risk premium and achieving parity in risk-adjusted compensation for managers performing similar jobs.²

Incentive contracts that incorporate subjectivity are, by definition, incomplete in the sense that there is no pre-commitment as to the monetary rewards that will be granted for observed performance (Christ et al. 2012). In contrast to an explicit formulaic bonus plan tied to objective performance measures, the use of subjectivity in the determination of incentive pay can be characterized as an *implicit* contract on which *ex ante* expectations about the *ex post* provision of monetary rewards are based (Aranda et al. 2019). We propose that this implicit contract can be used to achieve parity of risk-adjusted compensation for managers who have a common explicit incentive contract but are exposed to differing levels of compensation risk.

We predict that supervisors' subjective ratings of managers will be positively associated with three sources of compensation risk: (i) noise in objective performance measures that is associated with operational uncertainties in the local store operating environment (as proxied by sales variability), (ii) *ex ante* difficulty of the objective performance measure targets (as proxied by performance measure target relative to prior year actual performance), and (iii) *ex post* difficulty of achieving the objective performance measure targets (as proxied by the average relative performance-to-goal of managers for different supervisors). Consistent with our prediction, we find that, on average, supervisors give higher subjective ratings to managers with noisier objective measures of sales performance. We further hypothesize and find that

supervisors use higher *ex post* subjective ratings to provide risk premiums for managers exposed to greater compensation risk owing to higher *ex ante* objective performance targets. This is consistent with Aranda et al. (2019) who document in a field setting that subjective bonus awards are positively associated with performance target difficulty. Lastly, we hypothesize and find that supervisors use subjective ratings to pay a risk premium for heterogeneous compensation risk imposed by sales targets that are *ex post* revealed to be too difficult (cf. Bol and Smith 2011; Aranda et al. 2019). In all tests of these hypotheses, we control for manager ability and performance, for store characteristics, and for alternative ways to address inequities in compensation risk (i.e., salary level, salary changes and *ex post* discretionary bonus adjustments). In sum, we provide evidence consistent with supervisors using discretion in setting subjective ratings to provide a manager-specific risk premium for non-executive managers who are subject to an OSFA incentive plan.

Interviews with senior managers involved in the design of the compensation plan revealed that it was not the firm's intent for supervisors to use subjective ratings to mitigate store managers' exposure to compensation risk (as documented in prior research), even though economic theory espouses this benefit of subjectivity (e.g., Baker et al. 1994). Nor was it the firm's intent that subjective ratings be used to provide differential risk premium pay (as we show). Rather it was their intent that the subjective ratings play the theorized role of augmenting incomplete objective performance measures. Nonetheless, we hypothesize and find that in a non-executive setting such as ours where OSFA contracts are the norm, supervisors use discretion in subjective ratings to address heterogeneity in store managers' compensation risk that is associated with varied exposure to noise in performance measures and performance target difficulty. Moreover, as this aspect of subjective ratings was not anticipated in the design of the bonus program, we conclude that it was the result of supervisors exercising discretion in an unintended but optimizing fashion. One feature of our setting, namely, that supervisors' own performance evaluations are based on how well their store managers perform, may account for their self-motivated use of discretion to achieve parity in risk-adjusted compensation risk for their subordinates.

This study uses firm archival data to extend prior analytical and empirical executive compensation research to the common real-world setting of OSFA incentive contracts for groups of non-executive managers. While prior research focuses on the design of optimal contracts for a given manager that incorporates all aspects of the contract setting – for example, decision rights allocation, performance measures, task complexity, and environmental uncertainty – this study focuses on non-executive managers for whom OSFA incentive contracts are typical. We provide evidence of the use of discretion in incentive compensation aimed at restoring the optimality of individual-level incentives through differential risk premiums.

Prior research documents the use of subjectivity to mitigate the effects of operational uncertainty in incentive contracting through lower *ex ante* performance targets (Bol et al. 2010) or through subjective adjustments to the actual objective performance measure realizations (Bol and Lill, 2015). We extend this work by providing evidence that subjectivity is also used to pay for that uncertainty – that is, to provide for manager-specific risk premium pay. Our findings are similar to those of Aranda et al. (2019) and Gibbs et al. (2004), both of which show higher *ex post* subjective bonuses as a means to compensate managers for their pre-commitment to *ex ante* objective measure performance targets that are more difficult. Our study confirms the underlying premise of their study, that subjectivity serves as an implicit contract in which managers are rewarded *ex post* for compensation risk imposed on them during the evaluation period. We extend Aranda et al. (2019) by documenting that parity in risk-adjusted compensation is delivered indirectly via subjective ratings of managerial performance rather than directly through subjective bonuses. We also show that subjective ratings are used to provide differential risk premium pay for

² The subjective ratings used in our empirical setting can be viewed as one specific element of the more general construct of *ex post* subjective adjustments that firms may use for providing a compensation risk premium.

compensation risk generated from objective performance measure targets that are *ex post* revealed to be too difficult and more generally from local operational uncertainty. Our study thus helps to provide a more complete picture of the use of subjectivity in providing parity in risk-adjusted compensation.

2. Theory and Hypothesis Development

2.1. The incentive compensation risk premium

Separation of firm ownership and control engenders agency costs of managers making decisions that are not in the best interests of the firm's owners. The large body of agency-theory research shows that properly designed incentive contracts that tie compensation to performance outcomes can mitigate agency costs by aligning the incentives of the manager with those of the firm's owners. A well-documented finding of this research is that because the incentive compensation contract imposes compensation risk on agents, a compensation risk premium must be paid to the agent to induce him to accept the contract (Holmstrom 1979; Banker and Datar 1989; Prendergast 2002). Greater operating or environmental uncertainty increases compensation risk and necessitates a larger risk premium (Holmstrom 1979).

Extensive prior empirical research examines optimal incentive contract design for executives such as CEOs and CFOs (e.g., Lambert and Larcker 1987; Bushman et al. 1996; Core and Guay, 1999). For executives, it is reasonable and cost-effective to individualize incentive contracts, customizing them to the specific contract setting (e.g., the characteristics of the manager, the task complexity, and the operating environment, and the attributes of the performance measures being used). By contrast, most organizations use a common, one-size-fits-all incentive contract for groups of non-executive managers performing similar work (e.g., plant managers, sales managers, or retail store managers). It is not cost-effective to write individualized contracts for non-executive managers and, in fact, doing so may expose the firm to allegations of discrimination toward workers with similar jobs.

Non-executive incentive contracts are typically designed for groups of managers with comparable decision-making authority (e.g., business unit managers). However, while decision rights may be common for these managers, operational environments often vary. As a result, non-executive managers subjected to an OSFA incentive plan are exposed to varying levels of compensation risk which should be associated with commensurate variation in the risk premium paid. While an OSFA incentive contract that fails to address heterogeneous compensation risk through differential risk premiums could still be optimal on average for the group of managers covered by the contract, it may be suboptimal for any individual manager.

There are several mechanisms for paying a differential risk premium to non-executive managers, even in the presence of an OSFA incentive contract. Managers subject to higher compensation risk could be given a higher annual base salary as compared to managers subject to lower compensation risk. A disadvantage of this approach is base salary levels are typically permanent and non-reversible (Campbell and Kamlani 1997; Hall 2005). If risk changes over time or managers are relocated to a business unit with a different risk profile, a more flexible way of compensating for risk differentials may be favored.

Firms can also use contingent pay schemes that incorporate mechanisms to adjust for heterogeneous compensation risk. For example, supervisors can compensate managers facing greater environmental uncertainty through lower performance targets (e.g., Bol and Lill 2015) or through an attenuation of the target ratcheting following high performance (Bol et al. 2010). Lower targets, however, may also be associated with incentive loss because of reduced motivation to perform relative to the business unit's potential (Aranda et al. 2019). Moreover, *ex ante* efforts to ensure commensurate target difficulty across stores is challenging and often results in *ex post* differences in realized target difficulty.

Relative performance evaluation (RPE) schemes that tie contingent pay to the performance of a well-crafted peer group can likewise protect managers' pay from the effects of uncontrollable events and avoids the need for *ex ante* assessments of target difficulty (Albuquerque 2014). Such schemes, however, rely on a high degree of uncertainty being common across peers. Thus, the benefits of RPE are limited when there is heterogeneity in the operating environments faced by different managers.

Alternatively, we argue that in a setting where managers are compensated under an incentive plan with common performance measures and incentive weights, a more flexible and periodically reversible approach is to provide for differential risk premiums using *ex post* discretionary contingent payments – that is, monetary rewards tied to a subjective judgment by the supervisor. Such a flexible approach relies on the “implicit contract” of subjective performance evaluation. Aranda et al. (2019) argue that subjectivity in performance evaluation requires supervisors and subordinates to enter an implicit agreement whereby subordinates take action *with the expectation* that their efforts will be rewarded through the subjective part of incentive pay at the end of the evaluation period. While past interactions in which subordinates periodically observe supervisor behavior help to develop the implicit contract between the subordinate and supervisor, past interactions are not a necessary condition. To the extent that it is in the supervisor's best interest to optimize subordinate motivation, subordinates will form expectations (with or without explicit communication) of how the supervisor will account for the risk they are exposed to in forming subjective performance ratings. It is this expectation that reinforces the implicit contract and influences subordinate effort, and it will be in the best interest of the supervisor to honor the implicit contract. In this way, a contract that is *de facto* incomplete can still be effective in eliciting employee effort and performance. We posit that subjective performance ratings are a mechanism of implicit agreement that can be used to achieve parity in risk-adjusted compensation for managers performing similar jobs.³

2.2. The use of subjectivity in incentive compensation

The use of subjectivity in incentive compensation is ubiquitous in modern firms (Baker et al. 1994; Gibbs et al. 2004; Hayes and Schaefer 2005) and comes in a variety of forms, including subjective performance ratings, subjective weightings of multiple performance measures to create an overall performance assessment, discretionary bonuses, and discretionary adjustments to objectively determined bonuses (e.g., Ittner et al. 2003; Murphy and Oyer 2003; Gibbs et al. 2004; Fisher et al. 2005; Moers 2005; Bol 2008; Bol et al. 2010; Maas et al. 2012; Demeré et al. 2019). Subjectivity is useful in contracting when dimensions of performance are difficult or costly to capture with objective measures or when unforeseeable, non-contractible states of nature affect performance (Baker et al. 1994; Baiman and Rajan 1995). Subjectivity allows supervisors to “correct” for perceived deficiencies in objective performance measures (Gibbs et al. 2004, 2005; Bol and Smith 2011; Woods 2012; Hoppe and Moers 2011) and shield subordinate compensation from uncertainty.

Much of the prior subjectivity literature examines the decision to employ subjectivity and whether it is used more in settings that are exposed to uncertain, uncontrollable events (Bol 2008; Gibbs et al. 2004). However, the empirical evidence is mixed. For example, Hoppe and Moers (2011) find that discretionary bonuses are used for risk-reduction in CEO bonus contracts. Yet, no relation between the use of discretionary bonuses and measures of environmental risk are found in

³ While we focus on *ex post* subjective ratings this is a specific manifestation of the more general construct of *ex post* subjectivity. *Ex post* subjectivity can manifest in other ways including, for example, as *ex post* adjustments to objective measures.

Murphy and Oyer (2003) or Gibbs et al. (2004). In contrast to studies that predict a firm's use of subjectivity, we take the incentive contract as given and focus on whether supervisors use their discretion in subjective ratings to address inequities in compensation risk among managers with similar jobs.

2.3. Hypotheses

Objective performance measures provide a noisy indicator of effort in the presence of operational uncertainty, and when used in a compensation contract, noisy (i.e., less precise) objective performance measures impose compensation risk (Holmstrom 1979; Banker and Datar 1989). It follows that when managers are subjected to an OSFA compensation contract while facing varying levels of uncertainty associated with the operating environment, some managers are exposed to greater compensation risk than others. We predict that supervisors use discretion in setting subjective ratings to address inequities in compensation risk by awarding higher subjective ratings to managers who are exposed to greater operational uncertainty and thus greater noise in objective performance measures:

H1. Subjective ratings in common incentive contracts are positively associated with operational uncertainty.

We also consider the use of subjective ratings to compensate for risk induced by the target setting process. Performance targets provide an opportunity to calibrate incentives prior to the performance evaluation period. Targets can be customized to account for differences in the operational environment. Performance targets set for managers within an organization may vary, between managers and across time, in their degree of difficulty, either intentionally or unintentionally. This poses differential compensation risk in an OSFA incentive compensation plan. To the extent that increased target difficulty is intended or at least recognized, steps can be taken to mitigate the compensation risk heterogeneity.⁴ Aranda et al. (2019) find that for managers who have been assigned more difficult *ex ante* targets (i.e., current performance targets are high relative to prior year actual performance), supervisors use their discretion to *ex post* award a higher subjective bonus. They argue that the implicit agreement of the subjective bonus serves to elicit from managers a commitment to the higher performance targets. *Ex post* subjective ratings serve a similar role in fostering an implicit agreement. Accordingly, we predict supervisors use discretion in setting *ex post* subjective ratings as a means of compensating managers who are given higher *ex ante* performance targets:

H2. Subjective ratings in common incentive contracts are positively associated with *ex ante* performance target difficulty.

Hypotheses H1 and H2 make predictions regarding the use of *ex post* subjective ratings to address *ex ante* differential compensation risk. However, because foresight is not perfect, we propose that there will be a residual compensation risk differential at the end of the evaluation period. Our last hypothesis thus predicts that subjective ratings are used to provide an additional *ex post* mechanism to compensate subordinates for the varied exposure to compensation risk arising from performance targets that are deemed *ex post* to be inappropriate based on the realization of outcomes influenced by unforeseen and uncontrollable events that occurred during the period (Bol and Smith 2011; Merchant and Manzoni 1989; Gibbs et al. 2004). Consistent with this, Gibbs et al. (2004) provide survey evidence that firms are more likely to use subjective bonuses to "recalibrate incentives in situations in which stretch performance targets are not met (p. 429)." If such recalibration does not occur, the incentive effects of future period targets are weakened. Thus,

⁴ We note that intentional *ex ante* target difficulty can vary across time, for instance when in a certain time period, local circumstances favor a higher target for a store.

by compensating for a target that *ex post* is revealed as too difficult, contracting efficiency can be restored.⁵ Accordingly, we predict that:

H3. Subjective ratings in common incentive contracts are positively associated with subordinates' exposure to performance targets that are revealed *ex post* to be difficult.

3. Organizational setting

The research setting is a large U.S. specialty retailer that implemented a new performance-based bonus system that is based on both objective and subjective performance measures (see also Anderson et al., 2010). The firm has 74 company-owned retail stores that are organized in eight geographic districts, each with a district manager (DM) responsible for his stores' performance. Store managers (SM) report directly to a DM, and DMs report to the firm's Director of Operations. While corporate managers retain decision rights regarding many important aspects of the business (e.g., product portfolio and pricing, general advertisement, store locations, hours and layout), store managers have broad decision rights pertaining to store operations, including hiring and managing the workforce, maintaining the store, and initiating local sales promotions.

Prior to the introduction of the new bonus system, SMs received a base salary and were awarded an annual salary increase based on the DM's annual assessment of the overall performance of each of their SMs on a 1-5 scale. This subjective assessment was completely at the discretion of the DM. The firm supplied the DM with no explicit evaluation criteria. Thus, a DM could consider any information that he believed to be relevant to the SM's performance, including objective performance data (e.g., sales, growth, sales per employee hour) or observations from store visits.

There was no formal mapping of performance assessments to pay; however, SMs with good performance received a percentage of their salary as an additional performance payment at the end of each year. After the first year of employment the performance pay ranged from 5%-7% of salary. In subsequent years, SMs would receive a percentage increase of the prior year's performance payout percentage (i.e., prior performance payments were never at risk).⁶ Over time, this system resulted in an increasing share of total compensation being determined by the performance payment, which ratcheted based on prior years' performance payments and as a result, favored more tenured managers. Under this system, the average percentage of salary associated with performance payments was approximately 7%, with a minimum of 1% for a first year SM and a maximum of 23% for a long-tenured SM.

The program became untenable as the labor market for SMs became more competitive. Corporate managers realized that newly hired SMs were unlikely to be motivated by an incentive plan that favored seniority and for which the mapping between performance and annual increments in performance pay was ambiguous. The incentive strength of the system was further weakened by the infrequent observations by DMs of SM activities (often only a couple of times per year) upon which the annual performance assessments were based.⁷ Dissatisfaction with the old system was voiced by two SMs who criticized the casual,

⁵ We note that this prediction is not specific to OSFA compensation contracts, but for such contracts this prediction provides an additional impetus to restore contracting efficiency *ex post*.

⁶ For instance, a SM who received a year 1 performance payment of 5% of base salary, and whose year 2 evaluation stipulated a 10% increase, would receive a year 2 performance payment of 5.5% (5% + (5%*10%)) of base salary; another SM who performed identically in year 2 but who had received a year 1 performance payment of 15% of base salary would receive a year 2 performance payment of 16.5% (15% + (15%*10%)) of base salary.

⁷ Although new store openings provided some opportunities for advancement to DM, new positions were limited and many SMs were unwilling to relocate, so promotion was a weak incentive mechanism.

infrequent basis of past evaluations:

“I know that we should ALWAYS be ready for visits. However two visits per year is not always an accurate indicator of how a store really looks day to day.”

“We are doing everything in our power to improve the store appearance. But when you only get one visit per quarter ... or get a call from your district manager asking ‘What would you rate yourself this month?’... how can that be a true measure?”

After careful deliberation and consultation with DMs and SMs, the firm adopted a new bonus system that was a radical departure from the previous system which rewarded seniority, relied on subjective assessments of SM performance with no common understanding of the aspects of performance that were to be evaluated, and used these assessments in an undisclosed manner to arrive at an SM’s annual performance payment. The new bonus plan provided greater transparency with a formula that translated objective and subjective performance into bonus pay and a checklist that enumerated the items that the DM was to consider in the new subjective rating. Under the new system, the bonus is awarded quarterly and calculated according to a formula that uses four quarterly performance measures: three objective measures (sales performance, sales per hour, and expense control), and one subjective rating of the SM by the DM that is guided by criteria set by the firm. The new bonus system allows SMs to earn a maximum bonus of 20% of base salary each quarter.⁸ The subjective rating is tied to the bonus award via a pre-determined formula; thus subjectivity is limited to the rating itself, not its weight in the formula. SMs’ base salary was not revised structurally with the introduction of the plan, and thus was not explicitly seen as a way to address compensation risk that the new plan would impose upon SMs.⁹

The three objective performance measures account for 75% of the quarterly bonus. Sales performance is weighted 40% in bonus determination, and is measured as sales relative to a sales performance target that is set through participative goal setting between the DM and SM. The bonus payout associated with sales performance varies linearly within an incentive zone (i.e., a threshold of 88.4% and cap of 107% of the goal). The second objective performance measure, sales per hour (SPH), is measured as sales divided by the number of employee hours, and has a bonus weight of 25%. After the quarterly sales target is negotiated, thirteen weekly SPH targets are set. To motivate SMs to optimize revenue and staffing costs, these targets included a lower (-3%) and upper boundary (+20%).¹⁰ Overall SPH performance is computed as the number of times that the store ‘hits’ its weekly SPH target in thirteen weeks. For example, missing one weekly target by performing outside of the boundaries results in a linear bonus decrease (i.e., one-thirteenth times the 25% bonus weight). The third objective measure, expense control, is weighted 10% of the total bonus and requires SMs to reduce controllable expenses as compared to the same quarter in the

⁸ The bonus plan used 10% of base salary as the “baseline” bonus for meeting goals; however, exceptionally good performance (i.e., achieving the ceiling on all four performance dimensions) allowed an SM to earn up to 20% of base salary and unusually poor performance (i.e., achieving the floor performance on all dimensions) would result in no bonus payment. Six very senior SMs for whom a 20% maximum would have constituted a decrease in total compensation as compared with the prior system, had a payout range of zero to 40%. In the first year after adoption, the average bonus for SMs was almost 13% of base salary with a minimum of 5% and a maximum of 22%.

⁹ This observation is supported by tests in which we regress SM base salary and the percent salary change after the plan was introduced on the hypothesized predictor variables and control variables (except those relating to post-plan performance). The measures of sales variability and target difficulty are not significantly associated with either base salary or salary changes.

¹⁰ A number below the lower threshold (-3% of the target) could indicate costly overstaffing, while a number above the higher threshold (+20% of the target) could indicate understaffing and potentially poor customer service.

prior year.

The final component of the SM bonus, which accounts for the remaining 25% of the bonus, is a subjective rating by the DM that is intended to augment the objective performance measures. The new quarterly subjective ratings are different in purpose and content than the annual performance assessments in the old bonus plan. In field interviews with the designers of the compensation plan, the inclusion of the subjective rating in the new bonus plan was explained as a means for evaluating informative but specific and difficult-to-measure dimensions of performance (e.g., store appearance, employee engagement) that are not captured by the easy-to-measure short-term financial performance dimensions. The Director of Operations highlighted the widespread belief that the elements of performance embedded in the subjective rating are more predictive of future sales when she said that she counseled SMs to “focus on your subjective [assessment], that’s what’s going to drive sales.”

DMs were expected to conduct an intensive assessment of each store each quarter. The corporate Operations Department developed a checklist of key performance elements that DMs should consider in determining subjective ratings. For example, DMs are encouraged to consider: voluntary and involuntary staff turnover, store appearance and cleanliness, accuracy of pricing, adherence to store layout directives, stock levels, and so forth. For the first time, DMs were required to visit all stores on a regular schedule to complete evaluations. Although the checklist directed the DM’s attention to specific performance elements, it did not stipulate how the elements are to be weighted or measured, leaving this to the DM’s discretion (cf., Long et al. 2015). Moreover, the DM supplies only the final overall subjective rating to the corporate office without any supporting documentation on the different elements in the checklist.

Importantly, the firm does not instruct DMs to adjust subjective ratings to compensate for heterogeneity in the task environment, in the level of operational uncertainty, or in target difficulty. The discretion that DMs enjoy in setting the subjective rating, however, creates an avenue for addressing heterogeneity of compensation risk among SMs. A comment from one SM indicates that such discretion may have been welcomed:

“I have been involved in other companies’ bonus programs...[that] were based on the same principles yet had flexibility to be more store-specific, knowing that every store does not have the same problems.”

Another SM who was asked to comment on the firm’s intended design, in which the subjective rating was merely an evaluation of difficult to measure performance dimensions, pointed out the shortcoming of failing to accommodate store-level differences:

“I think the plan is not fair and is weighted to high performing stores with no or very few considerations for outside factors, such as location and amount of advertising.”

Thus, the hypothesis that DMs exercise discretion to counter inefficiencies in compensation contracts for individual SMs when the OSFA contract is a poor fit with the work environment, is a hypothesis of supervisors *instinctively* (perhaps with the added motivation of their own incentive compensation plan) addressing contract failures using discretion in a manner that was not prescribed (though not *per se* undesired) by the firm.

4. Research Methods

4.1. Data and sample

The empirical analysis uses the firm’s archival data on the performance measures in the incentive pay formula as well as manager and store characteristics. We also incorporate survey data collected from the SMs and DMs to perform measurement construct validation and a

robustness test. We collected ten quarters of archival data, five quarters from the pre-plan period (used to measure managerial ability, sales variability and *ex ante* target difficulty, as described below), and five quarters from the post-plan period (used to test the hypotheses). Eliminating observations with missing data leaves 242 observations with complete archival data for testing the hypotheses. Because the *ex ante* target difficulty measure generates a loss of 20 observations, we estimate our models with and without this variable. We take the larger sample ($N = 262$) as a basis for reporting descriptive statistics, noting that the smaller sample provides similar insights.

4.2. Variable measures

4.2.1. Dependent and independent variables

The dependent variable is the SM's quarterly subjective rating (*Rating*), which varies for each SM each quarter (q) and is scored on a 1 to 5 scale. The independent variables are organized according to their corresponding hypothesis below.

Hypothesis H1 predicts that the subjective rating is positively associated with operational uncertainty faced by an SM. In our survey data, one store manager commented, "the SPH targets are difficult to accomplish in stores that have highly fluctuating sales on a weekly basis." Following this logic, we use variability in store sales, *Sales Variability*, as our proxy for operational uncertainty and calculate this as the 3-month standard deviation of monthly sales within each quarter divided by that store's average sales for that quarter. To avoid contemporaneous influences on this measure, we use the lagged measure of sales variability for the same quarter of the prior year.¹¹ The correlation between the lagged and contemporaneous measure of sales variability (0.94, $p < 0.01$) is very strong and indicates persistent sales performance measure variation that supervisors can incorporate into their subjective ratings. Using a lagged measure, however, results in 19 store-quarter observations for new stores without prior year sales data in the respective quarter. To avoid a loss of observations on the new store indicator, we impute these 19 missing values with contemporaneous sales variability, noting that results are very similar without imputation.

Hypothesis 2 predicts that *ex ante* objective performance measure target difficulty is associated with higher subjective ratings. Our measure of *ex ante* target difficulty, which varies between stores and quarters, is measured as the sales target of the quarter divided by the sales level of the same quarter in the prior year ($Ex\ ante\ TD = Sales\ Target_t / Sales_{t-4}$). We use the sales level from the same quarter prior year because of large seasonal variation. This results in a loss of 14 observations. In addition, we exclude six outlier observations from new stores where sales targets (more than) doubled compared to prior year sales. This results in 242 observations available to test H2.

Hypothesis H3 predicts that *ex post* realized objective performance measure target difficulty is associated with higher subjective ratings. We compute a time-varying measure of *ex post* target difficulty at the DM level, *Ex post TD*, as the quarterly average STG performance across all SMs of a DM.¹² We base this variable measurement on the assumption that a DM will infer greater *ex post* target difficulty if he observes that the average STG of his SMs is low relative to the average STG of peer DMs. The measure is multiplied by minus one so that higher values indicate more difficult targets.

¹¹ We compute sales variation within the quarter as this aligns with the quarterly rating, and seasonality results in distinct sales patterns between quarters that make variation across longer time periods less informative.

¹² Computing an *ex post* measure of target difficulty at the sales manager level is infeasible because observed performance is confounded with the inputs provided (effort and ability) to realize that performance.

4.2.2. Control variables

We control for several variables that can affect subjective ratings and which may be correlated with the independent variables of interest, including: task complexity, whether or not a store is new, the store's economic environment, the level of objective performance measures, unusually high (windfall) sales, the SM's managerial ability, the preplan bonus payments, the incidence of SM change, and the DM's and SM's experience with the new bonus plan. We also control for alternative options for supervisors or firm management to provide a risk premium through *ex post* discretionary adjustments and salary changes.

Task complexity: As a proxy for task complexity we control for store size, *Store Size*, measured as square footage of the store, scaled by 1,000. Larger stores in our field setting are more complex to manage than smaller stores because of larger and more diverse operations involving more departments, a greater product variety, a more heterogeneous customer base, and a greater need to coordinate and supervise more subordinates. While assignment of store managers to stores was not random (e.g., more capable SMs being placed in larger stores), the new bonus plan was not accompanied by reassignment of SMs.

New store: Newly opened stores lack historical information about unique operational characteristics (e.g., characteristics of the labor force), market opportunities, and purchase habits of local customers. Even experienced SMs assigned to a new store need to learn how to operate the store, hire and supervise new employees, and learn about the local market. One store manager put it this way, "Being a new store ... I don't think I should be held responsible for hitting SPH on the mark every week, because the projections are being set off another store's projections." In a similar vein, another store manager argued, "New stores suffer when plans are set too high and not adjusted accordingly." *New Store* is an indicator variable set to 1 if in a quarter the store is open for less than a year, and zero otherwise.

Store economic environment: We use two market demographic measures to capture variation in the stores' economic environment and performance potential. Specifically, we use population density (i.e., square root of number of households, scaled by 100; *Number HH*) and median household income (*HH Income*) in the store's immediate area. Both control variables are supplied by the firm but purchased from a third party data source at a single point in time; thus they vary by store but not over time.

Objective performance level: The DM observes the objective performance measure outcomes at the time that he formulates the subjective rating. If subordinate ability affects all aspects of performance, the rating may be associated positively with the objective performance measures (Murphy et al. 1993). A positive correlation may also obtain if the DM relies on these data inappropriately to set the subjective rating (i.e., a "spillover effect" as in Bol and Smith 2011). We control for performance on the objective performance dimensions using two of the objective performance measures that jointly determine 65% of the SM's bonus: Sales to goal (*STG*), and sales per hour (*SPH*).¹³ *STG* is measured as the percentage of sales above or below the target (i.e., (actual sales-goal)/goal). *SPH* is measured as the ratio of weeks in a quarter that SPH is between the lower and upper limit (i.e., number of weeks out of 13).¹⁴ The theoretical range is 0-1. Sales goals are established quarterly

¹³ We exclude expense control, the third objective performance element of the bonus plan, because it was discontinued during our period of study (i.e., it exists for only the first three of the five quarters) and interviews suggest that problems with this measure had surfaced even earlier. Specifically, there was a sense that stores that had been profligate prior to the implementation of the bonus plan were rewarded because they found it much easier to reduce expenses. Including the expense control measure as additional control variable to our analyses results in similar results and inferences regarding our hypothesis tests, despite the significant reduction in sample size. The expense control measure is not significantly associated with the subjective rating.

¹⁴ For a few observations the number of weeks differs slightly because data were not available for all 13 weeks, for example because the store was opened during the quarter.

and even unforeseen events do not prompt adjustment of the goal within the period.

Windfall sales: An important feature of subjective ratings is that they are determined *ex post*, making them useful as a way to incorporate post-contractual information. The effect on bonus compensation of unusually high (low) sales that may be due to good (bad) luck can be filtered out with a lower (higher) subjective rating. We control for this noise filtering role of subjectivity by including a variable identifying exceptionally high “windfall” sales. Recall that the bonus system places a cap on managers’ bonus realization of 107% of the sales target. Exactly meeting the target provides store managers with a 10% bonus payout, and exceeding the cap by as much as 7% can increase the bonus percentage to a maximum of 20%. Accordingly, we include as a control an indicator variable, *Windfall Sales*, that takes the value 1 when STG is above 7%, and 0 otherwise. This holds for 5.4% of the observations.¹⁵

Managerial ability: Managers of higher ability can be expected to perform better on both objective and subjective performance dimensions. Moreover, managers of differing ability are assigned in a non-random manner to stores with characteristics that are correlated with the independent variables (e.g., the assignment of more capable managers to large stores). Recent studies have operationalized managerial ability as the fixed effects in a revenue performance model (e.g., Demerjian et al. 2012). Since SMs’ primary responsibility is generating store sales, we similarly estimate SM fixed effects to proxy for SM ability. As described in Appendix A and reported in Table A1, we estimate a sales performance regression model using data from the five quarters before the new bonus plan was in place (N = 224). In the model, we control for *Industry Growth* and four store characteristics that affect its sales potential: *Store Size*, *New Store*, population density (*Number HH*) and median household income (*HH Income*) in the store’s immediate area. *Industry Growth* captures the contemporaneous sales trend in the firm’s specific retail segment based on indices provided by the International Council of Shopping Centers (ICSC 2005). The SM fixed effect captures cross-sectional variation in managerial ability prior to the implementation of the plan. Table 1 indicates a large increase in explanatory power ($F_{63,155} = 13.14$, $p < 0.01$) of including SM fixed effects in the base model with controls only, indicating that manager influence on sales performance in the pre-plan period is significant. We use the estimated coefficient, α_{sm} , as the measure of *Managerial Ability* to discriminate common antecedents that cause correlation between the independent variables and the subjective ratings.

Preplan bonus: Because the new incentive system puts all bonus pay at risk, SMs face the possibility that their total pay will decline under the new system. We control for any tendency of DMs to adjust subjective ratings to ensure SM total pay does not decline under the new system (Woods 2012) by controlling for preplan bonus levels. Preplan bonus level is also associated with SM seniority (or tenure) in the firm, and thus additionally controls for subjective rating adjustments to ease the transition to the new system for more senior SMs.

SM change: Across all quarters, several SMs leave or join the firm, or switch to other stores. Disruption of store management likely will affect performance on a number of dimensions. We have no directional prediction for how this would influence the DM’s subjective rating, and different effects are likely depending on the circumstances that prompted the change (i.e., voluntary versus involuntary separation, promotion versus lateral move). We control for SM change with an indicator for whether the store has a new SM in a particular quarter (*SM*

¹⁵ We also investigated unusually low sales that may reflect unanticipated “headwinds.” In untabulated robustness tests, we include in our model an indicator variable identifying observations with STG below -7%. This variable is not significantly related to subjective ratings in any of our models and reported results are substantially unchanged with its inclusion. Since this is a comparatively rare event in the period that we studied (only 10 observations), we chose not to include this analysis in our reported results.

Table 1
Descriptive statistics.

	Hypothesis	Mean	Med	Std. Dev	Min	Max
<i>Rating</i>		3.50	3.60	0.55	0.50	4.75
<i>Sales Variability</i>	H1	0.26	0.20	0.15	0.01	0.64
<i>Ex ante TD</i>	H2	1.02	1.02	0.10	0.69	1.66
<i>Ex post TD</i>	H3	-0.19	0.34	2.82	-11.02	3.89
<i>Store Size (1000s)</i>		51.37	49.52	6.78	35.67	67.16
<i>New Store</i>		0.19	0.00	0.40	0.00	1.00
<i>Number HH (100 s)</i>		3.94	3.84	0.77	2.00	6.34
<i>HH Income</i>		42,891	41,074	9,901	26,054	69,145
<i>STG</i>		0.05	-0.57	4.75	-9.63	23.20
<i>SPH</i>		0.63	0.62	0.18	0.08	1.00
<i>Windfall Sales</i>		0.05	0.00	0.23	0.00	1.00
<i>Managerial Ability</i>		-0.09	-0.24	1.13	-1.91	3.18
<i>Preplan Bonus</i>		1.21	0.98	0.89	0.30	4.20
<i>SM Change</i>		0.03	0.00	0.18	0.00	1.00
<i>Plan Experience</i>		0.38	0.00	0.49	0.00	1.00
<i>% Δ Salary</i>		0.04	0.04	0.02	0.00	0.07
<i>Formula Override</i>		0.02	0.00	0.14	0.00	1.00

Notes: This table provides descriptive statistics for all variables. For one observation the subjective rating was lower (0.5) than the theoretical range (1-5) because of a temporary exception in the quarter. Deleting this observation from the analysis has no appreciable impact on the results. For confidentiality reasons, we disguise the bonus information in this table by multiplying the *Preplan Bonus* variable by a scalar. Variable definitions are provided in Appendix B. N = 262 for all variables except for *Ex ante TD* (N = 242).

Change), coded as one, and 0 otherwise.

Plan experience: With the introduction of the plan, DMs are subject to learning effects in the rating task. However, the DM’s increased rating experience may be confounded with SMs’ development of skills and performance on the checklist activities that the firm intends as the basis for the DM’s subjective rating. On balance we expect subjective ratings to improve with time as both DMs and SMs learn. *Plan Experience* is measured as an indicator variable that is set to one in the last two of the five quarters, and zero otherwise, to differentiate between early and later stages of the new incentive plan.¹⁶

Alternative mechanisms to mitigate compensation risk heterogeneity: We control for two alternative mechanisms that firms may use to achieve parity in risk-adjusted compensation among store managers. First, we control for the percentage change in salary from before to after the introduction of the bonus plan, *% Δ Salary*. While salary change likely reflects an SM’s performance achievements, the measure also provides a control for any additional, and potentially differential, risk premium that is paid via a salary differential.¹⁷ Second, in a few instances, discretionary overrides of the bonus formula were made by the corporate office in order to adjust the SM bonus (internally referred to as ‘formula override due to temporary exception’). Adjustments typically took place in unusual situations such as when an SM changed stores during a quarter. While infrequent, discretionary adjustments by the corporate office may substitute for DMs adjusting subjective ratings to compensate for risk, so we control for the occurrence of adjustments with an indicator variable, *Formula Override*. By including these control variables, we test for a differential risk premium provided by the subjective rating over and above the risk premium supplied by these alternative mechanisms.

¹⁶ Shifting the cutoff to the final three quarters or using a linear time measure (1–5) does not change the results.

¹⁷ Note that the new incentive plan was not accompanied with a revision of SMs’ base salaries, and DMs do not have the authority to decide on salary changes. Also, while the *change* in salary best reflects the compensation of new risk imposed by an OSFA incentive plan, adding the post-plan salary level as a control variable provides similar results as reported. Salary level is positively correlated with subjective ratings and with *Managerial Ability*, indicating that more capable managers receive higher salaries.

Table 2
Variable correlations.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
1. Rating	1.00															
2. Sales Variability	0.09	1.00														
3. Ex ante TD	0.17	0.06	1.00													
4. Ex post TD	-0.00	0.02	-0.05	1.00												
5. Store Size	0.35	-0.01	0.21	-0.07	1.00											
6. New Store	0.19	-0.01	0.22	0.01	0.38	1.00										
7. Number HH	-0.33	-0.02	-0.24	-0.05	-0.23	-0.11	1.00									
8. HH Income	-0.15	0.05	0.03	0.00	-0.03	-0.01	0.09	1.00								
9. STG	0.11	-0.06	0.10	-0.53	0.15	-0.04	-0.09	-0.02	1.00							
10. SPH	0.23	-0.18	0.14	-0.06	-0.05	-0.02	-0.16	0.04	0.23	1.00						
11. Windfall Sales	0.01	-0.03	0.00	-0.41	0.15	0.01	-0.06	-0.01	0.60	-0.01	1.00					
12. Managerial Ability	0.18	-0.05	-0.11	0.06	0.02	-0.03	-0.24	-0.29	-0.01	0.13	-0.05	1.00				
13. Preplan Bonus	0.10	-0.02	-0.06	0.04	0.05	-0.07	-0.28	-0.08	0.01	0.10	0.03	0.59	1.00			
14. SM Change	-0.02	-0.13	0.15	0.02	-0.01	0.01	0.09	-0.02	-0.09	0.06	0.05	-0.08	-0.08	1.00		
15. Plan Experience	0.10	-0.32	-0.00	0.07	0.03	0.05	0.04	-0.03	-0.08	-0.01	-0.15	-0.00	0.02	0.02	1.00	
16. % Δ Salary	0.34	-0.06	0.16	0.07	-0.07	0.11	-0.10	-0.21	0.03	0.22	0.01	0.09	-0.06	0.01	0.03	1.00
17. Formula Override	-0.12	-0.10	0.26	0.04	0.13	0.14	0.04	-0.09	-0.12	-0.10	0.09	-0.06	-0.09	0.28	-0.11	0.03

Correlations equal to or greater than |0.11|, |0.13|, and |0.16| are significant at 0.10, 0.05 and 0.01, respectively. Variable definitions are reported in Appendix B. N = 262 for all variables except for Ex ante TD (N = 242).

Appendix B summarizes all variable measures.

4.3. Descriptive statistics

Table 1 provides descriptive statistics for the variables. The mean subjective rating is 3.50, values span the full scale (from 0.5-4.75), and approximate a normal distribution. Mean Ex ante TD is 1.02 which indicates that sales targets increase on average by 2% compared to sales from the same quarter prior year. However, these numbers vary tremendously with a minimum value of 0.69 and maximum of 1.66. Mean STG is approximately zero, indicating that, on average, sales targets are met. Variation around the mean is substantial (i.e., up to 10% below and 23% above the target). STG exceeds the cap in the bonus system (Windfall Sales) for 5% of the observations. Mean SPH is 0.63, indicating that on average SPH targets are met in 8.2 of 13 weeks in a quarter, with an observed minimum of one and a maximum of thirteen weeks. In 19% of the observations, stores were open for less than 12 months (New Store = 1), consistent with the strong growth of the company. In addition, in 3% of the observations the SM moved to another store (SM Change = 1).

Table 2 reports correlations. In relation to our hypotheses, subjective ratings are positively correlated with (H1) Sales Variability (ρ = 0.09, p < 0.10, one-tailed) and (H3) Ex ante TD (ρ = 0.17, p < 0.01, one-tailed), but not with Ex post TD. Correlations among the independent and control variables provide no concerns about multicollinearity.

5. Results

5.1. Estimation method

The subjective rating varies at the quarterly level (q), store manager (or store) level (SM), and the district manager level (DM). Specifically, the 262 observations used represent five quarters nested within 58 SMs which are in turn nested within 8 DMs. We use multilevel modeling to identify sources of variation in subjective ratings across levels. The ability to estimate how variables at different levels explain variation in subjective ratings is an important advantage of multilevel modeling over a fixed-effects approach.

Since we aim to assess the influence on subjective ratings of variables that vary across quarters, stores, and store managers, we estimate a two-level model that includes DM fixed effects to control for

systematic rating differences among DMs.¹⁸ The estimated multilevel model is (cf. Bryk and Raudenbush 1992):

$$\text{Level 1: Rating}_{q,SM,DM} = \alpha_{0,SM,DM} + \alpha_{1,SM,DM} (\text{Sales Variability}_{q,SM,DM}) + \alpha_{2,SM,DM} (\text{Ex ante Target Difficulty}_{q,SM,DM}) + \alpha_{3,SM,DM} (\text{Ex Post Target Difficulty}_{q,SM,DM}) + \alpha_{4,SM,DM} (\text{New Store}_{q,SM,DM}) + \alpha_{5,SM,DM} (\text{STG}_{q,SM,DM}) + \alpha_{6,SM,DM} (\text{Windfall Sales}_{q,SM,DM}) + \alpha_{7,SM,DM} (\text{SPH}_{q,SM,DM}) + \alpha_{8,SM,DM} (\text{Formula Override}_{q,SM,DM}) + \alpha_{9,SM,DM} (\text{SM Change}_{q,SM,DM}) + \alpha_{10,SM,DM} (\text{Plan Experience}_{q,SM,DM}) + e_{q,SM,DM} \quad (1)$$

$$\text{Level 2: } \alpha_{0,SM} = \beta_0 + \beta_{1,DM} (\text{Store Size}_{SM}) + \beta_{2,DM} (\text{Number HH}_{SM}) + \beta_{3,DM} (\text{HH Income}_{SM}) + \beta_{4,DM} (\text{Managerial Ability}_{SM}) + \beta_{5,DM} (\% \Delta \text{ Salary}_{SM}) + \beta_{6,DM} (\text{Preplan Bonus}_{SM}) + \mu_{0,SM}$$

5.2. Model Estimation Results

Table 3 reports the estimation results.¹⁹ The benchmark model in Column (1) includes SM random effects, DM fixed effects, and the control variables.²⁰ The coefficient estimates show no association between Store Size and subjective scores in the new bonus plan. In contrast, managers who are responsible for a recently opened store (New Store) receive a higher subjective rating (coefficient of 0.23, p < 0.01), suggesting that SMs are compensated for the uncertainties that accompany operating a new store. SMs receive a higher rating when the economic environment reduces sales opportunities as indicated by a lower number of households (Number HH) in the store area (-0.11,

¹⁸ An alternative would be to use a fixed-effects specification (i.e., using indicator variables for each SM and DM) which, while relying on a weaker assumption about the correlation of SM-level effects and the error term, disregards the between-store variation related to manager and store characteristics (Kreft and De Leeuw 1998). Estimating OLS regressions with SM and DM fixed effects, and clustering of standard errors by SM, provides similar results and inferences as reported. Modeling DM random effects (i.e., level 3 variation) in the multilevel specification instead of DM fixed effects also provides similar results and inferences as reported.

¹⁹ Two SMs in the dataset each had only one quarterly observation, and one had only two quarterly observations. Deleting these observations from the analysis provides similar results as reported.

²⁰ A random effects only model (untabulated) shows most variation occurs at the SM level (level 2), with an intraclass correlation of 61% (39% at Level 1). The control variables significantly explain Level 2 variation (random effect decreasing from 0.20 to 0.04), and relatively less Level 1 variation (random effect decreasing from 0.13 to 0.11).

Table 3
Hypothesis tests.

Variables	Expectation	(1) N = 262	(2) N = 242	(3) N = 262
<i>Intercept</i>		2.64*** (5.89)	1.72*** (2.90)	2.37*** (5.30)
<i>Sales Variability</i>	H1: +		0.63*** (4.05)	0.66*** (4.30)
<i>Ex ante TD</i>	H2: +		0.58** (1.99)	
<i>Ex post TD</i>	H3: +		0.02** (2.07)	0.02** (1.66)
Control variables				
<i>Store Size</i>		0.01 (1.30)	0.01 (0.99)	0.01 (1.16)
<i>New Store</i>		0.23*** (2.65)	0.21*** (2.00)	0.24*** (2.87)
<i>Number HH</i>		-0.11** (-2.33)	-0.07 (-1.37)	-0.10** (-2.17)
<i>HH income</i>		-0.00 (-0.82)	-0.00 (-0.77)	-0.00 (-0.88)
<i>STG</i>		0.00 (0.77)	0.01 (1.41)	0.01 (1.41)
<i>SPH</i>		0.41*** (2.67)	0.54*** (3.45)	0.53*** (3.59)
<i>Windfall Sales</i>		-0.16 (-1.18)	-0.09 (-0.61)	-0.11 (-0.82)
<i>Managerial Ability</i>		-0.01 (-0.24)	0.01 (0.10)	-0.01 (-0.19)
<i>Preplan Bonus</i>		0.15** (2.55)	0.14** (2.20)	0.15** (2.60)
<i>SM Change</i>		0.17 (1.28)	0.17 (1.31)	0.22* (1.73)
<i>Plan Experience</i>		0.08* (1.84)	0.15*** (3.12)	0.15*** (3.30)
<i>% Δ Salary</i>		11.25*** (5.03)	10.10*** (4.12)	11.05*** (4.99)
<i>Formula Override</i>		-0.46** (-2.36)	-0.39 (-1.65)	-0.36* (-1.89)
<i>Fixed DM effects</i>	Yes	Yes	Yes	
Level 1: Residual		0.11*** (9.84)	0.10*** (9.40)	0.10*** (9.87)
(Wald)				
Level 2: SM effect		0.04*** (2.76)	0.05*** (3.00)	0.04*** (2.94)
(Wald)				
R ²		0.54	0.55	0.57
AIC / -2LL		271/225	248/196	255/205

Notes: This table presents the results (coefficient estimates with t-statistics in parentheses) of a multi-level model with *Rating* as the dependent variable (see equation (1)).

*, **, *** indicates significance at 0.10, 0.05, 0.01 (one-tailed for hypothesized effects; two-tailed otherwise). R² is computed as the percentage $\Delta(\epsilon + \mu + \nu)$ compared to the random effects only model. Variable definitions are provided in Appendix B.

$p < 0.05$), but there is no significant impact on ratings of *HH Income*.

The coefficient estimates show a significant positive association between *SPH* performance and subjective ratings (0.41, $p < 0.01$), but not of *STG* performance. This result may indicate that in establishing subjective ratings, DMs anchor on objective performance measures, in particular those that are considered more controllable by the SM (i.e., *SPH* with weekly targets in which SMs can influence staffing levels instead of quarterly *STG* that is subject to exogenous sales fluctuations). The coefficient on *Windfall Sales*, while negative as expected, is not significant.²¹ Although *Managerial Ability* is positively correlated with subjective ratings (see Table 2), its coefficient in the multivariate model is insignificant.

Consistent with DMs using subjective ratings to mitigate the impact of the new system, ratings are higher for SMs with larger preplan

²¹ As a sensitivity test, we re-estimate our model setting *Windfall Sales* equal to 1 when *STG* exceeds 5% of the cap (13% of the observations). While the coefficient on *Windfall Sales* remains negative and insignificant, the positive effect of *STG* becomes significant ($p < 0.10$).

bonuses (0.15, $p < 0.05$). There is no significant difference in subjective ratings for SMs changing to other stores (*SM Change*). However, a positive coefficient of *Plan Experience* (0.08, $p < 0.10$) indicates that ratings increase over time, likely due to learning by SMs and the development of strategies to enhance subjective performance.²²

Our results further show that discretionary bonus formula overrides (*Formula Override*) are associated with lower subjective ratings (-0.46, $p < 0.05$). This is consistent with a substitution between these alternative mechanisms for providing differential compensation risk premiums. In contrast, the percentage change in salary is positively associated with subjective ratings (11.25, $p < 0.01$). A salary change may not only capture an alternative way to compensate for risk, but also reflect overall SM performance that has an upward influence on subjective ratings.

Test of hypotheses: Column (2) reports the estimation of our full model (i.e., equation (1)) used to test our hypotheses. The AIC and -2LL indicate significantly improved model fit and the independent variables explain a larger part of variation in the subjective ratings ($R^2 = 0.55$).

Consistent with H1, *Sales Variability* has a positive and significant coefficient (coefficient of 0.63, $p < 0.01$, one-tailed). Importantly, this effect is incremental to the effects on subjective ratings of store and local market characteristics. This provides evidence of supervisors using subjective ratings to provide differential risk premiums to managers facing greater operational uncertainty.

Hypothesis H2 predicts a positive relation between *ex ante* target difficulty and subjective ratings. We find, consistent with this prediction, a positive and significant coefficient on *Ex ante TD* (coefficient of 0.58, $p < 0.01$, one-tailed). Consistent with H3 that predicts a positive association between subjective ratings and *ex post* revealed target difficulty, the coefficient on *Ex post TD* is also positive and significant (coefficient of 0.02, $p < 0.05$, one-tailed). Thus, we find evidence that both *ex ante* target difficulty and *ex post* revealed target difficulty are uniformly positively associated with the DM's subjective rating.²³

Because the construction of *Ex ante TD* results in the loss of sample size, in Column (3) we estimate our model without this variable as a robustness test. We again find evidence consistent with hypotheses H1 and H3 that subjective ratings are uniformly higher for SMs of stores with higher sales variability and higher *ex post* revealed objective performance targets.

In sum, we find evidence in support of our hypotheses that operational uncertainty, as reflected in increased variability of sales, and target difficulty (measured both *ex ante* and *ex post*) are associated with uniformly higher subjective ratings. This is consistent with our theory that operational uncertainty and differences in target difficulty impose differential compensation risk across managers subjected to an OSFA compensation contract. Variation in compensation risk, absent intervention, results in inequity in risk-adjusted compensation. We interpret our findings to suggest that subjective ratings can be used to provide parity in risk-adjusted compensation. Importantly, these results obtain after controlling for managerial ability and performance, store characteristics, prior bonus levels, SM change and plan experience, and for the potential substitution of alternative mechanisms to provide parity in

²² To obtain qualitative evidence on how SMs learn to improve subjective performance, the survey asked them to formulate their strategies to improve the rating. Strategies mentioned more than once are (1) holding assistant managers accountable to a greater extent, (2) going over checklists every day, and (3) setting tighter targets for assistant managers to improve.

²³ In a robustness test, we also add as a control variable the SM's tenure as a store director (i.e., the number of years the SM has been in that position). This not only reflects SM job experience, but, because DM tenure is typically longer than the tenure of any subordinate SM, also provides a close approximation of the length of the relationship with the DM (cf., Bol 2011). SM tenure was gathered from the SM surveys and is available for 52 SMs, providing a sample size of 234 quarterly observations. Results are comparable to those of the main model and SM tenure is not significantly related to the subjective rating.

risk-adjusted compensation.

6. Conclusion

Theory suggests that *ex post* subjectivity in determining incentive compensation provides a useful mechanism, not only to reward hard-to-measure dimensions of performance, but also to incorporate post-contractual information into incentive compensation. That is, the effect on incentive compensation of unusually high (low) performance that may be due to good (bad) luck can be filtered out with a lower (higher) subjective performance assessments. This study considers an additional use of *ex post* subjectivity in providing for parity in risk-adjusted compensation under ‘one-size-fits-all’ (OSFA) incentive contracts commonly used for non-executive managers.

We theorize that workplace heterogeneity is associated with variation in objective performance measure noise arising from operational uncertainty and with differences in the difficulty of performance targets across business units. This workplace heterogeneity, coupled with a ‘one-size-fits-all’ incentive contract results in differential compensation risk across managers. Importantly, this effect arises because the firm economizes on contracting by using an OSFA compensation contract for managers who have similar but not identical work environments. Absent any mechanism to address this compensation risk heterogeneity, a contract that may be optimal at the firm level, can be suboptimal when considered from the individual manager level.

We hypothesize that supervisors use discretion in setting subjective ratings to address heterogeneity in compensation risk by delivering an agent-specific risk premium for managers who are subject to the OSFA contract. Specifically, we predict and find that supervisors give uniformly higher subjective ratings to managers who are exposed to greater compensation risk arising (i) from noise in objective performance measures, and (ii) from variation in the difficulty of targets set for objective performance measures (as measured both *ex ante* and *ex post* of the performance evaluation period). Importantly, these results obtain after controlling for alternative mechanisms to provide parity in risk-adjusted compensation.

We interpret our results as evidence that supervisors use discretion

Appendix A

Recent studies operationalize managerial ability as the fixed effects in a revenue performance model (e.g., Demerjian et al. 2012). Since SMs’ primary responsibility is generating store sales, we estimate SM fixed effects to proxy for SM ability. Specifically, we estimate a sales performance regression model (equation 2 below) using data from the five quarters *before* the new bonus plan was in place (N = 224). We control for *Industry Growth* and four store characteristics that affect sales potential: *Store Size*, *New Store*, *Number HH* and *HH Income*. *Industry Growth* captures the contemporaneous sales trend in the firm’s specific retail segment (International Council of Shopping Centers ICSC, 2005).

$$Sales_{SM,q} = \alpha_{sm} + \alpha_1 Store\ Size_{SM} + \alpha_2 New\ Store_{SM,q} + \alpha_3 Number\ HH_{SM} + \alpha_4 HH\ Income_{SM} + \alpha_5 Industry\ Growth_q + e_{SM,q} \quad (1)$$

Table A1 provides the estimation results. Column (1) includes only the five control variables ($R^2 = 25\%$), and Column (2) adds the SM fixed effects, α_{sm} ($R^2 = 88\%$). The large increase in explanatory power ($F_{63,155} = 13.14, p < 0.01$), indicates that manager influence on sales performance in the pre-plan period is significant. The estimated coefficient, α_{sm} , is the measure that we use for *Managerial Ability*.

We conduct two validity tests of our *Managerial Ability* variable. First, we compare the measure with survey data from the DMs that we collected at the beginning of the new bonus system, before the first subjective ratings were assigned. In the survey, DMs indicated, for each of their SMs, the extent to which they agreed (1 = strongly disagree, 7 = strongly agree) that the SM: (1) knows the technical parts of his/her job extremely well, (2) is competent at his/her job, (3) is very good at his/her job, and (4) is one of the best store managers at the firm. Five DMs responded to these questions for 45 SMs. Factor analysis of the items reveals one factor (Cronbach $\alpha = 0.85$). The correlation of *Managerial Ability* with the DM’s assessment of the SM ability is strong and highly significant ($\rho = 0.52, p < 0.01$). Second, correlation analyses show that *Managerial Ability* is significantly associated with the pre- and post-plan salary levels ($\rho = 0.45$ and $\rho = 0.48$, both $p < 0.01$), dollar salary increase ($\rho = 0.16; p < 0.05$), SM promotion in the post-plan period (e.g., to a larger store) ($\rho = 0.18, p < 0.01$), and post-plan SPH performance ($\rho = 0.13; p < 0.05$).²⁴ These validation tests also indicate that the measure captures performance variation related to the SM instead of store characteristics.

²⁴ *Managerial Ability* is not correlated with post-plan STG performance ($\rho = -0.01; p > 0.10$). However, weekly SPH is the more controllable of the two objective performance measures and better performing SMs may also receive higher sales goals, resulting in relatively similar STG performance for SMs of differing ability.

Table A1
Estimation of *Managerial Ability*.

Variables	(1) N = 224	(2) N = 224
<i>Intercept</i>	-1.33 (-1.85)*	-5.53 (-0.18)
<i>Industry Growth</i>	0.00*** (8.86)	0.00*** (8.62)
<i>Store Size</i>	0.04 (2.37)**	0.09 (9.73)***
<i>New Store</i>	-1.43 (-7.21)***	-1.87 (-6.98)***
<i>Number HH</i>	0.20 (2.58)**	0.31 (0.64)
<i>HH Income</i>	-0.00 (-7.68)***	-0.00 (-0.09)
<i>Manager fixed effects</i>	No	Yes
<i>R²</i>	0.25	0.88

Notes: This table presents the results (coefficient estimates with t-statistics in parentheses) of an OLS regression with quarterly sales, *Sales*, as the dependent variable:

$$Sales_{SM,q} = \alpha_{sm} + \alpha_1 Store Size_{SM} + \alpha_2 New Store_{SM,q} + \alpha_3 Number HH_{SM} + \alpha_4 HH Income_{SM} + \alpha_5 Industry Growth_q + e_{SM,q} \quad (1).$$

*, **, *** indicates significance at 0.10, 0.05, 0.01 (two-tailed). Standard errors are based on White adjusted standard errors. A difference test between model 1 and 2 indicates that managerial influence on sales performance is significant ($F_{63,155} = 13.14, p < 0.01$). Manager fixed effects from this model provide our measure of *Managerial Ability*. Variable definitions are provided in Appendix B.

Appendix B. Variable measures

Measure	Description
<i>Rating</i>	The DM's quarterly subjective rating for an SM on a scale of 1-5.
<i>Sales Variability</i>	A time-varying measure of operational uncertainty, measured as the 3-month standard deviation of monthly sales within each quarter for a store divided by the average sales in that quarter for that store. <i>Sales Variability</i> is calculated for the same quarter prior year (q-4). For missing values the number is imputed.
<i>Ex ante TD</i>	A time-varying measures of <i>ex ante</i> target difficulty, measured as the SM's sales target for the quarter divided by that SM's actual sales level for the same quarter in the prior year ($Ex\ ante\ TD_{q,SM} = Sales\ Target_{q,SM} / Sales_{q-4, SM}$).
<i>Ex post TD</i>	A time-varying measure of <i>ex post</i> target difficulty at the DM level, measured as the quarterly average STG performance of the SMs reporting to the DM.
Control variables	
<i>Store Size</i>	Store square feet, divided by 1000.
<i>New Store</i>	Indicator variable that is 1 when the store was opened less than a year ago and 0 otherwise.
<i>Number HH</i>	Square root of the number of households in the store area, scaled by 100.
<i>HH Income</i>	Median income of households in the store area.
<i>STG</i>	Quarterly sales to goal performance: (sales-goal)/goal.
<i>SPH</i>	Quarterly SPH hit rate (0-1): number of weeks SPH within range / 13 weeks.
<i>Windfall Sales</i>	Indicator variable that is 1 in a quarter in which STG exceeds the target by 7% or more.
<i>Managerial Ability</i>	Store manager fixed effect from a regression (using data from before the new bonus plan introduction) of quarterly sales (<i>Sales</i>) on <i>Industry Growth</i> , and on firm characteristics of <i>Store Size</i> , <i>New Store</i> , <i>Number HH</i> , and <i>HH Income</i> .
<i>Preplan Bonus</i>	Bonus value in year before the new plan started.
<i>SM Change</i>	Indicator variable that is 1 in a quarter when the store has a new SM, and 0 otherwise.
<i>Plan Experience</i>	This construct is operationalized in three different ways; 1) L2Q: indicator variable that is 1 for last two quarters, and 0 otherwise (used in our primary hypothesis test). 2) L3Q: indicator variable that is 1 for last three quarters, and 0 otherwise. 3) Linear: 1 for first quarter, 2 for second, ...5 for fifth quarter.
<i>% Δ Salary</i>	The percentage change in the salary from before the bonus plan introduction to after the bonus plan introduction.
<i>Formula Override</i>	Indicator variable that is 1 in a quarter where an override of the bonus formula was made, and 0 otherwise.

References

Anderson, S.W., Dekker, H.C., Sedatole, K.L., 2010. An empirical examination of goals and performance-to-goal following the introduction of an incentive bonus plan with participative goal setting. *Management Science* 56 (1), 90–109.
 Aranda, C., Arellano, J., Davila, A., 2019. Subjective bonuses and target setting in budget-based incentive contracts. *Management Accounting Research* 43, 45–60.
 Albuquerque, A.M., 2014. Do growth-option firms use less relative performance evaluation? *The Accounting Review* 89 (1), 27–60.
 Baiman, S., Rajan, M.V., 1995. The informational advantages of discretionary bonus schemes. *The Accounting Review* 70 (4), 557–579.
 Baker, G.P., Gibbons, R., Murphy, K.J., 1994. Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics* 109 (4), 1125–1156.
 Banker, R.D., Datar, S.M., 1989. Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research* 27 (1), 21–39.
 Bol, J.C., 2008. Subjectivity in compensation contracting. *Journal of Accounting*

Literature 27, 1–24.
 Bol, J.C., 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review* 86 (5), 1549–1575.
 Bol, J.C., Keune, T.M., Matsumura, E.M., Shin, J.Y., 2010. Supervisor discretion in target setting: An empirical investigation. *The Accounting Review* 85 (6), 1861–1886.
 Bol, J.C., Lill, J., 2015. Performance target revisions in incentive contracts: Do information and trust reduce ratcheting and the ratchet effect? *The Accounting Review* 90 (5), 1755–1778.
 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.
 Bryk, A.S., Raudenbush, S.W., 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, London.
 Bushman, R.M., Indjejikian, R.J., Smith, A., 1996. CEO compensation: The role of individual performance evaluation. *Journal of Accounting and Economics* 21 (2), 161–193.
 Campbell, C.M., Kamani, K.S., 1997. The reasons for wage rigidity: Evidence from a

- survey of firms. *The Quarterly Journal of Economics* 112 (3), 759–789.
- Christ, M.H., Sedatole, K.L., Towry, K.L., 2012. Sticks as carrots: An experimental investigation of contract frame and effort in agency relationships. *The Accounting Review* 87 (6), 1913–1938.
- Core, J., Guay, W., 1999. The use of equity grants to manage optimal equity incentive levels. *Journal of Accounting and Economics* 28 (2), 151–184.
- Demeré, B.W., Sedatole, K.L., Woods, A., 2019. The role of calibration committees in subjective performance evaluation systems. *Management Science* 65 (4), 1562–1585.
- Demerjian, P., Lev, B., McVay, S., 2012. Quantifying managerial ability: A new measure and validity tests. *Management Science* 58 (7), 1229–1248.
- Fischer, P.E., 2000. Managing employee compensation risk. *Review of Accounting Studies* 4 (1), 45–60.
- 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.
- 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.
- Gibbs, M., Merchant, K.A., Van der Stede, W.A., Vargus, M.E., 2005. The benefits of evaluating performance subjectively. *Performance Improvement* 44 (5), 26–32.
- Hall, R.E., 2005. Employment fluctuations with equilibrium wage stickiness. *The American Economic Review* 95 (1), 50–65.
- Hayes, R.M., Schaefer, S., 2005. Bonuses and non-public information in publicly traded firms. *Review of Accounting Studies* 10 (4), 431–464.
- Holmstrom, B., 1979. Moral hazard and observability. *The Bell Journal of Economics* 10 (1), 74–91.
- Hoppe, F., Moers, F., 2011. The choice of different types of subjectivity in CEO annual bonus contracts. *The Accounting Review* 86 (6), 2023–2046.
- ICSC, 2005. ICSC Chain Store Sales Trends, May 5. International Council of Shopping Centers, New York.
- Ittner, C.D., Larcker, D.F., Meyer, M., 2003. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review* 78 (3), 725–758.
- Kreft, I., De Leeuw, J., 1998. *Introducing Multilevel Modeling*. Sage Publications, London.
- Lambert, R.A., Larcker, D.F., 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research* 25, 85–125.
- Long, J.H., Mertins, L., Vansant, B., 2015. The effects of firm-provided measure weightings on evaluators' incorporation of non-contractible information. *Journal of Management Accounting Research* 27 (1), 47–62.
- 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.
- Merchant, K.A., Manzoni, J., 1989. The achievability of budget targets in profit centers: A field study. *The Accounting Review* 64 (3), 539–558.
- Moers, F., 2005. Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society* 30 (1), 67–80.
- Murphy, K.R., Jako, R.A., Anhalt, R.L., 1993. Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology* 78 (2), 218–225.
- Murphy, K.J., Oyer, P., 2003. Discretion in executive incentive contracts. Working paper, University of California and Stanford University.
- Prendergast, C., 2002. Uncertainty and incentives. *Journal of Labor Economics* 20 (2), S115–S137.
- Prendergast, C., 2011. What have we learnt about pay for performance? *The Economic and Social Review* 42 (2), 113–134.
- Woods, A., 2012. Subjective adjustments to objective performance measures: The influence of prior performance. *Accounting, Organizations and Society* 37 (6), 403–425.