The Role of Subjectivity in Mitigating Incentive Contracting Risks

Anne M. Lillis*
University of Melbourne

Mary A. Malina
University of Colorado Denver

Julia Mundy
University of Greenwich

https://doi.org/10.2308/TAR-2017-0652

Running head: The Role of Subjectivity in Mitigating Incentive Contracting Risks

We are grateful to the Editor, Elaine Mauldin, and two anonymous referees for their constructive comments. We also thank Margaret Abernethy, Jan Bouwens, Mandy Cheng, Henri Dekker, Jennifer Grafton, Graeme Harrison, Michelle Hoggan, Kerry Humphreys, Chung Yu Hung, Xue Jia, Katherine Schipper, Paula Van Veen-Dirks, participants at the EIASM Performance Measurement and Control Conference, Nice, September 2015, workshop participants at ESSEC Business School, Vrije Universiteit Amsterdam and Macquarie University. Julia Mundy gratefully acknowledges financial support from the ICAEW's charitable trusts. Finally, we are grateful to the participating firms and managers for their time, insights and interest in the project.

***Corresponding author**

Anne M. Lillis, University of Melbourne, Department of Accounting, Melbourne, Victoria, Australia; Mary A. Malina, University of Colorado Denver, Business School, Denver, Colorado, USA; Julia Mundy, University of Greenwich, Centre for Governance, Risk & Accountability, London, United Kingdom.

The Role of Subjectivity in Mitigating Incentive Contracting Risks

ABSTRACT

We examine how subjectivity in performance measurement and reward systems (PMRS) is used to mitigate incentive contracting risks. Drawing on data from 38 interviews with supervisory and subordinate managers in four firms, we provide a more comprehensive explanation of the role of subjectivity in risk mitigation than is evident in the prior literature. We provide empirical evidence of the importance firms place on the use of subjectivity to mitigate the risk of incentive misalignment and employee sorting errors relative to its well-documented role in mitigating employee compensation risk. We find that incentive misalignment arising from unanticipated behavioral responses to performance measures is a particularly important risk, managed through subjective performance assessments. The extent of subjectivity we observe poses a significant risk of errors and bias. We observe that both vertical and horizontal information gathering and review by calibration panels are key strategies to mitigate the downside risk of subjectivity.

Keywords: subjectivity; incentive alignment; employee sorting; compensation risk; mitigating errors and bias in subjectivity; calibration

JEL Classification: M49, M55, M52 M51

I. INTRODUCTION

Recent global scandals reflect the dark side of incentive-driven behavior in organizations[1]. Incentives can lead to egregious behaviors because the underlying performance measures cannot be fully aligned with firm value creation, giving rise to a range of incentive contracting risks for firms and employees. Subjectivity in performance measurement and rewards systems (PMRS) is commonly viewed as a partial solution to these risks (Baker, Gibbons, and Murphy 1994). We draw on field studies in four large global firms to investigate the role of subjectivity in managing three incentive contracting risks. First, we investigate the role of subjectivity in addressing incentive misalignment, in which firms reward valued efforts that would otherwise be unrecognized or, conversely, avoid rewarding employees for behaviors or decisions that produce sub-optimal outcomes for the firm (Roberts 2010; Bol 2008; Gibbs, Merchant, Van der Stede, and Vargus 2004). Second, we examine the use of subjectivity to mitigate the risk of employee sorting errors that arise because performance measures also influence decisions about employee retention, promotion and task assignments (Prendergast 2002; Baker, Jensen, and Murphy 1988; Moers 2005). Third, subjectivity is used to mitigate employee compensation risk when performance measures are influenced by uncontrollable factors (Datar, Kulp, and Lambert 2001).

While subjectivity is introduced into PMRS to enhance incentive alignment and employee sorting and to reduce compensation risk, it also introduces its own risk of errors and bias associated with unverifiable subjective assessments (Moers 2005; Bol 2008; Luft, Shields, and Thomas 2016). We investigate the relative salience of incentive misalignment, employee sorting errors and compensation risk as rationales for the use of subjectivity (the

[1] E.g., Volkswagon in Germany; Wells Fargo in the United States; Barclays Bank in the United Kingdom; findings from the Royal Commission into Banking and Financial Services in Australia.

upside of subjectivity). We also investigate the strategies firms use to mitigate errors and bias associated with subjectivity (the downside).

Our primary motivation for this study is to gain rich field-based insights into the role of subjectivity within PMRS. Subjectivity is ubiquitous in PMRS within firms (Bol 2008; Woods 2012; Kampkotter and Sliwka 2015), yet we do not fully comprehend its role in mitigating incentive contracting risks (Gibbs 2016). In particular, the literature is very clear on incentive misalignment as an important contracting problem, and the theory literature proposes subjectivity as one solution. However, there is remarkably little empirical evidence of the use of subjectivity to enhance incentive alignment and employee sorting, relative to its role in managing compensation risk. Our study is designed to exploit the advantages of field site access to examine the rationale underpinning subjectivity in multiple PMRS practices (measures, ratings and rewards) occurring in combination, as well as the organizational practices that limit the downside risk of subjectivity. In doing so we answer Gibbs' (2016) call for case-based insights into the structures, patterns, methods of decision making and policies firms adopt that impact compensation.

We find that the four firms in our study use subjectivity in ways that have not been fully captured in empirical studies. We find that the firms rely extensively on ex post subjective assessments of subordinate behavior to minimize incentive misalignment and employee sorting errors. We observe a subjective overlay on objective measures and considerable discretion in weightings on objective measures and subjective assessments. Further, we find that subjective assessments tend to be non-lenient in that they either confirm or downgrade objective performance results that subordinates deliver in ways that are not consistent with broader firm goals. The subjectivity we capture reflects "ex post discretional adjustments based on factors other than the performance measures specified ex ante" (Bol 2008, 2; Gibbs et al. 2004). We find that the firms rely on these ex post subjective

assessments to both validate or realign single period bonus awards, and to determine task assignments, retention and promotion (employee sorting) decisions. In contrast to prior empirical research, the use of subjectivity to address employee compensation risk is a minor theme in our data.

The extent of subjectivity observed in these firms creates uncertainty for subordinates. Subordinates are unable to predict their bonus based on objective performance measures without also anticipating the impact of a subjective overlay on those measures. The firms manage this uncertainty, and the associated potential for errors and bias, by informing and validating subjective assessments. All four firms gather information on subordinate behavior vertically (through the direct supervisor), horizontally (through interdependent roles internal to the firm), and in some cases, externally through clients. In addition to information gathering, three of the firms adopt a calibration process in which peer and senior managers review and challenge subordinate ratings. The fourth firm offers an interesting contrast with less use of formal rating processes, no formal calibration process, more discretion at the bonus reward stage of the PMRS, and less transparency in its application of subjectivity. Despite these differences, the findings from this firm suggest that the rationale for subjectivity is consistent with the other firms, and reliance on high-level review of ratings is important to reduce errors and bias in subjective assessments.

Our findings contribute to the literature in several ways. First, our findings suggest that the use of subjectivity to address the risk of incentive misalignment and employee sorting errors may be dominant rationales for subjectivity in PMRS. While the use of subjectivity to address employee compensation risk is well understood, incentive misalignment is likely a more critical problem, but is less studied (Baker 2002; Feltham and Xie 1994). The use of subjectivity in PMRS to mitigate the firm risk of incentive misalignment is particularly difficult to assess empirically using archival methods because many forms of subjective

assessments are largely hidden to explicit scrutiny (Bol 2008; Gibbs et al. 2004; Ittner, Larcker, and Meyer 2003), and do not leave a clear trace from subjective assessments through to bonus outcomes[2]. Further, while the literature addresses the role of subjectivity in promotion decisions (e.g., Grabner and Moers 2013; Bol and Leiby 2018), it does not address the role of subjectivity in enhancing relative performance evaluation and employee sorting more generally. Economists argue that unreliable assessments of individual contributions to the firm result in employee sorting errors (retention, promotion and task assignments) that are more costly to the firm than single-period bonus payments (Baker et al. 1988; Lazear and Shaw 2007; Prendergast 2002). By focusing on identifiable subjective measures, or on subjectivity that can be traced from discretionary weights on objective performance measures, the accounting literature likely underestimates the level of subjectivity that underpins performance ratings. These ratings in turn influence both bonus determination and employee sorting. Our field study provides new insights into a broad range of subjective inputs to performance measurement and the utility firms gain from these subjective assessments.

Second, the literature has recently begun to consider the role of information gathering that underpins subjective assessments (Maas, van Rinsum, and Towry 2012; Bol, Kramer, and Maas 2016; Gillenkirch and Kreienbaum 2017), as well as the functioning of calibration panels as a mechanism to reduce errors and bias in performance measurement (Demeré, Sedatole, and Woods 2019; Kampkotter and Sliwka 2015; Arshad, Cardinaels, and Dierynck 2020; Grabner, Künneke, and Moers 2020; Bol, Braga de Aguiar, Lill, and Coelho 2018). We provide new insights into the way managers gather information horizontally, exploiting

[2] For example, flexible subjective assessments that draw on a range of objective performance measures and other information sources, and which may or may not alter "objective" performance ratings, are common in practice but very difficult to observe in archival data. In contrast, discretionary weights explicitly applied ex post to a limited set of objective measures impact on bonuses and are directly traceable in archival data.

interactions among interdependent roles internal and external to the firm to inform and validate subjective assessments both at the initial rating stage and in calibration panels.

Third, our findings provide insight into the use of subjectivity to reduce uncertainty about the veracity of performance measurement where there is potential for incentive misalignment or employee sorting errors. In many cases initial subjective assessments by supervisors confirm the ratings determined from objective measures. Further, calibration panels then typically confirm supervisors' initial ratings. Even when the information gathering underpinning subjective assessments produces few changes in performance ratings, it is valuable in reducing uncertainty regarding the performance signal from objective measures. This use of subjectivity to reduce uncertainty is hidden in archival data because the subjective assessments do not manifest in altered bonuses. In the context of calibration, our findings support those of Demeré et al. (2019) and Bol et al. (2018) by noting that ratings and bonuses change in a minority of cases during calibration. We extend this literature by exploring the value of calibration to firms and employees even when ratings and bonuses do not change.

Overall, the findings from our study suggest that extensive reliance on subjective PMRS can improve incentive alignment and employee sorting for managerial roles. However, exploiting the upside benefits of subjectivity requires that firms are also able to utilize reliable observations about employee performance and discipline evaluating managers against errors and bias.

II. LITERATURE REVIEW

Subjectivity and Contracting Risks

Incentive Alignment and Compensation Risk

Gibbs et al. (2004) and Bol (2008) identify three points in the PMRS process at which firms can exploit ex post subjective assessments. First, performance measures are defined ex

ante but measured ex post using non-verifiable information (for example, measures of "leadership", "collaboration" and "effective training of subordinates"). Second, the firm relies on objective performance measures, but subjectively re-weights these measures ex post to allow for changed conditions. Third, contracts allow for flexible discretionary adjustments based on a combination of performance measures and other information available ex post. Such adjustments allow that agent effort, action choices and outcomes are clearer ex post than they are ex ante (Baker et al. 1988). These approaches are not mutually exclusive and tend to be used in combination (Bol 2008; Gibbs et al. 2004).

These three forms of subjectivity, and their underlying rationales, have attracted different levels of attention in the literature. The theory literature focuses particularly on opportunities to either subjectively re-weight performance measures as conditions change, or to determine discretionary bonuses ex post to mitigate compensation risk (Baiman and Rajan 1995; Baker et al. 1994; Baker 2002; Feltham and Xie 1994). The empirical literature largely follows the theory literature by focusing on employee compensation risk and the associated need for discretionary adjustments, or discretionary performance measure weights (e.g., Liu and Leitch 2013; Bol, Hecht, and Smith 2015; Höppe and Moers 2011; Anderson, Dekker, Sedatole, and Wiersma 2020). Some economists suggest that while compensation risk is a common rationale for subjectivity in the literature, it is not a central issue in most incentive contracts in practice (Baker 2002; Prendergast 2002; Feltham and Xie 1994). Incentive misalignment is typically more challenging and costly to remedy than compensation risk (Feltham and Xie 1994).

Accounting studies that do attempt to operationalize the use of subjectivity to enhance incentive alignment are constrained by limited proxies in the context of both the incidence of subjectivity and the nature of incentive misalignment. Several studies focus on discretionary weights on measures defined ex ante (e.g., Ittner et al. 2003; Höppe and Moers 2011).

However, the use of discretionary weights only partially mitigates incentive misalignment as it assumes that the pre-contracted set of performance measures themselves are complete. Early studies focus on 'individual measures' or unexplained compensation outcomes as a proxy for discretionary rewards at the CEO level (Bushman, Indjejikian, and Smith 1996; Hayes and Schaeffer 2000). Recent studies examine the use of subjectivity in specific settings such as those requiring creativity (Grabner 2014), pursuit of a quality strategy (Van der Stede, Chow, and Lin 2006) or autonomous motivation (Kunz 2015), implying that subjectivity enhances incentive alignment in these settings.

Gibbs et al. (2004) is one of the first studies to access archival field data to assess the rationale underpinning subjective bonuses within firms. They find evidence of the use of subjectivity to address compensation risk, but they do not find a consistent association between subjective bonuses and incentive misalignment attributable to performance measure incompleteness, short-term focus or manipulability. They acknowledge that their archival proxies for the weaknesses in quantitative metrics are crude (including, for example, number of measures as a proxy for completeness). The field site Gibbs et al. (2004) study provides a distinctive subjective bonus which is identifiable and incremental to a formula bonus.

Several recent studies shed light on less constrained forms of subjective assessments, which do not reflect discretionary weights, specific subjective measures defined ex ante, or explicit discretionary bonus payments. In a footnote to Woods' (2012) study, he notes that in his setting, subjective adjustments to objective measures are designed to "align individual work with the organization's mission and priorities and account for factors that are necessary for effective, efficient work accomplishment that the measures alone may not achieve"[3]. This observation points to a broad subjective overlay on objective performance measures. Deller

[3] See footnote 4 on page 405. Woods (2012) sets out to examine supervisor behavior given a change in performance measurement practice. He identifies the incidence of subjectivity with a 0/1 indicator variable (present or absent).

and Sandino (2020) also assess the use of subjectivity in a field setting where subjective assessments are intended to reward honesty and consistent adherence to company values and goals[4]. These are important observations because they draw attention to a behavioral aspect of incentive alignment which has not featured in the subjectivity literature. The opportunity to assess performance subjectively ex post, drawing on non-contractible but decision-relevant information is valuable to firms, particularly to address the potential for incentive misalignment. Woods' (2012) and Deller and Sandino's (2020) observations of relatively unconstrained subjectivity address the risk of incentive misalignment highlighted by Baker et al. (1988, 598): "The principal knows, in general terms, what he wants the agent to do, but the range of possible actions that the agent can take and the range of possible outcomes is enormous". However, these flexible discretionary adjustments are difficult to capture empirically. Deller and Sandino (2020) claim to be the first to test the use of subjectivity to reward employees for "doing things the right way" in a tournament setting, suggesting there is little empirical evidence to date on the way firms use subjectivity to mitigate the risk of employee behaviors that adversely impact firm value.

Data limitations are a significant impediment in efforts to capture incentive alignment as a rationale for subjectivity. For example, in the context of complex tasks, the incompleteness of objective performance measures is a prime rationale for subjectivity, theoretically, intuitively and anecdotally, yet it is very difficult to capture empirically. Incomplete objective performance measures manifest in a range of ways which are well understood in the literature, including self-interest or a lack of co-operation in the firms' interest, gaming, shirking, as well as a range of behaviors inconsistent with firm values (Roberts 2010). However, the literature cannot consistently capture the use of subjectivity to address incomplete objective performance measures (e.g., see Anderson et al. (2020); Gibbs

[4] The focus of their study is a tournament scheme where subjectivity is applied to distinguish high performers.

et al. (2004)[5]). Another reason for the lack of empirical evidence on the use of subjectivity to address the incompleteness of objective performance measures is that the literature provides little insight into the most flexible form of subjectivity documented by Gibbs et al. (2004) and Bol (2008), which allows for flexible discretionary adjustments drawing on multiple information sources beyond performance measures specified ex ante. Both Woods' (2012) and Deller and Sandino's (2020) site descriptions suggest that this form of subjectivity could be important to firms in using ex post information about employee effort, performance and task execution to ensure incentive alignment. Beyond the descriptive evidence in these studies, there is little evidence in the literature of this use of subjectivity to address the risk that objective measures are deficient in ways that cannot be fully anticipated when contracting.

Employee Sorting

In addition to distributing rewards, the multi-purpose PMRS that underpin incentive contracts are also used to rank and sort employees. Routine performance measurement is an important input to employee sorting, for example, in determining employee retention (Woods 2012; Cichello, Fee, Hadlock, and Sonti 2009) and promotion prospects (Campbell 2008; Grabner and Moers 2013; Bol and Leiby 2018; Ederhof 2011). Unreliable assessments of individual contributions to firm value result in distorted assessment of employees' *relative* value to the firm. Employee sorting errors can be particularly costly because employees are likely concerned with the way their assessed performance affects their future in the firm, and the firm is concerned with retaining the right people and effective task assignments. These concerns may override issues about single period bonus payments (Baker et al. 1988;

[5] Anderson et al. (2020) proxy a control variable, task complexity, by store size and Gibbs et al. (2004) proxy performance measure completeness by number of objective performance measures. Neither study consistently documents the expected positive association with subjectivity.

Prendergast 2002; Moers 2005; Rankin and Sayre 2011; Campbell 2008; Gibbs 1995; Bol 2008).

Firms invoke subjectivity to facilitate relative performance evaluation and avoid costly sorting errors (Deller and Sandino 2020; Campbell 2008; Grabner and Moers 2013). The literature on the role of subjectivity in employee sorting focuses particularly on the subjective weights applied to non-financial measures (Campbell 2008) or current performance (Grabner and Moers 2013) in order to capture the different demands of current and higher-level roles. The focus of these studies is on the way objective and subjective performance measure inputs to promotion decisions may deviate from those used to determine current compensation (Bol and Leiby 2018; Cichello et al. 2009). However, they do not investigate the role of subjectivity in current performance measurement as a key input to relative performance evaluation and employee sorting more generally.

Subjectivity as a Source of Additional Firm-Level and Employee-Level Risks

Subjectivity in PMRS enables firms to improve incentive alignment and employee sorting, and reduce compensation risk compared with PMRS based solely on objective measures. However, subjectivity also has a well-established downside risk. Subjective assessments are potentially riddled with errors and bias (Bol 2008; Moers 2005; Luft et al. 2016). When managers make subjective evaluation decisions they weigh up personal costs and benefits, and they may lack incentives to invest in information acquisition to support accurate subjective assessments (Bailey, Hecht, and Towry 2011; Maas et al. 2012; Bol 2011). Biased or error-prone subjective assessments exacerbate rather than mitigate incentive misalignment by rewarding the wrong behaviors or outcomes and by reducing the clarity of performance expectation signals. Biased or error-prone assessments also introduce costly employee sorting errors because they impact on the attraction, development, and retention of

talent (Bol et al. 2016; Kampkotter and Sliwka 2015; Ahn, Hwang, and Kim 2010). Errors and bias undermine the potential for subjectivity to improve incentive contracting.

Optimizing the Upside/Reducing the Downside of Subjectivity

Firms face tensions in managing the upside of subjectivity, which can improve incentive contracting, and the downside in which subjectivity introduces additional risks due to bias and errors. Firms can optimize the upside value of subjectivity by adopting strategies to mitigate the risk of errors and bias in subjective PMRS. Much of the incentive contracting literature is built on the principle of unobservable agent action (Feltham and Xie 1994). However, supervision of subordinates is an important feature of large organizations and supervisor observations are an important input to subjective performance assessments. Observation and information gathering about subordinates (albeit imperfect) facilitates the use of informed subjectivity (Fisher, Maines, Peffer, and Sprinkle 2005; Maas et al. 2012; Bol 2016).

Recent empirical studies consider the role of monitoring as a key input to subjective assessments. Höppe and Moers (2011) find that board monitoring intensity is a determinant of the accuracy of unverifiable signals and is therefore critical to the value of subjective performance measurement of CEOs. Several experimental studies examine performance feedback and information gathering about subordinates as an input to performance measurement. For example, Fisher et al. (2005) introduce private supervisor information arising from observability of subordinate action. Gillenkirch and Kreienbaum (2017) find that supervisors invest in information gathering about subordinates which is more consistent with norm enforcement than efficient contracting and Maas et al. (2012) examine information gathering in team settings. Bol et al. (2016) consider information accuracy and transparency as mechanisms to reduce centrality bias. These studies all capture experimentally an aspect of information gathering which facilitates subjective performance measurement, but they do not

address the range of opportunities that exist within firms for supervisors to observe subordinate behavior directly and to exploit interdependencies within the firm to gather information which informs ex post subjective assessments.

Recent literature also examines calibration panels as a bias mitigation strategy within firms (Demeré et al. 2019; Grabner et al. 2020; Arshad et al. 2020; Kampkotter and Sliwka 2015; Bol et al. 2018). These studies document the role of calibration in standardizing performance benchmarks across supervisors and disciplining supervisors against leniency, but do not yet provide insights into the formal and informal information-gathering that underpins calibration.

Research Questions (RQ) Guiding our Field Study

In summary, we draw on field study data to shed light on several dimensions of the incentive contracting problem. First, we investigate the rationale for subjectivity. Based on extant literature, we distinguish three risks: incentive misalignment, employee sorting errors, and compensation risk. Our study design enables us to investigate subjectivity through a combination of PMRS practices, including the use of subjective measures, ex post subjective ratings which draw in information beyond performance measures, subjective assessments by calibration panels and subjective rewards which may or may not flow from ratings. Importantly, our field sites provide the opportunity to investigate subjective rating practices informed by both objective performance measures and other performance-relevant information. We are also able to investigate the information environment that underpins subjective assessments, the use of subjectivity to reduce uncertainty and the simultaneous management of both the upside and downside risks of subjectivity.

In RQ1, we investigate the rationale for the subjective PMRS practices adopted in our field study firms (the upside).

RQ 1: How is subjectivity used within PMRS to mitigate the incentive contracting risks of:

a) Incentive misalignment,

b) Employee sorting errors, and
c) Compensation risk (uncontrollable influences on performance)?

In RQ2, we investigate the strategies used to mitigate the risk of errors and bias introduced by subjectivity (the downside).

RQ 2: How do firms manage the risks of errors and bias introduced by subjectivity?

III. DATA COLLECTION AND ANALYSIS

Our study is designed with the aim of generalizing to theory[6]. That is, we collect data that speak to our research questions and allow for both literal and theoretical replications[7] across field study firms. The four firms we study are large, global, and diverse, each from a different industry: equipment manufacturing (EquipMan), food manufacturing (FoodMan), investment management (InvestServ), and professional services (ProServ)[8]. We selected firms that use some level of subjective evaluation in various stages of the determination of performance-based rewards. This excluded, for example, sites in which bonuses are based on sales or allocated on the basis of job title or position. We also required access to multiple supervisors and their subordinates. Further, we required that all interviewees had managerial roles. Initial discussions with a range of firms indicated that the four firms included in our study met these criteria, while also exhibiting sufficient differences in their PMRS to challenge us to consider the impact of different contexts (Lillis and Mundy 2005).

We conducted in-depth semi-structured interviews with two related levels of managers within each organization, predominantly through a "snowball" design. Our initial

[6] As expected in a qualitative field study, we make no attempt to generalize our findings to other firms. Rather, we draw on rich field insights that enrich and refine the theory base of the subjectivity literature.
[7] Literal replications refer to the potential to corroborate evidence by observing similar manifestations of the focal research phenomenon (e.g., subjectivity in PMRS) in different settings with relatively similar conditions. Theoretical replications refer to the potential to observe different manifestations of the phenomena in different settings characterized by different theoretical conditions (Yin 2014).
[8] All four companies have global operations and interviews were conducted in different countries. EquipMan interviews were conducted in the United States, InvestServ and ProServ interviews in the United Kingdom, and FoodMan in Australia. Given the global presence of these companies, the nature of the research questions, and minimal cultural differences among the US, UK and Australia, we do not believe cultural differences impact on our analysis.

entry point into each firm provided us with interview access to at least one senior manager at a business unit level, who then referred us to peer-level managers, as well as to their direct reports who were themselves supervisors for the next level below[9]. Upon completion of the scheduled interviews, we had evidence that our respondents within each firm consistently described the PMRS, and that no new themes were emerging. Further, early comparisons across the four firms suggested differences in the use of subjectivity across PMRS practices. Subsequent analysis revealed both common and idiosyncratic themes, at which stage we believed we had sufficient data to speak to our research questions. For simplicity, we refer to the two participating levels of management as supervisor and subordinate, while recognizing that both levels occupy management roles in their respective organizations. A copy of the interview guide is shown in Appendix A.

The interviews were audio-recorded and transcribed verbatim. Our data are drawn from interviews with 38 managers at two hierarchical levels (12 supervisory and 26 subordinate managers). We conducted interviews in three of the firms predominantly on site[10] while interviews in EquipMan took place by phone due to distance and/or manager availability. In some cases, we also had access to company documentation.[11] We have removed all names and other identifiers in order to preserve the anonymity and confidentiality of the firms and their employees. Table 1 reports demographic data for all participants.

Overview of Field Sites

[9] We acknowledge the potential for selection bias in the identification of respondents. This is inevitable in a field study of this type, where access in itself is a privilege. We acknowledge the potential effect of this bias in the limitations section.

[10] Three of the 12 interviews at FoodMan were conducted by telephone due to geographical distance. All other interviews at FoodMan, ProServ and InvestServ were conducted on site.

[11] Documentation included performance report pro-formas, evidence of bonus formula components, and individual performance measures and how they were captured. These documents, which were confidential and only provided for brief on-site viewing, were useful for validating and clarifying our understanding of each firm's PMRS, prompting site-specific questions and aligning our terminology with local practices.

The four firms in our study differ in the extent to which they adopt a formulaic approach to the determination of performance-based rewards[12]. At one extreme, InvestServ has separate bonus pools for each team, no prescribed formula for the allocation of bonuses, and a very opaque process. At the other extreme, FoodMan and EquipMan adopt a bonus formula which maps directly to a bonus allocation. ProServ is somewhat similar to FoodMan and EquipMan, but with added discretion in adjusting the bonus allocation. Importantly for the purposes of our study, these bonus schemes mask multiple layers of subjectivity in all four firms. In all cases, participants describe PMRS that encourage supervisors to override objective measures with subjective judgments regarding the way subordinates achieve objective performance results. Three of the firms have formal calibration processes, where the subjective judgments of other stakeholders can lead to the adjustment of initial performance ratings or bonus allocations submitted by supervisors for individual subordinates[13]. Bonus potential varies by firm, level of management, and over time as profitability influences the size of the bonus pool. Among the subordinate-level managers we study, bonus potential ranges from around six percent to 20 percent in EquipMan, ProServ and FoodMan; and is significantly higher in InvestServ in accordance with financial industry norms. Overall, the pay at risk in these firms is consistent with the level of pay at risk in other studies of discretionary bonuses using firm-level data (Moers 2005; Bol 2011; Gibbs et al. 2004).

Coding Scheme

In order to structure our analysis of a large volume of interview data, we commence with a coding scheme that classifies segments of narrative data to one or more key constructs

[12] The online Appendix provides a more detailed description of the PMRS in each of the four firms.
[13] The fourth firm, InvestServ, has a less formal process whereby senior management reviews and approves bonus allocations.

for analysis. Table 2 contains our coding scheme, including both code definitions and an example of each code from our field narratives.

We begin by identifying the PMRS practices in which subjectivity may be invoked as subjective measures, ratings and rewards[14]. Subjective measures are explicit performance indicators that are assessed based on non-verifiable information. Examples include "leadership" and "teamwork". Due to the complexity of the subjective rating processes in our data, we find it useful for analysis to separate two rating stages. In the first stage, supervisors rate subordinates based on information drawn from formal performance measures and other information they access. This information is typically condensed into a numeric scale. The second rating stage is a calibration process which introduces a further layer of subjective influence on ratings, based on peer supervisor and/or higher-level management review of ratings. Following two rating stages, we identify a rewards stage. While performance ratings flow through to rewards, we allow for discretion in the determination of bonuses or other rewards post-ratings.

Next, we distinguish three risks addressed by subjectivity: incentive misalignment, employee sorting errors and compensation risk. Respondents do not, of course, use the terminology of information economics. They do, however, explain the rationale for subjectivity in their narratives, and we code these explanations against each incentive contracting risk (RQ1)[15]. We also code narratives that reflect strategies adopted to mitigate errors and bias in the subjectivity within each PMRS process (RQ2). These strategies include information gathering to validate subjective assessments of individual subordinates (in both

[14] We acknowledge that target setting may also involve subjectivity. However, the subjective performance measurement and reward literature tends to focus on performance measures, ratings and bonus awards so we restrict our domain of observables to these key PMRS processes.

[15] Not all participants provided a rationale. Among those that did articulate a rationale, we were alert to the possibility of rationales for subjectivity outside incentive misalignment, employee sorting errors and compensation risk. We did not identify any rationales outside these three, other than a few isolated examples. Examples provided only once in our data include making subjective assessments based on '*what has been assessed in the past*' (EquipMan supervisor) and in an administrative function '*doing what needs to happen and there haven't been any [human errors]*' (InvestServ subordinate).

initial and final rating stages) as well as the function of calibration panels in collating information across subordinates. We use a qualitative data software package (NVivo) to code the data. Each author initially coded the interviews from their respective field sites but all coding was verified through multiple iterations by all three authors.

Analysis of Qualitative Data

Qualitative studies generally convey their evidence through the use of quotations along with researcher interpretations of broader patterns in data. It is difficult to convey the weight of evidence using this approach, and quotations can only constitute examples from extensive narratives that form the field study data base. We adopt a systematic method to identify narratives in which discussion of specific subjective practices (measures, initial ratings, final ratings, rewards) co-occurs with discussion of specific risk rationales and/or strategies to mitigate the risk of errors and bias. This technique, established in other fields (e.g., Guest and McLellan 2003), and introduced to the accounting literature by Malina and Selto (2015), allows us to identify patterns in the data based on the relative frequencies with which relevant codes co-occur in narrative segments. We rely on this analysis to guide us in relation to the salience of co-occurring themes. We thus avoid the influence of interesting outliers or unique observations[16] when analyzing prevalent patterns in the data.

We first obtain the raw frequency of each code by firm (shown in Table 3). These raw code frequencies are meaningless in themselves, but they form the basis of the quantification of co-occurring themes that follows. We interrogate the data with tools within the software to create a matrix indicating the frequencies with which coded narratives referring to each subjective practice co-occur with coded narratives relating to each rationale, and the incidence of strategies to mitigate errors and bias. Code co-occurrence is defined as "isolated

[16] We acknowledge that these unique and outlier observations may be interesting as case study examples of unusual situations or catalysts to research, but they are not necessarily a reliable basis for identifying recurring or potentially generalizable patterns in data.

pairs of codes applied to text segments and associated with a unique respondent" (Guest and McLellan 2003, 191). Co-occurrences proxy for interactions of concepts underlying the codes. Code co-occurrence reports show how "thematic domains, concepts or ideas are distributed within a data set, beyond simple frequencies" (Namey, Guest, Thairu, and Johnson 2008, 145). Table 4 presents the resulting co-occurrences in the data.

Table 4 identifies prevalent patterns in relations among the key constructs by indicating the simple counts of co-occurrences between codes across all four firms. Because more commonly used codes are likely to co-occur more frequently, we normalize the co-occurrences by constructing co-occurrence ratios, or C-ratios, (Malina and Selto 2015).[17] We use C-ratios to distinguish the relative salience of co-occurring themes. Table 5 presents the C-ratios in aggregate across the four firms. There is no standard for a level of C-ratio that is "significant" so the numbers in Table 5 are meaningful only in relative terms. It is important to note that, while an interviewer may influence the range of themes discussed by participants, the salience of *co-occurring* themes is driven by participants.[18] In Appendix B we provide the C-ratios by firm.

IV. FINDINGS

We address our two research questions by analyzing the narratives within each cell. Table 5, rows A through C report co-occurrence ratios which relate to RQ1, while rows D and E relate to RQ2. We address each row in sequence. Under RQ1, we investigate the subjective practices deployed by managers at each stage of the PMRS (measures, initial

[17] Co-occurrence exists when a qualitative datum, in whole or in part, is simultaneously covered by more than one code. Specifically, Code 1 (of frequency n_1) co-occurs with Code 2 (of frequency n_2) when Code 1 text overlaps or is overlapped by, contains or is contained by, or is identical to Code 2 text (with a frequency of $n_{1,2}$). Co-occurrence is measured by the co-occurrence ratio between two codes, or C-ratio, defined as $C_{1,2} = n_{1,2} / (n_1 + n_2 – n_{1,2})$. The C-ratio falls between 0 (no co-occurrence) and 1 (complete co-occurrence).

[18] More specifically, the risk rationale underpinning subjective practices was in all cases coded based on participant responses to open-ended "why" questions as shown in our interview protocol. Researchers did not refer to the economic rationales. Furthermore, the range of subjective practices, while available as prompts, were also more generally provided by respondents in response to open-ended questions as to "where discretion enters the process". Finally, we attach no meaning to the absolute number of thematic references or the magnitude of the C-ratios other than to observe relative salience and thus guide our discussion.

ratings, final ratings and rewards) and examine whether the rationale captured in the narratives about these practices reflects incentive alignment (Row A), employee sorting (Row B), or compensation risk (Row C). We organize our findings by these three risk rationales numbering the cells in each row A1 to A4, B1 to B4 etc. for easy reference.

Findings RQ1: How is Subjectivity Used within PMRS to Mitigate the Incentive Contracting Risks of Incentive Misalignment, Employee Sorting Errors and Compensation Risk?

Row A: Subjective PMRS Practices to Mitigate the Risk of Incentive Misalignment

Our analysis of the theme of incentive misalignment co-occurring with subjectivity in each PMRS practice results in C-ratios in Table 5, row A. We classify narratives within our data as relating to incentive misalignment when subjectivity is applied to assess an individual's contribution to the firm, including their contribution within their designated role and their compliance with firm-level behavioral expectations (an example is provided in Table 2). Table 5 indicates that the co-occurrence of subjective initial rating and incentive misalignment is a particularly salient theme in our data (cell A2). Further, the C-ratio for subjective initial rating and incentive misalignment is highly salient in all four firms (see Appendix B). Subjective measures, final ratings and rewards each indicate some co-occurrence with the theme of incentive misalignment.

Subjective measures/incentive misalignment - Cell A1. Discussion of subjective measures is not a strong theme in any of the firms in our study. Subjective measures are used, but participants do not necessarily consider them "subjective" because the firms require that the measures are as SMART[19] as possible. Thus, respondents refer to subjective measures infrequently in response to questions about where subjectivity or discretion enters the PMRS

[19] SMART is an acronym for Specific, Measurable, Actionable, Responsive and Time-bound. These measurable attributes help to clarify the criteria on which performance is assessed, but as the "correctness" of these measures cannot be determined by a third party, they are non-verifiable (Bol 2008). Examples from our data include the use of 20 bullet points describing traits and behaviors that underpin '*demonstrate integrity*' (ProServ subordinate) and 11 items to demonstrate '*capability*' (FoodMan subordinate).

process. We classify such measures as subjective because they rely ex post on unverifiable criteria. Examples from our sites include *'leadership'* and *'championing core values'* which figure prominently in all EquipMan narratives. Subjective measures emerge in our data as salient only in the context of correcting incentive misalignment. This is not unexpected. Subjective measures reduce (but do not eliminate) incentive misalignment by enhancing the completeness of subordinate performance measurement across difficult-to-measure task aspects:

> *[You need to] concentrate all your time on business development and bringing in new revenues but also delivering the work that you've got and doing the wider team stuff; because if everybody said 'no' to that, nothing would happen. (ProServ subordinate)*

Subjectivity in initial ratings/incentive misalignment - Cell A2. One of the strongest recurring themes throughout our interviews is distrust in the capacity of ex ante, relatively objective performance measures (and SMART subjective measures) to align incentives, particularly in relation to corporate values, collaborative team-based contributions, and long-term business growth. Subjectivity is invoked across all four firms to transform performance measures into a single numerical rating for each employee. Supervisors assess subjectively the behaviors subordinates adopt to generate measured performance outcomes on each performance measure and weight these behavioral assessments highly in ratings. Both supervisor and subordinate participants report that explicit performance measures capture '*what*' subordinates achieve, but do not capture the behaviors subordinates employ to achieve their objectives. Supervisors refer to these behaviors as the '*how*' – the extent to which results are achieved in ways which are '*consistent with firm values'* and are *'sustainable'* (FoodMan supervisor), and consistent with the firm's desired '*core competencies'* (ProServ subordinate). An EquipMan subordinate summarizes the importance of initial ratings in mitigating the risk of incentive misalignment arising from objective measures:

> *Going for pure performance metrics, you lose the view of 'how' people are behaving to deliver the work plan. Somebody may do it at the expense of everybody else in the company, especially technical stuff that encourages collaboration, information sharing and development of others; the elements cannot just be measured that easily. (EquipMan subordinate)*

Several participants explicitly link undesirable subordinate behaviors to poor firm outcomes:

> *I know you achieved your number but the way you achieved it was disruptive, it created the wrong shape to our profit and loss line. (FoodMan subordinate)*

Other examples that demonstrate the use of subjectivity in initial ratings to mitigate the risk of incentive misalignment include assessing whether the manager '*created new markets and drove sensible business development that won't immediately yield value'* (ProServ supervisor), '*having some skin in the game so that you work collaboratively for a team outcome'* and '*acting like owners*' (FoodMan subordinates), *'having impact'* (EquipMan supervisor), and contributing to '*making money'* (InvestServ supervisor).

In formulating ratings, supervisors exercise discretion in the relative weights placed on objective outcomes and subjectively rated behaviors.

> *You can deliver the most spectacular outcomes but if you don't meet the 'hows' you won't get an exceptional outcome. And in fact, if you get partially meets or unsatisfactory in the 'how' you may get an unsatisfactory overall. So there is slightly more concern with the 'how' than the 'what'. (FoodMan supervisor)*

> *Somebody who maybe ticked all the boxes in terms of their projects but had awful relationships within the workplace, that's a tricky one. (InvestServ supervisor)*

There is some evidence that the influence of outcomes versus behaviors on ratings may shift in some roles, during certain economic conditions, or based on hierarchical level. For example, a ProServ supervisor perceives a higher implicit weighting on meeting revenue targets at Partner (supervisor) level, potentially with behavioral consequences.

> *We're very focused on revenue and there are examples of things that I would consider very poor personal behaviors that I would rate very low, where people have been given strong ratings because their revenue is significantly better. (ProServ supervisor)*

Importantly, where a subjective rating on behavior overlays objective performance measures, the subjective rating on behavior is likely to confirm or reduce the subordinate's rating, and rarely increase it. This is particularly evident in FoodMan and EquipMan. Objective measures produce a formulaic rating, and then supervisors question the behaviors that underpin the rating and either affirm the rating, if behaviors are aligned, or adjust it downwards where they perceive self-serving behavior:

> *A lot of people get marked down because for one reason or another [their actions] might not align to the values and culture of the organization. (FoodMan subordinate)*

The narratives coded to this cell are particularly salient to RQ1, as they speak directly to a focus on subjective ratings as a means of reducing firm risk of incentive misalignment arising from a range of sources. These include the management of interdependencies (assessing collaboration and teamwork), exploiting growth opportunities (driving business development) as well as unanticipated behaviors that are inconsistent with firm values. The narratives are overwhelmingly consistent in tone around this rationale and it is the most salient theme in response to open questions about the role of subjectivity in the PMRS in all four firms[20]. It is also important to note that the rationale relating to interdependencies in these narratives is not about determining the contribution of individual employees, which would be a controllability or compensation risk issue. Rather the focus is on ensuring that managers collaborate with peers to produce firm-level outcomes; this is an incentive alignment issue[21].

Subjectivity in final ratings/incentive misalignment – Cell A3. A further layer of subjectivity is applied at the final rating or calibration stage, which occurs in EquipMan, FoodMan and ProServ. The purpose of this stage is to arrive at a final rating for each

[20] As evident in Appendix B and explained under cell B4, InvestServ deploys subjectivity at the rewards stage rather than at the rating stage (so it is not coded to Cell A2). The economic rationale is, however, remarkably similar.

[21] The two perspectives on interdependencies as a contracting issue are noted in Gibbs et al. (2004), but the literature has addressed interdependencies primarily as a compensation risk issue.

subordinate through a review of initial ratings. In all three firms, calibration panels include the immediate supervisor of the subordinates under review as well as several peer-level and/or higher-level supervisors. The primary role of calibration is to mitigate the risk of errors and bias in subjective assessments, which we address under RQ2. However, our coding indicates that the calibration stage also has a significant role in enhancing incentive alignment by providing an opportunity to assess contributions to the firm that might not be evident within the supervisor/subordinate dyad.

> *I think there are reasons why some people won't have met their targets but they'll still have made a huge contribution to the business and I think you need to discuss that in a group. (ProServ subordinate).*

Participants refer to calibration as an opportunity to draw out the *'organizational impact'* of a subordinate (FoodMan supervisor) or to discuss how an individual's performance *'rolls up'* to the firm level (FoodMan subordinate). A ProServ supervisor suggests that calibration panels assess whether a subordinate *'has done a good job for their client…but should be making a wider contribution to the business'* than the immediate supervisor has taken into account in the initial rating.

Calibration panels are also able to assess how individuals have managed intra-firm interdependencies which are important to firm-level outcomes.

> *Our Board of Directors…have a kind of open discussion and dialogue to say, 'look, am I right in my read of this staff member, are they a good team player, or are they only a good team player for me?' (FoodMan subordinate).*

> *Mostly, your manager evaluates your performance, but, then at my level, they get buy-in from the other business leaders who work with you, so they have discussions about your performance and how you supported them. (EquipMan subordinate)*

Subjective rewards/incentive misalignment – Cell A4. In three of the firms we study, subjective ratings determine bonus awards. Thus, we see little subjectivity introduced at the rewards stage. In these firms, subjective rewards are less salient than ratings or measures in addressing incentive misalignment (see Appendix B). However, incentive

misalignment is a highly salient rationale in the determination of bonus awards in InvestServ. At the end of the following sub-section related to Employee Sorting, we explain the approach taken at InvestServ. In this firm subjectivity is "bundled" in the determination of subjective rewards rather than formal ratings. The economic rationale for subjectivity at InvestServ is remarkably similar to the other firms, but the subjectivity is located in a different PMRS practice.

Row B: Subjective PMRS Practices to Mitigate the Risk of Employee Sorting Errors

Table 5, row B shows C-ratios resulting from our analysis of co-occurring PMRS practices and the theme of employee sorting. We classify narratives as relating to employee sorting when subjectivity is used to rank order individuals to differentiate them for the purpose of bonus, retention, promotion, or development (an example is provided in Table 2). Examination of Table 5, row B suggests that the employee sorting rationale underpins subjectivity in both final ratings (cell B3) and subjective rewards (cell B4).

Subjective final ratings/employee sorting – Cell B3. Cell B3 captures the use of subjective assessments by calibration panels to rank and sort employees. Subjectivity in the final rating stage mitigates the risk of employee sorting errors in two related ways. First, in the three firms with calibration processes, relative performance evaluation is used to compare employees at similar levels to peers elsewhere in the firm.

> *We make sure that you just don't evaluate one person on their behavior by themselves but if you benchmark against a peer group then you can tell whether that person should be a little bit above water or underwater and that way it becomes a little bit more objective relative to a peer group (FoodMan subordinate)*

> *If we look at this group, you can start to see the top performers who have been rated a 2, kind of as it were, float to the top. You can see that there's another group here, 2 lows, and then you start…[asking]… what's the difference between them and a 3. And so some will come down into the 3 high. There isn't a set of clear metrics where you can absolutely say, from the start, this person is a 3 or a 2. (ProServ supervisor*)

Several participants stress the importance of reducing errors when assigning high ratings:

There's more a tendency potentially to down-rate than overrate because the consequence of that in terms of you overrate somebody and they're promoted or whatever the case and then that individual doesn't perform. People are very concerned about that in our organization, so I think they err on the side of caution actually most of the time in terms of not overrating individuals. (FoodMan subordinate)

If we were a little harsher on who's truly a '1' you could…reward those at the top of that curve higher. And those are the people you clearly want to retain more than anyone else. (EquipMan subordinate)

As well as sorting individual employees through peer benchmarking, the firms also apply advisory distributions to the broad cohort subject to calibration. These advisory distributions further refine the sorting of employees by providing an expected bell curve for the final ratings. Participants in ProServ speak of having to hit a '*target distribution'*.

Ten per cent have got to get a four. So if you don't think you've done too badly, but frankly the rest of your peer group's done brilliantly, you could be subject to a four which is a pretty tough blow. (ProServ subordinate)

Final ratings drive both bonuses and performance feedback in EquipMan, FoodMan and ProServ, so subjectivity has a critical role in this process to sort employees based on their value to the business.

Subjective rewards/employee sorting - Cell B4. In EquipMan, FoodMan and ProServ, rewards largely follow ratings in a formulaic way. However, we also observe several different subjective interventions designed ensure that the "right" people are rewarded within the constraints evident in these firms. It is notable that discretion we observe in directly determining rewards relates to employee sorting, with little emphasis on incentive alignment or compensation risk. Subjective bonuses are not common at FoodMan and EquipMan but other subjective financial and non-financial rewards are salient in both firms. We observe some adjustments at the salary level, but managers "wrestle" within firm-level constraints in exercising discretion in relation to bonuses.

I've got someone on my team who is paid significantly in excess of the midpoint, but that's a conscious decision which reflects a number of things: the person's performance, potential etc. (FoodMan supervisor)

> *You've got some very strong people with specific technical knowledge or experience that would be a huge loss to the company if we lost them…they're not going to move faster than everybody, they're so senior now. They're not going to be the CEO. But how do you retain them? Maybe they're really good and you want high performance to continue, so that's the thing we wrestle with. (EquipMan supervisor)*

In the absence of significant discretion over performance–based bonuses, EquipMan and FoodMan both adopt alternative rewards to ensure the right people are rewarded and retained. These include meaningful '*values in action'* award payments (FoodMan) as well as non-financial rewards such as mentoring, training or travel (FoodMan). Sometimes rewards are promised in the future:

> *It's a retention mechanism. You're doing great now, we love you and want to keep you, and in two or three years we'll give you money and options for stock. (EquipMan subordinate)*

ProServ allows for greater discretion in bonus determination than either FoodMan or EquipMan, with discretionary bonuses decided at a higher level of management:

> *Those couple of individuals at that border, who were quite hard done-by by the forced distribution, those people were given a bit more bonus than a '3' would automatically result in. (ProServ subordinate)*

> *If the formula says they ought to get 3%, they probably won't because the view will be taken that they're already paid too much for what they do, so they won't get that (ProServ subordinate)*

There is a second salient theme in this cell, which relates specifically to InvestServ. We address this contrasting finding in the next subsection.

InvestServ. InvestServ does not have a formal rating and calibration process, but rather applies much more discretion in the determination of bonuses. Consequently, we observe a "bundling" of discretion in subjective rewards at InvestServ which follows a very similar rationale to that documented in relation to ratings in the other firms[22]. Participants discuss similar themes of incentive alignment, by rewarding individuals for their attitude,

[22] See Appendix B, C-ratios in column 4 applying to InvestServ.

flexibility and going the extra mile for the company. Echoing earlier themes related to employee sorting, they refer to bonuses for excellence suggesting '*it's a tool for staff retention'* (InvestServ supervisor). Several participants capture the application of direct bonus discretion to keep employees, *'I've been head-hunted' and they'll say, 'well, we'll up your proportion [of the bonus pool]'* (InvestServ subordinate). Others refer to the alignment of rewards with an employee's long-term potential net contribution to the firm:

> *The other thing I take into account is if that person isn't happy, and they're going to move because of money; how much is it going to cost to replace them? You've got the time and effort and training involved, and all of that has kind of got a price attached to it …I also think about the potential of the individual, so they are taking all their exams, they're clearly bright, then I'll think, okay, I really need to keep them. I may have to overpay them for a couple of years. (InvestServ supervisor)*

In several cases within one team, the transparency or accountability around the exercise of direct discretion over rewards was questioned. One participant referred to a higher-level manager rewarding a subordinate as '*a very valuable guy to the firm and he wants to keep him sweet. I thought hmmmm Ok that's quite interesting to know'* (InvestServ subordinate). While InvestServ has a form of executive oversight over bonus payments, it lacks the formality and transparency observed in the other firms. There is no explicit calibration, but managers argue the case with senior managers for their subordinates to '*get this guy what I think he deserves'* (InvestServ subordinate). The lack of transparency and formalization at InvestServ suggests greater risk of errors and bias and less predictable compensation outcomes for employees than we observe in the other firms. We return to this theme in the discussion section.

While the compensation literature focuses largely on subjective bonuses, our findings in this cell report that three of the firms apply little discretion at the rewards stage. The limited discretion that is available is focused on a sanity check on bonus outcomes emerging from the rating and calibration process. InvestServ on the other hand applies considerable discretion at the rewards stage and avoids much of the administrative burden that surrounds

performance ratings and calibration. Importantly, in the context of RQ1, the rationale behind subjectivity is similarly focused on both incentive alignment and employee sorting, whether dispersed through PMRS practices or bundled in the determination of bonuses.

Row C: Subjective PMRS Practices to Mitigate Compensation Risk

Table 5, row C shows C-ratios resulting from our analysis of co-occurring PMRS practices and the theme of using subjectivity to address compensation risk due to the impact of uncontrollable factors on subordinate performance. As is evident in row C, the use of subjectivity to address compensation risk is not a strong theme in our data. It is a moderately salient theme in arriving at initial ratings (cell C2) in EquipMan and FoodMan, and in subjective rewards (cell C4) in InvestServ (see Appendix B). The narratives relating to compensation risk (cell C2) are consistent with the extant literature on this theme:

> *Folks having an understanding, discretion, realization that things do change and there are surprises, exceptions, challenges, changes in the business…. (EquipMan subordinate)*

> *Markets like ours...can go to the region and to [Head Office] ultimately and say 'I've had an extraneous circumstance that was beyond our control and I'm applying for relief for our business unit against the budget'. (FoodMan supervisor)*

Others refer to making judgements as to whether to '*re-contract'* or not (EquipMan subordinate) and the development of '*informal agreements'* around targets as project portfolios change during the year (FoodMan subordinate).

Importantly, this is a considerably less salient theme than incentive misalignment, and less salient than employee sorting in all four firms. Furthermore, it is notably absent as a theme in ProServ. We suggest contingent determinants of the relative importance of compensation risk as an avenue for further research.

Findings RQ2: How do Firms Manage the Risks of Errors and Bias Introduced by Subjectivity?

To be rendered informative, subjective assessments must be relatively free of bias and error. Under RQ2 we outline the strategies perceived by participants to mitigate errors and

bias when making subjective assessments. This question of how firms manage the downside risk of subjectivity is particularly relevant given the extent of subjectivity in PMRS practices that we report in our analysis under RQ1. To examine risk mitigation strategies salient to RQ2, we draw on the narratives underpinning the cells in Table 5, rows D and E. We focus particularly on cells D2, D3 and E3, which have high C-ratios.

Risk Mitigation in Initial Ratings/Individual Subordinates – Cell D2.

An important element in subjective adjustments to ratings based on adverse behaviors is the reliable observation of a subordinate's behavior. Participants refer to the importance of direct observation feeding into the behavioral dimension of performance ratings. References to supervisor observations include '*what she sees and what she hears, how I'm demonstrating [those behaviors]*' (EquipMan subordinate), *'it's very much around observation and feedback from key members'* (FoodMan subordinate) and *'because of how we're structured… [knowing] exactly what each person is… doing each day'* (InvestServ subordinate). In addition, all four firms adopt formal processes of seeking feedback from internal stakeholders, colleagues on interdependent tasks, junior and senior staff, which is incorporated in initial ratings. Sources include 360-degree feedback (FoodMan), internal surveys (ProServ), and gathering information from clients (ProServ). Examples include:

> *I'll seek a lot of input from stakeholders… I think I talked to 8 or 10 people for every one of my direct reports on how well they were doing. (EquipMan subordinate)*
>
> *The reality is that… if the guy's lazy his laziness will show through. You won't be able to pull the wool over management's eyes or let them have the perception that he's a really hard worker, a star performer, but in reality he's not. We're all going to find out about it.' (InvestServ subordinate)*
>
> *Don't promise [a rating]… you say to them, this is what I'm observing, I'm going to check in with the other managers. (FoodMan supervisor)*

Contrary evidence. The data coded to this cell also capture flaws in the feedback process which renders it an imperfect risk mitigation strategy:

> *You are required to request feedback from stakeholders as well as direct reports. To be honest, the challenge with that is, rarely is it that useful, because people normally give good feedback, right? (EquipMan supervisor)*
>
> *The biggest thing is people not prioritizing getting feedback from people they know they're not going to get good feedback from…. But I think that the moderation group are quite savvy to that. (ProServ subordinate)*

Risk Mitigation in Final Ratings – Cells D3 and E3.

These highly salient cells capture the role of calibration in mitigating the risks of errors and bias in subjective assessments. While calibration introduces a further layer of subjectivity, it is evident in the narratives outlined below that supervisors, and more importantly subordinate managers, believe calibration resolves rather than exacerbates the risk of errors and bias.

Final Rating/Individual Subordinates- Cell D3. The calibration process allows for additional opportunities for informed subjectivity by surfacing a broader set of observations and feedback about individual subordinates than in the initial ratings process. Narratives highlight information sharing from peer and cross-functional supervisors:

> *So their results were outstanding and as soon as you got into calibration you had all the other functions saying yeah but nobody wants to work with that person, he's just burnt every bridge…. The person just tumbled back down on that basis. (FoodMan subordinate)*
>
> *If people in the room feel, why is S a 2, I thought she would be a 3, I've had issues with her this year, I don't think she's meeting expectations. (EquipMan subordinate)*

Importantly, while information tabled at calibration comes from outside the supervisor/subordinate dyad, participant managers talk about interdependencies within the firms that lead to well-informed observations and contributions to the discussions about individual subordinates:

> *We should know those 150 people [discussed at a calibration meeting]… but 10-20% we'll know really well 'cause we're cross-functioning with them' (FoodMan supervisor).*
>
> *There's always at least four or five people that know enough about the individual in question that we're talking about. (ProServ subordinate)*

These findings point to the use of calibration panels to draw on cross-functional interdependencies to inform subjectivity in a way that is not highlighted in other studies.

Final Rating/Across Subordinates - Cell E3. We find that calibration mitigates the risk of errors from the use of different performance benchmarks by facilitating comparison among supervisors. Participants highlight the role of calibration in ensuring that '*everybody's exercising the same level of fairness and judgement'* (ProServ supervisor*),* or of helping supervisors to see their own ratings as '*harsh'* compared to their peers (EquipMan subordinate; FoodMan subordinate). An EquipMan subordinate uses the term '*juried'* to describe the process, noting that calibration adds both another layer of subjectivity but also enhances objectivity in '*the way it is done'*. A ProServ supervisor and an EquipMan subordinate both refer to increasing '*fairness'* through a group perspective on individual subordinates' performance. Participants perceive calibration as '*weeding out the bias'* (EquipMan subordinate) and reducing the subjectivity of the rating process.

> *To me the calibration is kind of the secret sauce. If you can calibrate well, and you have the right environment and the time to calibrate well, then it's great. But if you don't, then all the other failure modes could slip in.* (*EquipMan supervisor)*

> *The whole concept of the [calibration] is to try to take judgement…or bias out of the system as much as we can. (ProServ supervisor)*

Calibration is intended to control leniency bias. This is achieved primarily by providing a forum in which ratings can be challenged for "evidence", not necessarily by managers with direct knowledge of the subordinate under discussion.

> *If you're one of those folks that, you know, everybody you got is fantastic, you've got to sit in that room and tell your peers why Joe and Jane are so much better than John and Sue. And you force the dialog back to things that are specific to what that employee has done. (EquipMan subordinate)*

While the role of calibration panels in standardizing benchmarks and mitigating leniency bias is documented in the literature (Demeré et al. 2019), our field data provide further insight into the way calibration is used to manage the downside risk of subjectivity. In

many cases, calibration implicitly confirms initial ratings. While both supervisors and subordinates were keen to discuss the circumstances in which calibration would alter ratings, it is clear that this is not the norm, and for the majority of cases before the calibration panel the process also performs a critical confirmatory function. We document the incidence of changed ratings at approximately 10% in ProServ[23] and it is clear in all three firms that the majority of ratings are robust to challenge[24]. The ratings of employees who meet expectations (the majority) are rarely challenged, '*sometimes*' ratings at the top of the distribution are reduced, and *'rarely'* are lower ratings increased.

> *Feedback across a range of stakeholders …either confirms or balances the view of the manager in terms of the final rating. (FoodMan supervisor)*

Overall, there is evidence of substantial investment of resources in risk mitigation practices to reduce errors and bias, but in many cases will not impact ratings in any observable way. While other studies also note that calibration frequently does not change ratings, our data provide insight into the importance to participants of uncertainty reduction when ratings do not change.

> *At the end of the day, we have been through a process, and all the assessment is done, it's a drill down on the individual and it's an assessment that says that person has been evaluated fairly based on their performance or potential. (EquipMan subordinate)*

> *I might say okay, in my head I'm forming a provisional rating. I carry that into the room, in the moderation meeting, and then I see corroboration for that, with other people, who have, you know, worked maybe more closely with these people, etc, (ProServ supervisor)*

Consensus-informed confirmation of initial ratings may also have positive benefits for subordinates.

> *When you vet it with other people and they all agree, it's a way to really incentivize and motivate people. (EquipMan subordinate)*

[23] While we do not have specific estimates for FoodMan and EquipMan, the perception of participants was consistent with altered ratings in a minimal proportion of cases.
[24] Our findings are consistent with Bol et al. (2018), who document adjustments in 10.9% of cases. Demeré et al. (2019) document adjustments in 25% of cases.

Reducing uncertainty around performance evaluation reduces the risk of incentive misalignment and employee sorting errors, which are the overriding concerns of the participants in our study.

Contrary evidence. While calibration is widely discussed as the primary strategy to mitigate the risks of errors and bias in subjective assessments, participants also identify several biases introduced by calibration. First, the focus of calibration is invariably on avoiding over-rating and challenging top performers. However, given the focus on top performers, calibration mitigates leniency only at the top rating level. Leniency biases in initial ratings that are present in the "meets expectations" category are less likely to attract attention at calibration. Consistent with Demeré et al. (2019), we observe (and managers confirm) that the disproportionate focus on challenging top performers is likely to increase centrality bias.

> *If you rate an individual average, usually people wouldn't really challenge… so maybe what the discretion is doing is actually averaging. (FoodMan subordinate)*

> *Sometimes we move people down (from the top group). We then move to the bottom group, same kind of discussion; occasionally we move somebody up, it's usually less likely. (EquipMan supervisor)*

Second, several managers identify sources of error and bias that are inherent in the calibration process. While declaring that the process is '*not about advocacy'* (ProServ supervisor), subordinates note the impact of variation in the ability or propensity of supervisors to skillfully and freely debate in calibration sessions. Supervisors that '*fight harder'* (ProServ subordinate; EquipMan subordinate) or '*sell their people very, very well'* (FoodMan subordinate) are influential. Alternatively, a FoodMan subordinate suggests that managers with biases against employees may choose '*not to be very vocal about their cases'* in calibration meetings. Subordinates also note the potential for information to be '*over-indexed'* (FoodMan subordinate), where challenges at calibration are driven by isolated events (EquipMan subordinate) or by people with little working knowledge of the

subordinate being rated, particularly in a global setting where peer supervisors and subordinates are geographically dispersed (EquipMan supervisor; FoodMan subordinate). At ProServ, participants comment on the lack of transparency – a theme which is not evident at EquipMan or FoodMan.

> *Your rating, you're not in that meeting so you don't see that…In practice it probably is fair, but the perception is where it brings it down. (ProServ subordinate).*

Our observations are consistent with Demeré et al.'s (2019) survey data, indicating that subordinates largely trust calibration outcomes (outcome justice) despite some concerns about transparency of process (procedural justice).

Finally, in addition to the biases articulated by our participants, analysis of our data reveals a further bias not acknowledged by managers. The strong belief in the effectiveness of calibration in weeding out bias and errors, and the perception of reduced subjectivity though calibration cited earlier, suggest that participants view consensus as a surrogate for accuracy.

> *Managers can use as much discretion as they like, but they have to recognize the reality – the [calibration] will at the end of the day catch them out if they want to be irrationally subjective. (ProServ supervisor)*

While the correlation between consensus and accuracy may be high, there may also be a false sense that the results of the calibration sessions are reliable and objective, when in essence calibration reflects a further layer of subjectivity that reduces some errors and biases and generates others.

V. DISCUSSION AND CONTRIBUTIONS TO THEORY

The analysis of PMRS practices in our field sites suggests that objective performance measurement has inherent weaknesses that the firms try to address with an extensive subjective overlay on objective performance measures. We draw our findings from a broad cross-section of roles and levels of management across four firms in a variety of industry settings. While we cannot claim the findings from these firms are generalizable, the firms are

not unusual in their task or management profile. The performance ratings and rewards we observe draw on objective performance signals and a broad range of subjective cues.

Incentive alignment and employee sorting emerge as the most salient rationales for subjectivity in the PMRS in our field study firms. The firms invest in extensive information gathering to test the extent to which subordinate objective performance measures capture contributions to the firm, and they indicate preparedness to reduce performance ratings (and rarely increase them) if values are not adhered to, or behavior is self-serving or disruptive to the business. These observations are consistent with addressing incentive misalignment through information gathering, and flexible subjective assessments rather than through the design of better performance measures, or through discretionary weights on objective performance measures. We find that the firms manage employee sorting in a similar way. They subjectively rank employees with a view to managing talent, identifying individuals for retention, promotion and development.

In contrast, we observe that the use of subjectivity to mitigate compensation risk is a less salient theme in our data. Nonetheless, this use of subjectivity is a moderately salient theme in both FoodMan and EquipMan as part of the initial rating process (see Appendix B). Compensation risk is similarly evident as a rationale in InvestServ in the determination of bonuses – which is in line with that company's "bundled discretion". It is not a salient theme at all in ProServ. When asked open questions about the role of subjectivity, respondent rationales focus on issues of incentive misalignment and employee sorting rather than compensation risk. The fact that we observe variation in the salience of this theme across our field study firms suggests that future research may investigate contingent factors which would alter the focus on compensation risk, incentive misalignment and employee sorting in different settings.

These findings complement the existing literature in several ways. We provide the first (to our knowledge) empirical evidence of the importance of the risk of incentive misalignment and employee sorting errors relative to employee compensation risk as the rationale for subjectivity in PMRS within firms. Firms struggle to align incentives through the ex-ante design of performance measures (Baker 2000). Further, many instances of misalignment result from unanticipated behavioral responses to objective performance measures[25]. Empirical evidence of this use of subjectivity is sparse in the accounting literature. Our field narratives complement Deller and Sandino (2020) in capturing the use of subjective assessments to manage the unanticipated behavioral misalignments that plague performance measurement practice. In addition, while the accounting literature examines the role of subjective adjustments in PMRS for employee retention (Woods 2012) and promotion (Grabner and Moers 2013; Bol and Leiby 2018), we find that employee sorting through relative performance evaluation is also an important rationale underpinning the use of subjectivity more generally in routine PMRS. Our findings suggest that by focusing on compensation risk, discretionary weighting of performance measures, or identifiable subjective measures, the accounting literature may underestimate the level of hidden subjectivity in performance rating practices, and the utility of this subjectivity in mitigating the firm risk of incentive misalignment and employee sorting errors.

An important aspect of the form of subjective rating practices we observe is that subordinates in our settings could not be influenced by the expectation of rewards associated with achievement of specific objective performance goals. Given the subjective overlay on objective performance measures, subordinates wanting to choose actions to maximize their

[25] Several documented scandals can be traced to behavioral responses to well-intentioned performance measures. Examples include opening fake accounts to meet sales targets at Wells Fargo in the United States (Fritter 2020) and Commonwealth Bank of Australia (Ferguson 2018), and Australian police falsifying random driver alcohol tests by testing themselves in order to meet daily test volume targets (Mills 2018). While firms anticipate a volume-driven response to such targets, they do not anticipate fraud.

bonus or promotion prospects would need to anticipate the way two successive subjective rating processes would either alter or confirm their performance rating. This substantially dilutes the decision influence of objective performance measures (Luft et al. 2016). These observations are unlikely to be unusual, with meta analyses in the personnel psychology literature suggesting that the correlation between performance ratings and objective measures is approximately 27 to 39 percent (Heneman 1986; Bommer, Johnson, Rich, Podsakoff, and Mackenzie 1995). To the extent that this level of subjectivity is commonly interposed between performance measures and rating-driven bonuses, the decision influence of accounting performance measures can be better understood by considering these discretionary adjustments.

Our findings also shed light on the information environment that underpins informed subjectivity. We find in all four firms that the application of subjectivity in the determination of ratings and rewards is informed by both formal and informal information gathering practices. This is particularly prevalent in gaining insight into the behaviors subordinates adopt to achieve their performance outcomes. Supervisors observe subordinates as part of everyday practice, but they also supplement these observations with information from the subordinate's team members, peers and stakeholders on interdependent tasks. This information gathering is an important facilitator of informed subjective assessments.

Höppe and Moers (2011) introduce information gathering in the form of board monitoring as a variable in incentive contracting. Other experimental studies identify specific forms of information gathering that facilitate subjective assessments (Fisher et al. 2005; Gillenkirch and Kreienbaum 2017; Maas et al. 2012). However, we offer the first field-based evidence of the relationship between observability, information gathering and subjectivity in PMRS in typical large firms. We note the importance of interdependent roles within firms in generating information outside the supervisor/subordinate dyad, which is critical to both

initial ratings and calibration. This finding complements Demeré et al. (2019), who highlight the important advocacy role of the immediate supervisor on calibration panels. Our findings also complement the literature that addresses supervisor incentives to invest in information gathering to support subjective assessments (Bailey et al. 2011; Maas et al. 2012; Bol 2011). While we do not have insights into the incentives to gather information in our field sites, we do observe widespread information gathering. We also document potential flaws in the quality of this information at the initial rating stage, where feedback may be selective or sanitized. Information quality appears more robust at calibration when it is subject to review by peers and higher-level mangers. Our observations complement Grabner et al. (2020), who capture a higher-level disciplining effect of calibration on supervisors that over-rate their subordinates, and Arshad et al. (2020), who note that calibration provides an incentive for supervisors to gather information to support initial ratings.

An important contribution of our findings is the use of subjectivity to reduce uncertainty about the veracity of performance measurement when there is a risk of incentive misalignment or employee sorting errors. Within the supervisor/subordinate dyad, initial ratings on objective measures are reviewed subjectively (drawing in information about subordinate behavior), but this review frequently confirms ratings based on objective measures. Further information is gathered at calibration and again, this information most frequently confirms initial ratings. This suggests a role for subjectivity in reducing firm-level uncertainty about objective ratings, which is quite different to the commonly understood role of subjectivity in mitigating the risk to employees of uncertainty about the impact of uncontrollable events (compensation risk). The use of subjectivity to validate objective performance measurement is not evident in archival data that relies on rating changes to infer the presence of subjectivity (e.g., Ittner et al. 2003; Woods 2012). We are not aware of the

prior documentation of uncertainty reduction as an economic role for subjectivity, yet it is important in a risk mitigation model.

While subjectivity is extensive across all four firms, it is introduced in different ways. In the broad narratives underpinning the PMRS description in EquipMan, FoodMan and ProServ performance measurement protocols rely on regular communication, feedback and "no surprises". Resultant bonuses are relatively predictable as the year progresses. On the other hand, the discretionary allocation of bonuses at InvestServ allows the potential for less transparency, which one team within the firm noted. Importantly, InvestServ also has far less resource-intensive practices to mitigate errors and bias in subjective assessments. Our findings point to potential trade-offs in determining the investment in resource intensive strategies to mitigate errors and bias, and the level of transparency in subjective assessments. Raising similar issues, Ittner et al.'s (2003) study documents a subjective PMRS characterized by lack of transparency, perceptions of favoritism and ultimately, a fall-back to a focus on financial performance. Contrasting Ittner et al.'s (2003) findings and the findings across our four participant firms, we suggest that the trade-offs firms make to optimize the upside benefits of subjectivity and the investment required to mitigate errors and bias, is a fruitful avenue for further research.

VI. LIMITATIONS

Our findings are naturally limited in scope and generalizability as they are a function of the PMRS in the small set of firms we study. Our sample comprises four stable PMRS in well-established and successful firms. Somewhat different patterns of behavior and uses of subjectivity may be observed in other settings. We have selected firms across a range of industries but the PMRS we observe are by no means exhaustive or necessarily typical. Even within that subset of firms we acknowledge that we do not have a complete picture of the four PMRS. We are constrained to observe the PMRS in these firms through the eyes of the

limited set of informants we were able to interview and any archival documentation we received. Furthermore, we were only able to speak with supervisors and subordinates identified by the firm. This bias in respondent selection increases the risk of exposure to a sanitized account of reality. However, the responses to our questions are varied, credible and supported by examples. We have cited, across all firms, negative aspects of both processes and outcomes.

Our method of data analysis also carries limitations. In order to align field narratives with theoretical constructs, a level of interpretation is necessary. While we rigorously validated each other's coding many times, we acknowledge that the links between narratives and theory are not definitive. On the other hand, the patterns we observe are relatively robust to differences in interpretation at the margin. We also rely on the relative salience of narrative themes to guide our analysis of the dominant economic rationales for subjectivity. We acknowledge that this is a suitable instrument for distilling themes in narrative data, but a blunt instrument to assess relative importance. We are unable to measure the relative economic value of subjectivity, the ultimate impact on the firm, or on compensation outcomes.

While we acknowledge these limitations, we believe this study makes a further contribution by bringing new methods to the question of understanding subjectivity in performance measurement. This cross-sectional field study provides a unique insight into the rationale for subjectivity, the forms it takes, its consequences and the perceptions that surround its use.

APPENDIX A
Interview Guide

Italicized questions represent further prompts to be used if necessary.

Section 1: Demographics:

- Role title and description
- Time in role
- Time with company
- Supervisor (e.g., profit center manager or higher): Number of direct reports (= number of people whose performance you evaluate?)
- Subordinate: Level of management you report to
  - Who evaluates your performance?
  - How many managers are at a similar level and subject to similar evaluation?

Section 2: Perceptions of corporate performance measurement practice:

Supervisors and subordinates: What is the corporate approach to performance measurement at your level of management? Note - subordinates were asked to focus on their perceptions as a subordinate, not as a supervisor for their own subordinates.

- Frequency of performance measurement and evaluation
- Process
  - How is an overall evaluation determined?
    - *Is there a formula that converts several measures to a performance "grade"?*
    - *What measures are used? Are these standardized across managers at the same level?*
    - *How does discretion enter the process? E.g., the measures themselves are qualitative? The weightings among measures can vary? Some objective measures are supplemented with some qualitative measures? The measures are objective, but the overall rating is subject to discretion?*
  - Are bonuses/rewards dependent directly on performance measurement?
    - *How?*
    - *Does discretion enter into the process of converting performance ratings into rewards?*
  - What do you think is the rationale behind the process adopted at X company?
    - *E.g., the balance between objectivity/discretion?*
    - *The relationship between performance measures, performance evaluation, and rewards (both direct bonuses and other rewards such as promotions)?*
- To what extent does the corporate practice of performance measurement bind you as a manager when you are evaluating subordinates?
  - Are there ways in which you can introduce additional discretion into the process?
    - *Do you try to introduce more or less discretion into the process? Why?*
    - *How do you do this?*
    - *Do you think other managers at your level introduce more or less discretion into the process?*
- What are the most critical information sources you access when formulating judgments and evaluating performance? *E.g., peer assessment, internal reports, your observations, the observations of others?*
- Do you think the process "works" to distribute rewards fairly relative to performance?
  - *Why/why not?*
  - *Is discretion a critical part of the performance measurement process? Why/Why not?*
  - *Would you prefer that performance measurement was more objective or subjective? Why?*
- How do you think discretion impacts the outcomes of performance measurement?
- Do you think other managers at your level hold similar or different views of the process? Explain differing views
  - *Do you think the process affects some of your peers in ways that it does not impact on you? E.g., impact on decisions, morale*

APPENDIX B
C-Ratios by Firm

		Subjective PMRS Practices			
		Subjective Measures	Subjective Initial Rating	Subjective Final Rating	Subjective Rewards
EquipMan					
Research Question 1: Rationale					
	Incentive Misalignment	0.22	0.30	0.12	0.01
	Employee Sorting	0.00	0.06	0.17	0.21
	Compensation Risk	0.02	0.14	0.01	0.00
Research Question 2: Mitigating Errors and Bias					
	Individual Subordinates	0.00	0.15	0.09	0.02
	Across Subordinates	0.05	0.05	0.39	0.00
FoodMan					
Research Question 1: Rationale					
	Incentive Misalignment	0.11	0.38	0.14	0.05
	Employee Sorting	0.00	0.04	0.11	0.10
	Compensation Risk	0.00	0.13	0.01	0.07
Research Question 2: Mitigating Errors and Bias					
	Individual Subordinates	0.00	0.10	0.26	0.00
	Across Subordinates	0.03	0.03	0.39	0.01
InvestServ					
Research Question 1: Rationale					
	Incentive Misalignment	0.04	0.33	0.00	0.28
	Employee Sorting	0.00	0.08	0.00	0.28
	Compensation Risk	0.00	0.03	0.00	0.11
Research Question 2: Mitigating Errors and Bias					
	Individual Subordinates	0.00	0.17	0.00	0.00
	Across Subordinates	0.00	0.00	0.00	0.00
ProServ					
Research Question 1: Rationale					
	Incentive Misalignment	0.10	0.44	0.08	0.01
	Employee Sorting	0.02	0.04	0.20	0.11
	Compensation Risk	0.00	0.03	0.00	0.00
Research Question 2: Mitigating Errors and Bias					
	Individual Subordinates	0.00	0.24	0.04	0.00
	Across Subordinates	0.02	0.05	0.31	0.00

Notes: C-Ratio = $n_{1,2} / (n_1 + n_2 - n_{1,2})$, where $n_{1,2}$ = co-occurrence frequency of two codes c_1 and c_2, with n_1 and n_2 being their occurrence frequency. PMRS is performance measurement and reward systems.

REFERENCES

Ahn, T. S., I. Hwang, and M. Kim. 2010. The impact of performance measure discriminability on rate incentives. *The Accounting Review* 85 (2): 389-417.

Anderson, S. W., H. Dekker, K. L. Sedatole, and E. Wiersma. 2020. When one size does not fit all: Using ex post subjective ratings to provide parity in risk-adjusted compensation. *Management Accounting Research*: in press.

Arshad, F., E. Cardinaels, and B. Dierynck. 2020. *Facing a calibration committee: The impact on costly information collection and subjective performance evaluation*. Working Paper, University of Manchester, Tilburg University and KU Leuven.

Bailey, W. J., G. Hecht, and K. L. Towry. 2011. Dividing the pie: The influence of managerial discretion extent on bonus pool allocation. *Contemporary Accounting Research* 28 (5): 1562-1584.

Baiman, S., and M. V. Rajan. 1995. The informational advantages of discretionary bonus schemes. *The Accounting Review* 70 (4): 557-579.

Baker, G. P., M. C. Jensen, and K. J. Murphy. 1988. Compensation and incentives: Practice vs. theory. *The Journal of Finance* 43 (3): 593-616.

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

Baker, G. P. 2000. The use of performance measures in incentive contracting. *American Economic Review* 90 (2): 415-420.

Baker, G. 2002. Distortion and risk in optimal incentive contracts. *Journal of Human Resources* 37 (4): 728-751.

Bol, J. 2008. Subjectivity in compensation contracting. *Journal of Accounting Literature* 27: 1-24.

Bol, J. 2011. The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review* 86 (5): 1549-1575.

Bol, J., A. Braga de Aguiar, J. Lill, and A. Coelho. 2018. *Peer-level calibration committees.* Working paper, Tulane University, University of São Paulo, University of Kansas and Universidade Federal do Ceará.

Bol, J., G. Hecht, and S. D. Smith. 2015. Managers' discretionary adjustments: The influence of uncontrollable events and compensation interdependence. *Contemporary Accounting Research* 32 (1): 139-159.

Bol, J. C., and J. Leiby. 2018. Subjectivity in professionals' incentive systems: Differences between promotion-and performance-based assessments. *Contemporary Accounting Research* 35 (1): 31-57.

Bol, J., S. Kramer, and V. S. Maas. 2016. How control system design affects performance evaluation compression: The role of information accuracy and outcome transparency. *Accounting, Organizations and Society* 51: 64-73.

Bommer, W. H., J. L. Johnson, G. A. Rich, P. M. Podsakoff, and S. B. Mackenzie. 1995. On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology* 48 (3): 587-605.

Bushman, R. M., R. J. Indjejikian, and A. Smith. 1996. CEO compensation: the role of individual performance evaluation. *Journal of Accounting and Economics* 21 (2): 161-193.

Campbell, D. 2008. Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research* 46 (2): 297-332.

Cichello, M. S., C. E. Fee, C. J. Hadlock, and R. Sonti. 2009. Promotions, turnover, and performance evaluation: Evidence from the careers of division managers. *The Accounting Review* 84 (4): 1119-1143.

Datar, S., S. C. Kulp, and R. A. Lambert. 2001. Balancing performance measures. *Journal of Accounting Research* 39 (1): 75-92.

Deller C., and Sandino, T. 2020. Effects of a tournament incentive plan incorporating managerial discretion in a geographically dispersed organization. *Management Science* 66 (2): 911-931.

Demeré, B. W., K. L. Sedatole, and A. Woods. 2019. The role of calibration committees in subjective performance evaluation systems. *Management Science* 65 (4): 1562-1585.

Ederhof, M. 2011. Incentive compensation and promotion-based incentives of mid-level managers: Evidence from a multinational corporation. *The Accounting Review* 86 (1): 131-153.

Feltham, G.A., and J. Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69 (3): 429-453.

Ferguson A. 2018. Dollarmites bites: the scandal behind the Commonwealth Bank's junior savings program. https://www.theage.com.au/business/banking-and-finance/dollarmites-bites-the-scandal-behind-the-commonwealth-bank-s-junior-savings-program-20180517-p4zfyr.html; last accessed 2 September 2020.

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

Fritter, E. 2020. The Price of Wells Fargo's Fake Account Scandal Grows by $3 Billion https://www.nytimes.com/2020/02/21/business/wells-fargo-settlement.html; last accessed 2 September 2020.

Gibbs, M. 1995. Incentive compensation in a corporate hierarchy. *Journal of Accounting & Economics* 19: 247-277.

Gibbs, M. 2016. Past, present and future compensation research: Economist perspectives. *Compensation & Benefits Review* 48 (1-2): 3-16.

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

Gillenkirch, R. M., and H. Kreienbaum. 2017. What guides subjective performance evaluation: Incentive alignment or norm enforcement? *Review of Managerial Science* 11 (4): 933-957.

Grabner, I., 2014. Incentive system design in creativity-dependent firms. *The Accounting Review* 89(5): 1729-1750.

Grabner, I., and F. Moers. 2013. Managers' choices of performance measures in promotion decisions: An analysis of alternative job assignments. *Journal of Accounting Research* 51 (5): 1187-1220.

Grabner, I., J. Künneke, and F. Moers. 2020. How calibration committees can mitigate performance evaluation bias: An analysis of implicit incentives. *The Accounting Review* 95 (6): 213–233.

Guest, G., and E. McLellan. 2003. Distinguishing the trees from the forest: Applying cluster analysis to thematic qualitative data. *Field Methods* 15 (2): 186-201.

Hayes, R. M., and S. Schaefer. 2000. Implicit contracts and the explanatory power of top executive compensation for future performance. *The RAND Journal of Economics* 31 (2): 273-293.

Heneman, R. L. 1986. The relationship between supervisory ratings and results-oriented measures of performance: A meta analysis, *Personnel Psychology* 39 (4): 811-826.

Höppe, F., and F. Moers. 2011. The choice of different types of subjectivity in CEO annual bonus contracts. *The Accounting Review* 86 (6): 2023-2046.

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

Kampkotter, P., and D. Sliwka. 2015. The complementary use of experiments and field data to evaluate management practices: The case of subjective performance evaluations. *Journal of Institutional and Theoretical Economics* 172: 354-389.

Kunz, J. 2015. Objectivity and subjectivity in performance evaluation and autonomous motivation: An exploratory study. *Management Accounting Research* 27: 27-46.

Lazear, E. P., and K. L. Shaw. 2007. Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives* 21 (4): 91-114.

Lillis, A. M., and J. Mundy. 2005. Cross-sectional field studies in management accounting research: Closing the gap between surveys and case-studies. *Journal of Management Accounting Research* 17 (1): 119-144.

Liu, X. K., and R. A. Leitch. 2013. Performance effects of setting targets and pay–performance relations before or after operations. *Management Accounting Research* 24 (1): 64-79.

Luft, J., M. D. Shields, and T.F. Thomas. 2016. Additional information in accounting reports: Effects on management decisions and subjective performance evaluations under causal ambiguity. *Contemporary Accounting Research* 33 (2): 526-550.

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

Malina, M. A., and F. H. Selto. 2015. Behavioral-economic nudges and performance measurement models. *Journal of Management Accounting Research* 27 (1): 27-45.

Mills, T. 2018. Police faked 258,000 breath tests in shocking 'breach of trust'. https://www.theage.com.au/national/victoria/police-faked-258-000-breath-tests-in-shocking-breach-of-trust-20180530-p4zii8.html; last accessed 2 September 2020.

Moers, F. 2005. Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting, Organizations and Society* 30: 67-80.

Namey, E., G. Guest, L. Thairu, and L. Johnson. 2008. Data reduction techniques for large qualitative data sets. In *Handbook for Team-Based Qualitative Research*, edited by G. Guest, K. M. MacQueen, 137-161. Lanham Md: Altimira.

Prendergast, C. 2002. The tenuous trade-off between risk and incentives. *Journal of Political Economy* 110 (5): 1071-1102.

Rankin, F. W., and T. L. Sayre. 2011. Responses to risk in tournaments. *Accounting, Organizations and Society* 36: 53-62.

Roberts, J. 2010. Designing incentives in organizations. *Journal of Institutional Economics* 6: 125-132.

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

Woods, A. 2012. Subjective adjustments to objective performance measures: The influence of prior performance. *Accounting, Organizations and Society* 37: 403-425.

Yin, R. K. 2014. *Case Study Research: Design and Methods*, 5th edition. London, Thousand Oaks: Sage Publications.

TABLE 1

Demographic Information

Participant Information	**EquipMan**	**FoodMan**	**InvestServ**	**ProServ**
Supervisors	2	4	2	4
Subordinates	10	8	4	4
Total number of interviewees	12	12	6	8
Average length of interview (mins)	38	47	57	53
Minimum interview length (mins)	23	34	48	40
Maximum interview length (mins)	49	61	71	60
Average tenure at firm (years)	22	7	11	15
Average length in role (years)	2	3	5	3

TABLE 2

Coding Scheme

Code Definition	Examples from Interview Data
Subjective Measures Measures used in performance measurement and reward system (PMRS) are non-verifiable in nature.	*Measures might be things like build relationships with potential clients X, Y and Z.*
Subjective Initial Rating Judgement by supervisor who compiles performance measures and other information to formulate an initial numeric rating of subordinate performance	*I have an employee who I've assessed as a 2, he's meeting expectations, he exceeds a few, but mostly he's meeting…. He struggles to come up to speed as fast as I think he should.*
Subjective Final Rating Judgment by calibration panel consisting of peer and senior managers to determine final assessment of performance.	*She would submit something, but then it's, you know, it's juried, if you will, or it's reviewed by other people who are familiar with my performance and they come to a consensus on my assessment, so, while it is probably more subjective, there is some objectivity to the way it's done.*
Subjective Rewards Judgment involved in determining bonus and other rewards, altering the direct relationship between ratings and rewards.	*I felt the [rating to bonus] formula was quite punitive ….so I had that adjusted manually, like an intervention.*
Incentive Misalignment Subjectivity applied to ensure that individuals are rewarded for adding firm value and not rewarded for behaviors or decisions that produce sub-optimal firm outcomes	*A lot of people get marked down because for one reason or another [their actions] might not align to the values and culture of the organization.*
Employee Sorting Subjectivity applied to rank order individuals to differentiate them for the purpose of bonus, retention, promotion, or development	*I think people really have a tough time understanding they'll work their tail off, they'll do everything and more than what is asked of them, right? But, yet they won't be rated amongst the top performers and we can't tell them how or what they can do to be rated a top performer. I think about the NFL or something, right. You might be getting paid okay, performing well for your team, but does that necessarily mean you're going to play in the pro bowl? Will you be the best of the best and then, and it's kind of the same here.*
Compensation Risk Subjectivity applied to take into account uncontrollable factors impacting on individual performance	*If somebody missed their numbers, maybe not a big miss, but they were lower than what the commitment or the plan was, but, say it was because the markets went down, but they continued to work with customers, continued to grow share, I'm not going to penalize them if the markets go down, right?*
Mitigating Errors and Bias Strategies implemented to mitigate risks of errors and bias arising from subjectivity, including • Information gathering about *individual subordinates'* performance, and • Review of initial performance ratings *across subordinates*	*I ask people outside the team that they worked with what had worked, what wasn't working, what did the person focus on?* *The reason we run calibration processes is because people may not be calibrated at the same level when they rate a person 1, 2 or 3.*

TABLE 3

Code Frequencies by Firm

Code	EquipMan	FoodMan	InvestServ	ProServ	Total
Subjective PMRS Practices					
Subjective Measures	32	28	8	19	**87**
Subjective Initial Rating	96	131	30	105	**362**
Subjective Final Rating	103	77	0	97	**277**
Subjective Rewards	29	84	72	44	**229**
Total Frequencies	**260**	**320**	**110**	**265**	**955**
Rationale					
Incentive Misalignment	52	101	39	72	**264**
Employee Sorting	40	26	23	37	**126**
Compensation Risk	12	17	8	3	**40**
Total Frequencies	**104**	**144**	**70**	**112**	**430**
Mitigating Errors and Bias					
Individual Subordinate	27	36	5	29	97
Across Subordinates	51	37	0	34	122
Total Frequencies	**78**	**73**	**5**	**63**	**219**

See Table 2 for code definitions.
PMRS refers to performance measurement and reward systems.

TABLE 4

Code Co-Occurrences - All Firms

		Subjective PMRS Practices			
		Subjective Measures	Subjective Initial Rating	Subjective Final Rating	Subjective Rewards
Research Question 1: Rationale					
	Incentive Misalignment	38	169	50	34
	Employee Sorting	1	23	53	51
	Compensation Risk	1	34	2	15
Research Question 2: Mitigating Errors and Bias					
	Individual Subordinate	0	62	39	1
	Across Subordinates	7	19	106	1

See Table 2 for code definitions.
Code co-occurrence measures the frequencies with which coded narratives referring to each subjective practice co-occur with coded narratives relating to each rationale, and the incidence of strategies to mitigate errors and bias. Co-occurrences proxy for interactions of concepts underlying the codes.
PMRS refers to performance measurement and reward systems.

TABLE 5

C-Ratios - All Firms

	Subjective PMRS Practices			
	Subjective Measures	Subjective Initial Rating	Subjective Final Rating	Subjective Rewards
Research Question 1: Rationale				
(A) Incentive Misalignment	(A1) 0.12	(A2) 0.37	(A3) 0.10	(A4) 0.07
(B) Employee Sorting	(B1) 0.00	(B2) 0.05	(B3) 0.15	(B4) 0.17
(C) Compensation Risk	(C1) 0.01	(C2) 0.09	(C3) 0.01	(C4) 0.06
Research Question 2: Mitigating Errors and Bias				
(D) Individual Subordinate	(D1) 0.00	(D2) 0.16	(D3) 0.12	(D4) 0.00
(E) Across Subordinates	(E1) 0.03	(E2) 0.04	(E3) 0.36	(E4) 0.00

C-Ratio = $n_{1,2}$ / (n_1 + n_2 - $n_{1,2}$), where $n_{1,2}$ = co-occurrence frequency of two codes c_1 and c_2 (from Table 4), with n_1 and n_2 being their occurrence frequency (from Table 3). For example, the C-ratio between the Incentive Misalignment code and the Subjective Measures code is calculated as 38 / (264 + 87 – 38) = 0.12, where n_1 and n_2 are found in Table 3 and $n_{1,2}$ is found in Table 4.
See Table 4 notes for description of co-occurrence frequency.
PMRS refers to performance measurement and reward systems.