Rendering subjectivity informative: A cross-sectional field study of subjectivity in performance measurement

Anne M. Lillis*
Department of Accounting
University of Melbourne
alillis@unimelb.edu.au

Mary A. Malina
University of Colorado Denver
mary.malina@cudenver.edu

Julia Mundy
Centre for Governance, Risk & Accountability
University of Greenwich
j.mundy@greenwich.ac.uk

DRAFT – PLEASE DO NOT QUOTE OR DISTRIBUTE WITHOUT PERMISSION

***Corresponding author**

"The theory that explains the uses and effects of subjective bonuses is in an early stage of development...many interesting issues arising with subjectivity are not easily studied using traditional datasets. By its nature subjectivity involves concepts that are qualitative and, of course, subjective. It raises behavioral issues such as trust, and conflicts in perceptions between the employee and supervisor. Such issues require a different approach, involving either field research or survey methods." (Gibbs et al, 2004:434).

Introduction

In this paper we report the findings of a cross-sectional field study examining the ways in which firms use subjectivity in performance assessments to enhance the informativeness of performance measurement and reward systems (PMRS). We also explore the ways in which subjectivity is managed within organisations to minimise contracting hazards. Subjectivity in performance measurement, evaluation and incentive determination is attracting considerable interest in the management accounting literature drawing on both economics and behavioral perspectives. On the one hand there is an acknowledged general preference for objectivity in performance measurement (Bol and Leiby, 2011; Moers, 2005; Merchant, 2006). A preference for objectivity is also implicit in the attempt to limit noise in performance measures in the incentive contracting literature (Holmstrom and Milgrom, 1991; Feltham and Xie, 1994; Lambert 2001). However, high quality, objective performance measurement is possibly the unachievable "holy grail" (Baker et al, 1994; Prendergast and Topel, 1996). In the incentive contracting literature subjective measures, ratings and adjustments are posited to serve a range of purposes that counter the deficiencies of objective measurement, but nonetheless subjective judgment is viewed as noisy, inaccurate and corruptible (Baiman and Rajan, 1995; Baker et al,1994; Moers, 2005).

Subjectivity in PMRS is a difficult phenomenon to study empirically as the performance judgments themselves are largely hidden and unobservable. Ittner et al (2003) suggest that prior studies are limited by the use of proxies and note that proxies for noise have little association with measurement choices. Bol (2008) highlights the data limitations that impede the study of subjectivity and concludes that research into subjectivity is still in its infancy. In some studies, the rationale for subjectivity is inferred according to the theories which justify its existence, or by otherwise unexplained bonus awards (eg Ittner et al, 2003). Alternatively, direct measures of subjectivity are

observed in simplified settings (eg Gibbs et al, 2004). Bol (2008) identifies two further limitations of extant empirical studies of subjectivity. First, they do not adequately distinguish the different types of subjectivity (measures, ratings, weighting and adjustment) which have different antecedents and consequences. Second, they tend to study a single type of subjectivity when it is likely that there are a range of subjective mechanisms in use that the same time. This literature does not allow us to disentangle the range of rational economic motivations that may or may not influence the use of subjectivity in different ways in PMRS, the impacts expected, the impact on perceptions of fairness and bias, and the behavioral implications for both supervisor and subordinate managers.

This paper explores the roles and impacts of subjectivity in performance measurement and reward schemes in a field study spanning four large diverse organizations. We draw on the informativeness principle to frame our theoretical contribution. Our findings, drawn from data collected from both senior and subordinate-level managers, challenge the treatment of subjectivity in performance measurement in the incentive contracting literature in several ways

- The incentive contracting literature views subjective performance measurement as a 'residual' solution, adopted when objective performance measures fail. We find subjective performance measurement to be pervasive and purposeful.
- Incentive contracts based on subjectivity are viewed as implicit contracts (Baker et al, 1994) as they lack an evidence base to render them explicit and enforceable. We find incentive contracts to be very explicitly based on subjective criteria, subjective weights and discretionary adjustments.
- While monitoring and observability are imperfect, as envisaged by the incentive contracting literature, they are not absent. Firms adopt a range of mechanisms to harness effort-based observations of agents' effort and action strategies.

We find that firms use subjective measures to enhance congruence by broadening the scope of performance measurement. However, we find that subjective ratings on both subjective and objective measures are critical to building a congruent PMRS. Subjective ratings are used to reduce the

propensity for gaming and self-serving behavior. Supervisors draw on a range of subjective cues to assess both the intensity and direction of agent effort and their action strategies. In particular, calibration committees have an important information-sharing role. Calibration introduces a multi-level RPE mechanism which disciplines lenient supervisors, reduces uncertainty inherent in performance ratings and facilitates a more extensive comparison pool for effective RPE.

Following the recommendations of Bol (2008) we see this field study as a unique opportunity to contribute to the empirical literature on subjectivity by observing the complex construction of PMRSs, probing the rationale for their design and their consequences, viewing objective and subjective PMRS elements holistically, and observing the intended and unintended interactions between elements of the systems. We draw on this rich description to supplement existing incentive contracting theory. In particular, we seek to contribute to the incentive contracting theory literature a deeper understanding of contemporary PMRS design attributes that enhance incentive contract efficiency, effectiveness and risk sharing. We believe this literature has understated the informativeness of subjective performance information, and the range of options available within firms to mitigate contracting hazards.

Our findings are also important for practicing managers. The design of comprehensive, congruent PMRS in practice is challenging. Managers experience tension in balancing the virtues of objective measurement against the risks of misleading signals from the PMRS inducing poor personnel decisions (promotions, task assignments) and inappropriate allocation of rewards due to incompleteness, undetected gaming and self-serving behavior. They also seek to design PMRS that reward subordinates credibly and fairly to enhance motivation and organizational commitment. Our study documents the ways in which subjectivity enters the PMRS, the perceptions of both superior and subordinate managers engaged in performance measurement and the assignment of rewards, and the range of mechanisms firms use to limit the corruptibility of subjective assessments. In addition, we shed light on the hidden subjectivity within apparently formulaic PMRS. Our field interactions indicate that managers are not fully cognizant of the extent of subjectivity in their PMRS. While our participant managers perceive PMRS characterized by objectivity and limited discretion, we observe

many layers of subjectivity, with bureaucratic risk mitigation strategies in place. Participant managers are aware of both the potential risks associated with subjectivity, and the adoption of firm-level risk mitigation, but they do not assess PMRS efficiency. They show limited awareness of residual risks which remain post-mitigation such as centrality bias. Our findings contribute to the intelligent design of comprehensive PMRS in practice.

The next section reviews both the contracting theory literature and empirical literature relating to subjectivity in performance measurement. We then identify the research questions guiding our field study. Subsequent sections outline our research method, findings and discussion. We conclude with the limitations of the study.

Literature Review

Contracting theory

In positioning the role of subjectivity in performance measurement, it is important to revisit the foundation assumptions of the incentive contracting literature. Drawing on information economics, this literature starts from the principle that perfect monitoring is impossible, agent effort is largely unobservable, agents' time horizon and personal goals are not aligned with those of the firm, and the outcomes of decisions are generally not a suitable basis for incentive contracting because agents' "privately taken actions affect the probability distribution of outcomes" (Holmstrom, 1979: 74). The foundation principle of performance measurement in the economics literature is the informativeness criterion, in which "any additional information about the agent's action, however imperfect, can be used to improve the welfare of both the principal and agent" (Holmstrom, 1979: 75). Also important to the principle of informativeness is that agents' risk is low when performance measures provide accurate, timely, and informative signals about their contribution to firm value (Holmstrom, 1979). In the absence of perfect monitoring of agent effort, in situations where decision outcomes are uncertain and influenced by private, hidden actions, firms adopt incentive schemes which incentivize the use of private information for the benefit of the firm (Holmstrom, 1979). The basis of these incentives is a set of information which is directly informative about agents' actions (Holmstrom, 1979).

Lambert (2001) identifies three key attributes of performance measures which impact on their capacity to reduce agency problems in situations where agent actions are not observable.

1. Informativeness of performance measures about agent's actions. This is a function of sensitivity to agent's actions and noise. A corollary of the informativeness requirement is that "very noisy" or "not very sensitive" signals may have a positive value as part of the incentive scheme, but they would be expected to carry very low weights.
2. Congruity – how the measure contributes to constructing an assessment of overall performance, embracing multiple goal-directed actions. This relates to the need to motivate both effort intensity and allocation of effort. For example measures that that are sensitive to the long-term impact of decisions might be used to complement measures that capture short-term action and impacts. These long-term and short-term measures are "incongruent in opposite ways" thus constructing an overall congruent measure (Lambert, 2001; Abernethy, Bouwens and van Lent, 2013). Other examples of incongruity are performance padding, window dressing, lobbying and influencing actions which are non-productive actions induced by performance measurement.
3. Ability to inform the principal not only about agent effort, but also about the agent's action strategy. This attribute of performance measurement reduces information rent that an agent is able to extract based on her information advantage. Agents possess superior information about a parameter that affects the outcome-generating process or the outcome itself. To the extent that ex post performance measures confirm or disconfirm agents' prior disclosures about their action strategies, those disclosures gain credibility. Discretionary bonus schemes allow principals to exploit non-contractible information to compensate agents more effectively and improve risk-sharing (Baiman and Rajan, 1995)

The broad takeaways from this literature are that performance measures will be informative if they are relatively noiseless, sensitive to agent effort, informative about agents' action choices, and induce decisions which are congruent with the goals of the firm. Performance information will vary along each of these dimensions, and additional information is valued if it enhances contracting efficiency and effectiveness.

The incentive contracting literature starts from the premise that, ceteris paribus, objective performance measurement is preferred, but that there are inherent weaknesses in objective measures in some settings (Bol and Leiby, 2011; Moers, 2005; Merchant, 2006). The incentive contracting theory literature assesses a range of incentive contracting solutions that may apply in settings where performance measurement is inherently noisy, incongruent or agent's action strategies are particularly unobservable. The choices include monitoring of agents' actions and design of performance based incentives that rely on performance measures that are informative regarding agent effort (Holmstrom, 1979; Lambert, 2001). Mitigating noise and/or incomplete capture of employee effort in contractual, objective performance measurement potentially benefits both the firm and employee by improving incentive alignment, reducing risk by filtering out uncontrollables and allowing for low-cost recalibration of rewards (Gibbs et al, 2004). This literature explores the limits of these solutions in a range of settings which challenge the design of "second best solutions" (see Lambert, 2001 for a review of this literature). Subjective performance measures may have information content and thus complement objective measures in contracting (Moers, 2005; Bol and Leiby, 2011; Bushman et al, 1996).

Baiman and Rajan (1995) suggest discretionary bonus schemes as a mechanism for principals to exploit non-contractible information in order to compensate agents more effectively and improve risk-sharing. They note particularly the confirmatory role of qualitative performance information. Similarly, Baker, Gibbons and Murphy (1994) conclude that subjective weights can be used in optimal contracts by "backing out" the distortion created by objective measures. Feltham and Xie (1994) suggest that noise in performance measurement induces the use of weak incentives because noise creates risk for the agent which would otherwise need to be compensated. They also suggest that subjectivity can reduce this risk. Prendergast (2002) explores the role of subjectivity and formal investigations in situations where uncertainty would otherwise impose risk on agents that would need to be compensated. Prendergast's solution is premised on the notion that the self-serving bias which is endemic to subjectivity will be exacerbated by incentive contracting because there is "money on the

line"[1]. In an inherently noisy performance setting, agents have little incentive to exert effort. Performance-based incentives (potentially including discretionary components) and formal investigations may increase agent effort.

Relative performance evaluation is popularly perceived as a means of improving contract efficiency through effective risk sharing, by gathering reliable information about 'common' states of nature influencing agent performance (Maher, 1987; Merchant, 1987; Holmstrom and Milgrom, 1991; Feltham and Xie, 1994). Holmstrom's (1979) analysis attaches value to monitoring (increasing observability) and the collection of ex post information regarding agent effort which cannot be inferred from the outcome (possibly in the form of a random audit). Other solutions include the use of low-powered incentives, strategic job design, and rigid bureaucracy (Holmstrom and Milgrom, 1991). Holmstrom and Milgrom (1991) explore contracting in multi-task settings and demonstrate the relevance of optimal job design. They establish the efficiency of task specialization with varying incentive designs associated with different task portfolios - high-powered incentives on measurable tasks and low-powered incentives or fixed wages on unmeasurable tasks. They acknowledge the practical limitations in grouping tasks this way and also discuss the role of rigid bureaucracies and rules in compensating for limitations on the capacity to measure performance.

In sum, the theory literature admits a role for subjectivity or discretion to counter the deficiencies of objective performance measurement. These roles are limited to informative subjective measures complementing objective measure and discretionary ex post adjustments (including re-weighting) to adjust for the impact of uncontrollable factors. The net effect of subjectivity is inconclusive. Used optimally, subjectivity can reduce the overall noise and incongruence in objective performance measurement and provide the principal with recourse to make ex post adjustments based on agents' observed action strategies (Baiman and Rajan, 1995). On the other hand, the economics literature acknowledges that subjective assessments are inherently noisy and corruptible, that they induce rent-seeking behavior (influence costs) and that they provide opportunities for principals to

[1] Interestingly, Prendergast also concludes that in risky environments, supervisor judgements are inherently noisy and the firm has little to lose by offering incentives because supervisors do not have sufficiently valuable information to use it in a self-serving way.

renege on implicit contracts (Baker et al, 1994;). Notably, Prendergast and Topel (1996) reject the conclusions of the early incentive contracting literature, which is "to make compensation less sensitive to supervisor evaluations and to use bureaucratic rules placing "excess" weight on noncorruptible signals" (p976) favoring discretionary organizational processes which draw in relevant performance information. They explore analytically the benefits of favoritism and conclude that a solution of placing weight on noncorruptible measures is dependent on the costs of favoritism.

Empirical Literature on Subjectivity in PMRS

There is now a rich empirical literature identifying the determinants of subjectivity, which are broadly consistent with the incentive contracting literature. Notable in the empirical literature is the move away from principal/agent to supervisor/subordinate relationships where supervisors are not residual claimants. This typical PMRS setting in supervisor/subordinate settings shares many common concerns of principal/agent modelling – issues of noise, completeness, controllability, ex post settling up, and corruptibility of subjective assessments remain salient. The potential for supervisors to renege on compensation promises is reduced, but the supervisor's self-interested preferences provide additional incentives for bias (Bol, 2008; Prendergast, 1999; Moers, 2005).

Researchers focus on the use of subjectivity to counter deficiencies in objective measures including incompleteness and noise (Gibbs et al, 2004; Ittner et al, 2003; Bushman et al, 1996; MacLeod and Parent, 1999; Hoppe and Moers, 2011; Woods, 2012) and susceptibility to gaming (Gibbs et al, 2004). Many recent studies draw on proprietary databases to access real compensation data to infer patterns in the way subjectivity is used, and to examine evidence of systematic bias. Several studies find evidence of anchoring of subjective assessments to objective measures or past performance, consistent with the argument that supervisors are not fully exploiting available information in rendering subjective assessments (Anderson et al, 2013; Bailey et al, 2011; Woods, 2012; Bol and Smith, 2011). Studies of bias identify evidence of both leniency and centrality in subjective assessments (Moers, 2005; Bol, 2011). More recently, researchers have examined bias mitigation strategies including calibration committees (Sedatole and Woods, 2013) and supervisor incentives (Grabner et al, 2014).

While the empirical literature on subjectivity is rich and diverse, we agree with Bol (2008) that our understanding of subjectivity within PMRS is still limited. Despite access to real compensation data, methods of analysis have restricted researchers to examining individual elements of subjectivity such as subjective measures (eg Moers, 2005) or discretionary bonus allocations (eg Gibbs et al, 2003; Bailey et al, 2011) or subjective performance ratings (Grabner et al, 2014; Anderson et al 2013). These finding are not necessarily cumulative as they address different elements of PMRSs; nor do they acknowledge that "many different types of supervisor discretion influence the compensation process" and they may interact (Bol, 2008:16). Further, extant research commonly assesses the impact of subjectivity through observed differences from objective ratings (Moers, 2005; Ittner et al, 2003; Bol and Smith, 2011). By effectively anchoring judgment regarding the quality of subjective assessments to objective data, this approach fails to recognize the potential for subjectivity to be based on superior information. This approach also discounts the possibility that subjective assessments will be informative, even if they are consistent with objective ratings. For example, subjective assessments may have a confirmatory role (Baiman and Rajan, 1995), reducing uncertainty about ratings.

The outcomes of subjectivity are also inconclusive in the behavioral literature. Behaviorally, subjective measures can improve the comprehensiveness of the PMRS and adjustments to either weightings or bonuses can enhance perceptions of fairness, trust and feelings of reciprocity (Bol, 2008). In addition, the behavioral literature recognizes the risk of bias (favoritism, centrality, leniency) and inaccuracy in subjective evaluations and the consequential uncertainty in compensation. The case documented by Ittner et al (2003) provides a powerful example of the use of discretion which is inconsistent with the informativeness hypothesis, and behaviorally destructive. They note lack of transparency, perceptions of favoritism and an excessive focus on financial performance. Moers (2005) finds evidence of leniency in subjective assessments.

It is notable that the costs of being wrong in performance assessments can be quite significant. 'Errors' in performance ratings (which are largely unidentifiable and can only be inferred from large-scale empirical data) are costly both in compensation costs and in task assignment, screening, selection and promotion (Prendergast, 1999; Moers, 2005).

Research questions guiding our field study

Our review of the incentive contracting and empirical literatures on subjectivity lead us to identify a paradox. In practice we see the common use of performance-based incentives in a broad range of settings and clearly in situations of multi-task complex roles, private information and hidden action (eg profit center managers). Furthermore, there is evidence that subjective assessments play a key role in these contracts despite their potential corruptibility. In this study we examine subjectivity in PMRS across four field sites. We draw on the incentive contracting literature to frame our theoretical contribution. In particular, we are not aware of any empirical evidence addressing directly the motivations of supervising managers, perceptions of subordinates and behavioral impacts of the exercise of discretion in performance measurement and the allocation of rewards.

Our overarching research question is to investigate how subjectivity is rendered informative in performance measurement and reward systems. Our objective is to document and interpret contemporary incentive contracting solutions in inherently noisy, multi-task settings. We seek to provide both a structure to the use of subjectivity in practice as well as richness to the empirical experience of subjectivity. In these settings, we document the relative reliance on objective (measurable) criteria and subjective judgments in contemporary performance measurement and reward practices, and analyze the rationales and impacts of these practices. We address three main research questions. Under each research question, we interpret our field study findings against the principles of the incentive contracting literature:

RQ1: How and why do supervisors deploy subjectivity in PMRS?

a. Congruity - completeness, management of interdependencies and decision time frames, and countering detected gaming
b. Informativeness - estimating outcomes, assessing effort, and facilitating ex post settling up
c. Encouraging truthful information exchange - reducing information rents, informing about agents' action strategies

RQ2: Are the risks of subjectivity acknowledged, and how are they mitigated?

a. Mitigation strategies - mechanisms to manage the inherent noise and corruptibility of subjective assessments

RQ3: To what extent is residual bias evident after risk mitigation?

a. Leniency
b. Favoritism
c. Centrality (rating compression)

Research Method

We adopt a field study research design which allows us to study subjectivity in PMRS holistically, in context, and with both supervisor and subordinate informants. We study these phenomena across our four field sites in order to increase the variation in PMRS and to identify recurring patterns in different settings (Lillis and Mundy, 2005). The four organizations we study are large and diverse, each from a different industry: equipment manufacturing (EquipMan), food manufacturing (FoodMan), investment services (InvestServ), and professional services (ProServ). In our selection of firms, we require that some level of subjective evaluation is used in the determination of performance-based rewards. This excludes, for example, sites in which bonuses are based on sales or allocated solely on the basis of job title or position.

We conducted in-depth semi-structured interviews with two consecutive levels of managers within each organisation. We first interviewed managers at a business unit level and then asked each of these supervisors to recommend us to several direct reports who were themselves supervisors for the next level below. This enabled us to analyse our data both within and across managers at different hierarchical levels, further increasing the potential for variation in our phenomena of interest. In one of the firms we collected data from a single division, while in a further two, data were drawn from two and three divisions, and in the fourth from three teams (departments) that work closely together but undertake different activities. The purpose of the interviews was to gain insights into: those elements of the incentive contract that introduce subjectivity into the contract, those elements over which supervisors have additional discretion, how supervisors obtain and use subjective information,

how they construct and use subjective performance measures, the rationale underpinning the various choices they make, perceptions of bias, and the behavioural implications of the incentive contracting process for both supervisor and subordinate. A summarised copy of the interview guide is shown in Appendix A.

The interviews were audio-recorded and then transcribed verbatim and in full. In total, we conducted 43 interviews, ranging from 23 minutes to 61 minutes. We obtained ethics approval from our respective institutions and informed consent from each of the participants. All names and other identifiers have been removed in order to preserve the anonymity and confidentiality of the firms and their employees. Demographic data for all participants are shown in Appendix B.

We used a qualitative data software package (NVIVO) to code the data according to our theoretical framework. All three authors coded the interview data. The coding process consisted of associating text with broad themes of interest. The use of high-level thematic codes increased inter-coder reliability[2]. Exceptions were resolved through further discussion between researchers. We then reorganized the data according to our themes of interest before systematically analyzing the transcripts (Miles and Huberman, 1994). We created summaries of key themes for each interviewee that included quotations from the interview transcripts. These summaries form the basis of the findings and subsequent analysis presented in this paper.

Findings and Analysis

The four firms in our study differ in the extent to which they adopt a formulaic approach to the determination of performance-based bonuses. At one extreme, InvestServ has separate bonus pools for each team, no formula for the allocation of bonuses, and a very opaque process. At the other extreme, FoodMan has a completely formulaic process with three components, each with a multiplier attached and a resultant numeric score which maps directly to a bonus allocation. The other two firms lie between these extremes. ProServ is somewhat similar to FoodMan, but with added discretion in adjusting the bonus allocation. EquipMan has no formula for performance ratings, a formula for

[2] High-level thematic codes include themes such as 'calibration', 'bias', 'uncontrollables'. Using very high-level coding categories, the scope for disagreement among coders is minimal, but the trade-off is the time spent on data analysis given the larger volume of data recalled under each theme.

allocation of bonuses and a transparent process. Importantly for the purposes of our study, these formulaic bonus schemes mask deep layers of subjectivity in all four firms. Apart from the use of subjective measures, three of the four firms override objective measures with subjective judgments regarding 'how' the objective results have been achieved (InvestServ is the exception as it doesn't have objective measures). All of the firms also have additional layers of calibration, where the subjective judgments of other stakeholders can lead to the adjustment of performance ratings of individual subordinates. Thus these field sites provide a rich setting in which to study subjectivity in PMRS. Appendix C provides a brief but more extensive outline of the PMRS in each of the four case firms. In the following subsections, our field study observations are organized and analyzed around the themes of our research questions.

RQ1a: How do supervisors deploy subjectivity in PMRS to enhance congruity?

Theory stresses the importance of congruity of a performance measure, or how it contributes to constructing an assessment of overall performance embracing multiple goal-directed actions. Incongruity occurs when a) performance measures incompletely capture valued task attributes, b) the subordinate places more importance on individual or local outcomes at the expense of the firm, or c) the subordinate takes actions to increase a reported measure without increasing the real outcome through gaming. We find two distinct forms of subjectivity are used to enhance congruity. Subjective measures are used to draw in important role expectations that are difficult to quantify, and subjective ratings on relatively objective performance criteria are adopted to adjust for 'how' subordinates have achieved objective outcomes. We find that subjective information allows supervisors to a) make an evaluation of performance covering all performance dimensions, b) determine how the outcomes were achieved, not just what was achieved, and c) manage and detect gaming.

In all four firms, subjective measures are used to enhance the completeness of subordinate performance measurement across all task aspects including leadership, achievement of project milestones, and other hard-to-measure support functions such as human resources, corporate affairs, and research and development. Our evidence suggests that in three of the firms, managers strive to quantify targets on subjective measures as far as possible, but recognize that quantification will be

incomplete. They supplement short-term objective measures with subjective measures that capture local versus company-wide goals, more qualitative role dimensions, and long-term versus short-term performance.

> *Some of mine might be ... support the role out of integrated business planning where necessary in Australia. It might be ensure development plans and KPOs are in place for all your team by a certain date. It might also be a safety target, ... successfully lead the annual planning process, ...Was it done correctly? Did you miss any information, errors, that kind of thing?* FoodMan

> *You can help individual employees understand the importance of setting good work plan objectives that are measurable, can be evaluated. ... It's important, you know, for them to be involved in the community, for them to be helping lead the team, drive* [EquipMan] *overall initiatives, that type of thing.* EquipMan

> *There's an awful lot you can do to create new markets and drive sensible business development activity that won't immediately yield value. Now, I think in a way we probably reward those that are more short-termist than long-termist and driving activity, or certainly those with mature business opportunities rather than more innovative business opportunities. So, do I implicitly weight them? I'm sure I do, I'm sure everybody does.* ProServ

In each of the case firms, when evaluating overall subordinate performance, including performance on objective criteria, the supervisor evaluates not only *what* was achieved, but *how* it was achieved with respect to the company's core values. It is the subjective assessment of the 'how' that stimulates managers to draw from a range of sources to assess whether objective outcomes are a reliable measure of contribution to firm-wide outcomes. The fact that the 'how' will be assessed and will influence the allocation of bonuses is explicitly contracted in three of the firms (InvestServ) is the exception). However, the weight placed on the outcome (what) versus effort (how) in making an overall assessment of subordinate performance is not based on a formula, thus giving supervisors discretion over both the rating and weighting. While the more-objective *what* and the more-subjective *how* need to be balanced by the subordinate, the allocation of effort between the two is not explicitly stated in any of the PMRS and may vary by subunit, by level of management and by individual 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

> *They measure the .. 'what' - tick the box, get the job done, that's fine but you know is it sustainable? Have you done it in accordance with the company's values and behaviors? Have you left a legacy? Have you done it in a way that doesn't breach the code of conduct? So that's more around the 'how'. .. I have seen in organizations that high performers particularly in sales and trading environments can have shocking personal behaviors but because they're a*

> *revenue-generator they're rewarded and kept in the organization 'cause the organization doesn't want to lose the revenue so they're really focused on the 'what'. Despite the fact they talk about the 'how' they just don't worry about the how. I think this organization .. the how is important.* FoodMan

> *What's not consistent is maybe the amount of weight we put on leadership versus other work plan items. I think that probably varies from leader to leader, depending on how much emphasis they put on the how and not the what.* EquipMan

In each case firm, we found evidence of subjectivity being used to mitigate and adjust for detected gaming behaviors that might occur with purely objective performance measurement. In particular, our evidence suggests that managers guard against the potential to achieve performance targets through self-serving actions, ignoring firm-wide impacts. The calibration process captures broader information about subordinates' action strategies from other stakeholders within the firm in order to mitigate the risk of rewarding self-serving actions.

> *You just pick a variety of numbers which you think are fair using those inputs. No formula whatsoever because all the nine people are very, very clever people. If you tell them what the formula is there are enough variables in there for them to game with comfortably…It's complete anarchy within our team in that you pick which stocks you want to analyze, so if you're very clever you would just pick the stocks that you think are most likely to be cheap and you would blow everyone out of the way.* InvestServ

> *[Calibration] is where you find out if people manage upwards whereas other perceive them as self-serving.* FoodMan

> *You begin to challenge individuals that you'd had interactions with. I've got a position on this guy, I worked with him on [x] project, I found him to be very brusque and abrasive in his style and not terribly collaborative yet you rated him 'exceeds'. Talk to me about how you got him there because I don't see the 'how'. Or ... sometimes it'll be the reverse ... you've been hard on this guy and yet when I asked him to give me some new reports he just nailed it, they were great, he had insights, it was timely, it really changed how I approached that thing. You know so you get it in both directions. It's a very passionate day, I'll tell you.* FoodMan

> *Going for pure performance metrics only, 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. Behavior-based, especially.* EquipMan

There is also some limited evidence of other information sources accessed to detect gaming (eg internal audits at FoodMan.

RQ1b: How do supervisors deploy subjectivity in PMRS to enhance informativeness?

According to theory, when an outcome is not observable, there is potentially a role for information that a) helps estimate the outcome, b) adjusts for uncontrollables and, c) removes the effects of common random factors (to isolate the intensity of subordinate effort). Managers in all four of the firms in our study supplement their own subjective evaluations about subordinate performance with information from colleagues and clients who have worked directly with the subordinate. Our evidence indicates that feedback is gathered formally and informally, in written or verbal format, and directly by questioning or indirectly by observing subordinate interactions with colleagues. The calibration process is an important information source. Supervisors also access other internal and external information sources to obtain subjective feedback relating to subordinate actions not reflected in performance measures, which is ultimately then embedded in subjective ratings.

> *I might be aware of feedback from clients, more likely colleagues… particularly if I feel like I'm marking someone down I tend to ask a few other people… because some of them are also marked on their relationships with others…an analyst does a lot of writing which a fund manager then reads and a lot of their job is to feed back to the analysts to give them effectively cuddles, so are you getting the good cuddles?* InvestServ

> *There's actually a mechanism in the system where you are required to request feedback from stakeholders as well as direct reports. You get much less constructive feedback written down and submitted through the system, and rarely do you get bad feedback. For me, that part of the feedback, the formal feedback, has not been as useful. I will certainly get informal feedback meaning I may ask the person specifically for feedback if it's someone I think has a good view into someone's performance.* EquipMan

> *Feedback from clients is given probably the most weight, but also internal feedback from direct reports and up and down and across the firm.* ProServ

Where subordinates are held accountable for achieving specific goals and performance is potentially influenced by circumstances outside the subordinate's control (whether favorable or adverse), supervisors seek to evaluate the extent to which the subordinate could have predicted or mitigated the outcome and whether she took reasonable steps to manage or exploit the situation. Adjustments are made to the evaluation or rating given by the supervisor and are particularly evident in those areas in which the subordinate goals include hard targets. The evidence gathered from our firms suggests that, while supervisors consider many factors that may have impacted the subordinate's performance, the process is largely implicit and does not seek to quantify the extent of the adjustment made by the supervisor.

I look at their teams and I think, okay, how much have you cost me this year in an error? Was it something that you weren't in control of and it would have happened anyway, or actually you should have really thought about that and you should have buttoned that down before the error recurred? How proactive were you in closing that down and putting in a control? InvestServ

Street justice. If someone doesn't get 100% of the objective typically speaking it's difficult to say they've achieved an objective unless there are really extenuating circumstances. Conversely someone might get 101%… and they have really pulled out all the stops to get there. So you might say it's basically just around 100% but given all the circumstances it was a very, very good effort so you might call it an exceeded so there's room for discretion there. FoodMan

They're not going to hit that revenue target so then the issue is how does that person show leadership in doing what they did to do when the market started going down? Did they see it? Were they proactive? Did they start cutting costs, stop hiring, get the team together with a plan to close the gap at the bottom-line level? There's all those kinds of things and I think those are some of the discretionary things that happen. EquipMan

The performance of a subordinate and his peers will often be affected by common random factors. All four case companies utilize relative performance evaluation (RPE) to filter out common noise. In three firms this process is explicit, consisting of two layers of RPE within the PMRS, both of which are subjectively determined. The first layer can be described as RPE within supervisors. The supervisor determines an overall performance rating for each of his direct reports relative to each other. The second layer of RPE can be described as RPE across supervisors. In a calibration session, the supervisor and her peers review the overall performance evaluations of all subordinates within the same compensation level. This level of RPE establishes expectations for common rating approaches and distributions across supervisors, as well as broadening RPE comparisons among subordinates. This is particularly important where the supervisor is not required to assign ratings according to a pre-specified distribution.

You can play with it as much as you like, but the moderation will catch you up. So you can have a Partner or a Director who's appraising a group of people and says, 'all of the people I appraise are brilliant, they're all 1s'…You arrive at the moderation meeting, and you immediately see this as an outlier, the guy's behaving like a trade unionist, not a manager…It's not impossible that every single one of the people that guy has rated as a 1 will end up as a 2, or lower. ProServ

If you're one of those folks that, you know, everybody you've 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 dialogue back to things that are specific to what they employee has done. EquipMan

You do go through a process of trying to calibrate to make sure that you don't have one manager who's got all their folks sitting over here and where a manager has a dispersion that isn't normal for the business they start to get challenged on certain individuals and they have to explain. FoodMan

In one firm (InvestServ), the process of RPE is entirely implicit, manifesting only through the allocation of bonuses by each supervisor. The supervisors do not know the bonus pools of other teams, so calibration takes place informally with their respective line managers. By providing each supervisor with their own bonus pool (determined at the end of the performance period), senior management has essentially undertaken an across-team calibration without the direct involvement of the managers, and added an additional layer of subjective adjustment in this firm.

There's sort of a sanity check, but I've got the final say so I go and speak to my boss who says 'that's a bit high, a bit low, other people are getting paid this'. I'm made aware of it but I don't have to act on it. InvestServ

This is one of two examples in our data of subjective adjustment applied directly to the bonus allocation. In the other cases, the adjustments are made to performance ratings, which then feed formulaically into the bonus determination. The second example is ProServ:

Maybe someone has genuinely performed exceptionally well, or they were only here for part of the year, and the bonus pro rata ends up looking a bit low for what they've achieved, so they need a little bit more. Because the bonus is generally set as a percentage of salary, you can be affected by someone who just happens, for historical reasons, to be paid a lot more, who's getting a much bigger bonus than someone else, who for, because they're a little bit younger, or they didn't come in from outside or whatever, the percentage of the bonus, because it's a percentage, is unfair. So we might take a bit off them and add a bit to them, so they end up on the same bonus. …[T]here are sort of things like that, where you just need to tweak it. ProServ

RQ1c: How do supervisors deploy subjectivity in PMRS to encourage truthful information exchange?

According to theory, principals can potentially reduce information rents by having agents communicate information signals to supervisors. This aspect of incentive contracting is potentially less salient in supervisor/subordinate relationships than it is in principal/agent models, as there is commonly limited information asymmetry between supervisors and subordinates given day-to-day interactions. For example at FoodMan information asymmetry is low. The risk of subordinates extracting information rents was not raised by supervisors. The company is characterized by a culture of communication. Targets are negotiated, performance measurement and reward practices are transparent and trust was generally considered to be high. The combination of close

superior/subordinate working relationships and calibration to capture unobserved behaviors limits the potential for information rents. Nonetheless at EquipMan and InvestServ we identify a variety of roles for subjectivity in the reduction of information rents. At EquipMan, the supervisor has discretion to re-contract some pieces of the work plan if circumstances change during the evaluation period. This encourages truth telling because re-contracting will require the subordinate to change her strategy for achieving the desired outcome.

> *If something's not working right, I might re-contract it with her and say, okay, I thought I was going to be at this point, but this has happened and this has happened. So we either agree to re-contract it or we won't. If we re-contract it and change the measure a little bit, you can continue to meet expectations, but if it's not re-contractible, then you're falling below what your commitment is, - you fall into the category of not meeting expectations.* EquipMan

At InvestServ, there is a high potential for information rents among the analysts in the risk and investment teams because they have a more detailed level of knowledge about the performance of projects and investments than do their respective supervisors. The design of the incentive contracts induces truth-telling. For example, while the analysts choose which stocks to analyze, their performance is not measured against the performance of these stocks. Rather, the supervisors subjectively evaluate the quality of the company notes they write. This disincentivizes the analysts from withholding information about the predicted performance of particular companies. It also reduces their incentive to mislead colleagues over which stocks they should analyze. Furthermore, as the bonus pot depends heavily on the collective performance of the team, analysts have an incentive to report 'truthfully' their views or doubts on the company notes written by their peers, a fundamental aspect of the team's processes. Notably, one senior analyst is told which stocks to analyze (because the supervisor requires some control over the stocks that are analyzed in each period). As this analyst has no reputational interest in the performance of these stocks, he therefore potentially provides a benchmark for truth-telling by the other analysts.

RQ2: Are the risks of subjectivity acknowledged, and how are they mitigated? (noisiness and corruptibility of subjective assessments)?

The literature has shown that subjectivity in performance measurement might result in biases such as favoritism, centrality, and leniency as well as inaccuracy in subjective evaluations. The extent to which our case study firms' PMRSs allow or require supervisors to incorporate subjective ratings

into subordinate assessments, even on relatively objective criteria, is evident in the prior sections. We find that our case firms acknowledge the risks associated with subjective ratings and have implemented numerous processes to combat them such as evidence-based assessments, calibration sessions and formulating expectations consistent with Bell-curve distribution of ratings.

> *There are continuous layers of scrutiny and check and moderation and balance.* ProServ

All four companies provide guidelines to reduce "free" judgment and, where included as part of a formal process, require supervisors to provide evidence supporting ratings given to subordinates.

> *So there is a guide when you look at core values that kind of talks about what does exceeding and not exceeding expectations look like? Again, it's not numbered but it's a list of statements that say what it looks like. If you think your employee is exceeding you can look at the list of things that say exceeding to see if they are doing these things or not. So it gives you a guide.* EquipMan

> *[I]n order to rate somebody, our partners are always having to…justify and have evidence. Because of the nature of our organisation, there's a certain integrity, you know, 'I've applied this rating and this is why', and so very rarely will we get someone who says, 'well, I think this person is this rating, and that's just because I've decided'.* ProServ

In three firms, subordinates and supervisors view the calibration process as a way to discipline supervisors against leniency and favoritism.

> *The same thing that makes it difficult for managers to give courageous and constructive feedback to individuals makes it hard for some managers to not write their direct reports as superstars. And you need a process of some kind,... let's get around a table and we do it* [calibration] *in two ways, we do it within each function and then the senior management team does it across functions*. FoodMan

> *Some people try and game the process a bit by saying, 'well, really, if I'm honest with myself, I would say that these people are threes, but I'm going to go and see if I can get away with getting them twos, because I've worked with these people and I like them and I don't want to have a difficult conversation with them'. But that is challenged very robustly in the moderation meeting, and I have seen some examples of people coming up with what are almost ludicrous suggestions, and they are just completely shot down and look completely stupid.* ProServ

Both subordinates and supervisors expressed that if there is consensus on a rating given to an employee, then it is as close as possible to being objective and accurate.

> *The ratings are subject to another level of scrutiny and then re-evaluated but at the same time it seems to almost reduce some of the subjectivity as well by controlling it.* FoodMan

> *She would submit something, but then you know, it's juried, if you will. It's reviewed by other people who are familiar with my performance and they will come to a consensus on my*

> *assessment. So, while it is probably more subjective, there is some objectivity to the way it's done. They look at it to insure accuracy and fairness and so forth.* EquipMan

In many cases, calibration also implicitly confirmed initial ratings. While both supervisors and subordinates were keen to discuss the circumstances in which calibration would alter ratings, it was clear that this was not the norm, and for the majority of cases before the calibration committee the calibration process also performed a critical confirmatory function.

> *Basically you ask for that feedback against the seven values and say is that person you know on par, below par, above par? ... Because then that gives us feedback across a range of stakeholders that either confirms or balances the view of the manager in terms of the final rating and that feedback is then shared with the individual.* FoodMan

Interestingly, despite the apparent high levels of subjectivity in the processes of two firms (subjective measures, ratings, weights, and occasional subjective ex post adjustments) several managers at FoodMan commented that "[FoodMan] *is at the low end of the discretion scale to begin with*" or "*there is discretion, but it is within small bounds*" or "*there is a lot of monitoring around the process to try to eliminate that discretion*". Similarly, at ProServ one supervisor noted:

> *We sort of have discretion and we don't have discretion…I have discretion in that I can make judgements about individuals but then the judgements I make are subject to scrutiny.* ProServ

These views seem to reflect a great deal of faith in both the transparency of the process and consensus-driven objectivity through the calibration process.

> *This has been the most objective I've ever seen. In the past we didn't have this breakdown so our performance rating would be typically based on a manager's assessment, maybe they did the calibration between them but it was a very closed doors, now it's a lot more calibrated across the business. So for example, in a marketing function they're going to probably rate it very similarly because what HR is doing is standardizing the calibration across the business as well. So in terms of quantitative, I think this is probably as close as we're going to get to have a quantitative measure.* FoodMan

The literature highlights the risk of a leniency bias and a centrality bias when using subjectivity in assessing subordinate performance. A leniency bias occurs when there is inflation, *i.e.*, the average rating is higher than actual average competency. A centrality bias occurs when there is compression, *i.e.*, the distribution of the ratings has a lower variance than the underlying distribution of actual competence. The risk of these biases was recognized by all four firms and two of the firms currently use 'advisory' distributions to reduce bias, while stopping short of demanding a fixed distribution.

> *It is better to stick to the curve, because then you can more clearly differentiate, and the point is that, if you don't differentiate, what's the point of appraisals, you might as well not bother. You might as well say, there's a 130 [employees], the pot's this, I'll just divide one by the other and there you go.* ProServ

> *Generally in any organization, people will get a green and that could be 'achieved objectives'. And then there's blue which is 'outperformed', and then there's red which 'needs improvement' or 'below expectations'. And generally with these things, you want 80 or 85% of people in the green, and then obviously allocating out your 7.5% into the blue and your 7.5% into the red as well... let's say everyone had met their targets, ... people go, alright, who are we going to put in the red. So, well who's done a job that's not 100% there? who can we mark down. So [calibration] is just one of those processes.* FoodMan

InvestServ's processes are largely opaque and heavily reliant on the sole evaluation of the supervisor. A high level of trust mitigates the potential adverse effects caused by the lack of transparency.

> *I have a total faith in the subjectiveness…I think [manager] is as objective as anyone could be, in that he has a load of different people on the team and he's looking at the work that comes out…I trust [him] to pay what he thinks people are worth, whether they're quiet as a mouse or…* InvestServ

There is mixed evidence, even within one case firm, as to whether calibration extracts upward adjustments as well as downward adjustments:

> *I like to think they challenge up as well as down but I suspect it's challenging down.* FoodMan

> *Managers probably won't rate their people blue [outperformed expectations] unless they've done a really exceptional job. But I think that if you're trying to fit people into a bell curve, you'd probably look at it and go, oh right, well we need some people to go into the blue area. Of all these people here, who's done an exceptional job this year? Right, move them up.* FoodMan

EquipMan no longer uses a forced (Bell curve) distribution within RPE. Managers felt that forcing a distribution distracted them from the true task at hand of evaluating and coaching people. However, informally, a natural distribution with a wide 'meets expectations' band is preferred.

> *When I came to the company 14 years ago, it was forced ratings. You just kind of knew you had to hit a distribution. You just focused on who falls in what bucket instead of individual coaching and feedback and therefore developing people. So I think we've moved a long way on that. And to me, I think that's the goal is you want your performance measurement system to help both your organization and the individual.* EquipMan

RQ3: What residual biases are evident after risk mitigation?

We find that the subjective elements of the firms PMRS work together to create a coherent, informative and 'relatively objective' system. Within each case company, we find evidence of subjective elements across the entire PMRS, including subjective measures, subjective weightings of measures, subjective ratings on objective measures and subjective distribution of rewards. The chain begins by placing careful attention on making subjective measures as objective as possible to aid in measurement and consistency. The subjectivity in the overall performance evaluation appears to be acceptable, given that those evaluations are subject to the calibration process. Firms rely heavily on the calibration process as the high-level 'sanity check' of the entire subjective 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

We find that, in fact, failure modes do slip in. There are several residual, unmitigated effects of subjectivity in our case firms. Einhorn (1974) describes consensus as a necessary, but not sufficient, condition for accuracy among a group of expert decision makers. Our case firms view consensus as a surrogate for accuracy and place heavy reliance on results of the calibration sessions. While the correlation between the two may be high, there may be a false sense of security that the results of the calibration sessions are accurate and objective.

Our firms noted several weaknesses in the calibration process that fail to mitigate the risk of using subjectivity in performance evaluation. These weaknesses include variation in the ability or propensity of supervisors to skillfully and freely debate in calibration sessions, lack of evidence-based information about lesser-known subordinates, and the potential for inappropriate weighting of information brought to calibration:

> *I could push back in the room and I'm comfortable in that and the environment is comfortable, but, I'm like in China or India with the leader in the room, and if the leader were the one saying that, will people really push back? Right? So I don't know. So I think cultural implications, when it comes to calibration could be interesting.* EquipMan

> *My manager in S* [location] *doesn't see me day-to-day. I will have a phone conversation with him maybe once a month and X sees me pretty much at least once a week, so X has a much better judgment of what I am like on the ground where Z doesn't. However, Z needs to put a word for me in the calibration session… From his limited knowledge about my performance, he'll make a judgment as to how I'm tracking.* FoodMan

There are also situations where it can be overused where you might have someone who has quite a strong opinion who then can voice that opinion and … people might then over-index in that space. So it's sort of power but it's got to be used with some sort of caution as well. FoodMan

Some people are more powerful than others and hence even in a moderation their voice carries more weight than others…And sometimes you can't measure the support that somebody gets from another partner that they will value a lot…It can often be an emotional thing…It's very difficult to take that bias out of the system. ProServ

As stated earlier in the paper, firms mitigate centrality bias by either explicitly or implicitly requiring a bell-curve distribution of ratings. However, the calibration meetings focus mostly on high and low performers which may, in fact, increase the amount of averaging.

"We look at every person who is a high performer and the manager talks about their work plan and core values and leadership together. We're questioning it in a good way, we talk about it and agree and sometimes we move people down. We then move to the bottom group and have the same kind of discussion. Occasionally we move somebody up, it's usually less likely." EquipMan

The calibration sessions I've been in typically people talk about the exceptions rather than everyone But when … when people were rated average/meet expectations, we don't typically discuss so what I'm thinking is maybe we're actually making everyone average. FoodMan

In one firm, a desire to minimize confrontation between supervisors and subordinates reduces the ability of the supervisor to provide a 'true' rating and leads to leniency bias and the potential for inefficient contracting.

The problem we have is actually the distribution curve, where 60% of people are in the 2 bracket [high performing], *and that causes problems, ...people absolutely hate being in the 3 bracket. …. In the past, we've had a 1 to 5, where the majority of people are told, you're in the 3 category…[A]t the moment we're saying the vast majority of our people are high performers, it kind of doesn't help you manage expectations.* ProServ

We also find evidence of favoritism towards specific subordinates that takes place after the supervisor's evaluation but before the final distribution of rewards. Adjustments are made by the supervisor's superior to allocate additional compensation to the subordinate beyond that initially determined by the supervisor's evaluation. Those who identify favoritism regard it as distorting the outcomes, implying that it results in inefficient contracting.

> *I was trying to work the numbers in such a way that would ensure that the people that did very well got a substantially better bonus and the one that did not so well got a lower bonus. [My Superior] wanted to keep them on the same as last year and I said 'but how else are we going to balance the books?' [My Superior] said to me, '[my Superior's boss] thinks that he's a very valuable guy to the firm and he wants to keep him sweet', and I thought, 'hmm, okay, that's quite interesting to know.'* InvestServ

> *There's always a suspicion that over the long term…there are relationships that build up between partners that, you know, maybe you've gone through the firm together and are friends. One could speculate that there's some bias there and that some people are protected and that you do not get the same level of objectivity… It's generally more senior partners where people think 'why are they still there?' But that partner will quite often have a lot of big networks in the firm, a lot of powerful people who say, 'I remember 15 years ago doing whatever with him and he's a great guy'.* ProServ

Managers in FoodMan talked about the importance of trust in their manager in terms of the way their performance is represented in calibration sessions, but also implied a level of influencing activity:

> *You make sure you equip your manager with enough information about yourself, bad or good, to paint the right picture for you, so you trust that he will do that and at least the best case has been put forward. I've also seen colleagues where there wasn't such trust in their relationship and what I've seen ..the manager would not … be very vocal about their cases in a peer forum. On the other hand, I've seen where … perhaps too much trust or even a personal relationship outside of work can actually lead to a over-judgment (sic) of someone or they're very super-protective of that person.* FoodMan

We also find evidence of systematic bias arising from the desire to limit downside risk associated with incorrect personnel decisions, both for the supervisor and subordinate.

> *Within our organization I would say there's more of a tendency to downrate than overrate because the consequences of that in terms of [people being] promoted … 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 most of the time in not overrating individuals and setting them up for failure.* FoodMan

Discussion

There are several themes which emerge from our findings. First, subjectivity is pervasive in the PMRS of the case firms we study. All use subjective performance measures, but more importantly, all make extensive use of subjective ratings, even when performance criteria are largely objective. All four firms place considerable importance on rating 'how' objective outcomes are achieved, and generally attach greater weight on subjectively assessing the 'how' than they place on objective outcomes. The subjective overlay on the rating of objective criteria effectively provides

supervisors an implicit capacity to not only adjust performance ratings on specific criteria, but also to re-weight criteria in the overall assessment as the weights on the 'what' and 'how' were perceived as largely discretionary. In contrast the level of subjectivity in bonus allocations was minimal in three out of the four firms. In these three firms, the formulaic approach to bonus determination led managers to describe their PMRS as formulaic rather than discretionary[3], without recognition of the layers of subjectivity impacting on the elements in the formula.

While subjectivity in these firms is more pervasive than residual, which challenges the view of subjectivity in the incentive contracting literature, the rationale for use of subjectivity is largely consistent with theory. Subjective measures render PMRS more complete in capturing complex role attributes, and subjective ratings are essential to ensure that congruous decisions are rewarded, and self-serving behaviors are not.

The extent of subjectivity in these PMRS would, in theory, be consistent with high compensation risk in these firms. The inability of subordinates to predict compensation outcomes due to the influence of subjective judgements would typically lead them to demand supplementary compensation. In our case firms, there is no mention of significant compensation risk or premium compensation by either subordinates or supervisors. Conversely, managers at both levels tend to speak favorably of the PMRS, drawing on descriptors such as fairness and transparency. There are, of course, some exceptions as noted above in relation to residual risks post-mitigation. However these are the exception from 43 interviews. This counter-intuitive finding is consistent with these firms adopting effective risk-mitigating strategies. We find that the extensive reliance on calibration committees in three of the firms serves multiple purposes, all of which are consistent with risk mitigation. These committees discipline supervisors against favoritism and leniency. They provide a multi-stakeholder evidence base to validate subjective ratings, and they provide a critically informative forum in which incongruent actions become identifiable. Other mechanisms such as

[3] For example, at FoodMan the bonus scheme had three elements – firm performance (three financial criteria), business unit performance and individual performance). Each element was included in the formula as a numeric amount, each element had a specific multiplier attached, and the result was the determination of a bonus which was final. There was no subjectivity applied to the determination of bonus payments.

rating guidance and implicit distributions are also used, but calibration is the dominant risk mitigation strategy. We do, however, observe that while calibration appears effective in reducing compensation risk by managing risks of favoritism and leniency it does seem likely to increase centrality bias in the absence of fixed distribution expectations. There is also potential for some offsetting increase in compensation risk due to the uncontrolled 'weighting' on information introduced to that process. The negative feelings expressed by some subordinates who feel unfavorably treated by the calibration process are consistent with perceptions of higher-level compensation risk.

One of the aims of this study is to provide a holistic view of the multiple layers of subjectivity evident in PMRS, and the interactions among them. We find that subjective measures are used, but are less pervasive in these firms than the use of subjective ratings across both objective and subjective criteria. The subjective overlay implicitly affects both weightings and ratings. The impact of subjectivity on bonus determination is direct in one firm, but indirect via the impact on performance measures that feed into bonus determination in the others. Subjective adjustments to counter subordinate self-serving behavior are more common than adjustments for the impact of uncontrollable. These elements work together to mutually enhance congruence and informativeness, with reduction in information rents appearing less relevant in these firms. Our field sites provide us with a unique opportunity to observe these elements working together.

Our findings both enrich and challenge current understanding of the treatment of subjectivity in performance measurement in the incentive contracting literature. Specifically, we seek to understand how managers use subjectivity to construct and use performance measures that promote congruity, are informative about an agent's effort, and that reduce the potential for agents to extract information rents. We find that, in the main, managers seek to design contracts that will enable them to determine how outcomes were achieved by their subordinates. While incentive contacting theory would suggest that subjective performance measures will be noisy, managers address this by drawing on information that can provide reliable and valid assessments of a subordinate's effort (level and direction) and action strategies, and triangulating this data from a range of sources. Managers are prepared to submit their decisions to the scrutiny of their peers, believing that this both enhances the

quality of their own decisions as well as reducing the potential for self-serving behavior by other managers. They undertake these processes within the context of a firm-level PMRS that, while providing constraints around the operational aspects of the process, also allow them scope in their choice and use of measures that are most appropriate for the nature of the work.

Our evidence indicates that managers expend care and time in their consideration of the performance of each subordinate and do not have a fixed approach nor claim to have 'easy' answers. While their decisions around different aspects of the performance measurement process are largely intuitive and opaque, our data provide evidence of the type of thinking that informs their evaluations; for example, in seeking to adjust for uncontrollable events they consciously ask themselves whether a subordinate's actions were appropriate for a particular situation. While we find minimal evidence of inefficiency at the level of individual manager, we acknowledge that this may be understated due to desirability bias in the collection of interview data. However, we are able to identify disciplining mechanisms that reduce corruptibility of subjective information, and thus tilt the scales in favor of informativeness and efficiency. We are unable to assess efficiency in terms of PMRS design and resource intensity. However it is clear from the system descriptions that they are significantly more resource intensive than PMRS based entirely on objective measures. Furthermore, the resources are consumed largely in bureaucratic processes that mitigate both the compensation risk for subordinates and the risk of errors in personnel decisions. Both supervisors and subordinates in our study are largely supportive of the effectiveness of those bureaucratic mechanisms. Interestingly, even in InvestServ, where bureaucratic processes were less structured and transparent, participants still considered the processes robust, trustworthy and unbiased. We are unable to assess whether compensation risk is optimally shared within these firms, but our interviews, reflecting the views of both supervisor and subordinate managers, did not suggest either disproportionate risk or excessive compensation cost. Holmstrom (1979) argues that an agent's risk is low when performance measures provide accurate, timely and informative signals about their contribution to firm value. Our data points to the use of safeguards around subjectivity which enhance the accuracy, timeliness and informativeness of subjective ratings.

In addressing the question of how subjectivity is rendered informative, we offer the following summary observations from our findings which represent propositions that complement existing theory on the role of subjectivity in incentive contracting:

- Subjectivity is both purposeful and pervasive in PMRS. The use of subjective measures and subjective ratings on objective measures are explicitly contracted. While attempts are made to identify relatively objective metrics (even for qualitative task attributes) the level of subjective ratings on these criteria, along with additional subjective judgements in calibration are not consistent with an inherent preference for objectivity.
- Subjective measures, ratings and weightings are critical to both congruence and informativeness of performance measures. While subjective measures only address congruence, subjective ratings and weightings jointly enhance completeness, allow for the informed assessment of effort (both level and direction) and allow for ex post settling up.
- There are a range of PMRS solutions adopted to limit the noise and corruptibility of subjectivity in PMRS, which do not involve fixed wages or job redesign. These include bureaucratic processes that discipline supervisors against leniency and favoritism, establish networks of information about supervisor and subordinate behavior, and reduce the risk of self-serving behavior by both supervisors and subordinates. These processes, which rely heavily on calibration, are effective in mitigating specific risks of leniency, favoritism and inaccurate assessments, but not centrality bias. Calibration may increase centrality bias or rating compression.

Limitations

Our findings are naturally 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 use of subjectivity may be observed in other settings. We have picked 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. Furthermore, we were only able to speak with both supervisors and subordinates identified by the firm. This bias in respondent selection increases the risk of exposure to a 'sanitized' account of reality. Furthermore, in a study such as this, probing performance measurement and reward practice, there is a high risk of desirability bias, and we do not doubt that our findings reflect this. However we do not believe we have reported a completely sanitized account of PMRS practices in these firms. The responses to our questions were varied, credible and supported by examples, firm-level guidelines and policies. We have cited across all firms negative aspects of both processes and outcomes. While we acknowledge these limitations, we believe this study is an effective response to calls in the literature to bring 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.

References

Abernethy, M. A., Bouwens, J., & Van Lent, L. (2013). The Role of Performance Measures in the Intertemporal Decisions of Business Unit Managers. *Contemporary Accounting Research, 30*(3), 925-961.

Anderson, S. W., Dekker, H., Sedatole, K. L., & Wiersma, E. (2013). *Halo effects in subjective performance evaluation*. Working paper.

Bailey, W. J., Hecht, G., & Towry, K. L. (2011). Dividing the Pie: The Influence of Managerial Discretion Extent on Bonus Pool Allocation. *Contemporary Accounting Research, 28*(5), 1562-1584.

Baiman, S., & Rajan, M.V. (1995). The Informational Advantages of Discretionary Bonus Schemes. *The Accounting Review, 70(4),* 557-579.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective Performance Measures In Optimal Incentive Contracts. *The Quarterly Journal of Economics, 109*(4), 1125-1156.

Bol, J. (2011). The Determinants and Performance Effects of Managers' Performance Evaluation Biases. *The Accounting Review, 86*(5), 1549-1575.

Bol, J., Keune, T. M., Matsumura, E. M., & Shin, J. Y. (2010). Supervisor Discretion in Target Setting: An Empirical Investigation. *The Accounting Review, 85*(6), 1861-1886.

Bol, J., & Leiby, J. (2011). The Forward-Looking Role of Subjectivity in Performance and Promotion Evaluations: Evidence from Professional Services. Working paper, University of Illinois Urbana-Champagne

Bol, J., & Smith, S. D. (2011). Spillover Effects in Subjective Performance Evaluation: Bias and the Asymmetric Influence of Controllability. *The Accounting Review, 86*(4), 1213-1230.

Bol, J. C. (2008). Subjectivity in Compensation Contracting. *Journal of Accounting Literature, 27*, 1-24.

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

Feltham, G.A., & Xie, J. (1994). Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations. *The Accounting Review, 69(3),* 429-453.

Gibbs, M. J., Merchant, K. A., Van der Stede, W. A., & Vargus, M. E. (2004). Determinants and Effects of Subjectivity in Incentives. *The Accounting Review, 79*(2), 409-436.

Gibbs, M. J., Merchant, K. A., Van der Stede, W. A., & Vargus, M. E. (2009). Performance Measure Properties and Incentive System Design. *Industrial Relations, 48*(2), 237-264.

Grabner, I., Kunneke, J., & Moers, F. (2014). *How to mitigate bias in performance evaluations: An analysis of the consequences of supervisors' evaluation behavior*.

Grabner, I., & Moers, F. (2013). Managers' Choices of Performance Measures in Promotion Decisions: An Analysis of Alternative Job Assignments. *Journal of Accounting Research, 51*(5), 1187-1220.

Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics, 10*(1), 74-91.

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law Economics and Organization, 7*, 24-52.

Hoppe, F., & Moers, F. (2011). The Choice of Different Types of Subjectivity in CEO Annual Bonus Contracts. *The Accounting Review, 86*(6), 2023-2046.

Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard. *The Accounting Review, 78*(3), 725-758.

Lambert, R. A. (2001). Contracting theory and accounting. *Journal of Accounting and Economics, 32*, 3-87.

Lillis, A. M., & Mundy, J. (2005). Cross-Sectional Field Studies in Management Accounting Research: Closing the Gap Between Surveys and Case-Studies. *Journal of Management Accounting Research, 17*, 119-144.

MacLeod, W. B., & Parent, D. (1999). Job Characteristics and the Form of Compensation. *Research in Labor Economics, 18*, 177-242.

Maher, M. W. (1987). The Use of relative Performance Evaluation in Organizations. In J. Bruns, William J & R. S. Kaplan (Eds.), *Accounting & Management: Field Study Perspectives* (pp. 295-315). Boston: Harvard Business School Press.
Merchant, K. A. (1987). How and Why Firms Disregard the Controllability Principle. In J. Bruns, William J & R. S. Kaplan (Eds.), *Accounting & Management: Field Study Perspectives* (pp. 316-338). Boston: Harvard Business School Press.
Merchant, K. A. (2006). Measuring general managers' performances: Market, accounting and combination-of-measures systems. *Accounting, Auditing & Accountability Journal, 19*(6), 893-917.
Miles, M. B., & Huberman, A. B. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
Moers, F. (2005). Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society, 30*, 67-80.
Prendergast, C. (2002). Uncertainty and Incentives. *Journal of Labor Economics, 20*(2), 115-137.
Prendergast, C., & Topel, R. H. (1996). Favoritism in Organizations. *Journal of Political Economy, 104*(5), 958-978.
Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature, 37(1*), 7-63.
Sedatole, K. L., & Woods, A. (2013). *Consistency and Organizational Justice: The role of calibration committees in subjective performance evaluation systems*. Working paper.
Woods, A. (2012). Subjective adjustments to objective performance measures: The influence of prior performance. *Accounting, Organizations & Society, 37*, 403-425.
WorldatWork, & Consulting, D. (2014). Incentive Pay Practices Survey: Publicly Traded Companies.

Appendix A – Interview Guides

<u>Supervisor</u> (eg profit center manager or higher):

Section 1: Demographics:

- Role title and description
- Time in role
- Time with company
- Number of direct reports (= number of people whose performance you evaluate?)

Section 2: Perceptions of corporate performance measurement practice:

What is the corporate approach to performance measurement at your level of management (from the perspective of the 'supervisor')?

- 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 measure are used? Are these standardized across managers at the same level?
    - How does discretion enter the process?
      - Eg the measures themselves are qualitative?
      - The weightings among measures can vary according to your discretion?
      - Some objective measures are supplemented with some qualitative measures?
      - The measures are objective, but the overall rating is subject to your 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?
    - Eg 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?
    - Eg 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?
    - How do you think discretion impacts the outcomes of performance measurement?
    - Would you prefer that performance measurement was more objective or subjective? Why?
    - Do you think other managers at your level hold similar or different views of the process?
        - Explain?

Subordinate (eg direct report to profit center or higher level manager):

Section 1. Demographics:

- Role title and description
- Time in role
- Time with company
- 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 process:

- 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 measure are used? Are these standardized across managers at the same level?
        - How does discretion enter the process?
            - Eg the measures themselves are qualitative?
            - The weightings among measures can vary according to your supervisor's discretion?
            - Some objective measures are supplemented with some qualitative measures?
            - The measures are objective, but the overall rating is subject to your supervisor's 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?
        - Eg the balance between objectivity/discretion?
        - The relationship between performance measures, performance evaluation, and rewards (both direct bonuses and other rewards such as promotions)?
- Do you think the process is applied consistently at your level of management?
    - Is there variation in the level of discretion introduced to the process by some managers when evaluating managers at your level?
        - What sort of variation?
        - What is the impact of this variation?
- 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?
    - How do you think subjectivity impacts the outcomes of performance measurement?
    - Would you prefer that performance measurement was more objective or subjective? Why?
    - Does the performance measurement practice adopted impact on your operating decisions in particular ways?
            - How? Can you give examples?
- Do you think other managers at your level would hold similar views of the process?
    - Explain potential differing views
- Do you think the process affects some of your peers in ways that it does not impact on you? Eg impact on decisions, morale

Appendix B – Demographic Information

	ProServ	**InvestServ**	**FoodMan**	**EquipMan**
Participant Information				
Number of Interviews	9	11	12	12
Average length of interview (mins)	53	51	47	38
Minimum interview length (mins)	40	35	34	23
Maximum interview length (mins)	60	60	61	49
Average tenure at firm (years)	15	10	7	22
Average length in role (years)	3	5	3	2

Appendix C – Summary of Performance Measurement Process for Each Firm

Case 1: ProServ

Structure: ProServ is a multinational partnership operating in the professional services industry. It has three business units of which one forms the focus of this study. Work activity is based on discrete projects fulfilled by teams that exist only for the duration of each project.

How ProServ incorporates subjectivity: Formal performance measurement takes place once per year. The business unit has six broad strategic priorities, for which each subordinate has a set of 12-18 goals negotiated personally with his or her supervisor. Across all goals, there are only three objective quantitative measures, two of which are financial. Measures may be assigned to other goals but these are not considered to be 'hard' targets. Other qualitative goals, such as 'Build Relationships' do not have specific measures attached to them. Subjectivity is thus evident in the supervisor's initial selection of goals, discretion over whether or not to include measures, and choice of associated targets.

At the end of the year the supervisor and subordinate meet to discuss the subordinate's performance. The discussion is based on the contents of two qualitative reports: a self-assessment prepared by the subordinate and a written narrative prepared by the supervisor. Both reports are informed by the non-anonymized opinions of managers, peers, and clients who have worked with the subordinate on various projects. After the meeting the supervisor provides HR with a provisional rating from 1 (high performer) to 4 (low performer) of the subordinate's overall performance. The rating is based on the supervisor's overall evaluation of performance against the goals (the 'what') as well as the behaviors evident in attaining those goals (the 'how'), but there is no separate assessment of each nor a formula to convert scores into an overall rating. Behaviors are evaluated according to a company-wide framework that contains a list of criteria against which supervisors exercise discretion in assessing how each goal was achieved.

The overall rating given to each subordinate is not discussed in the appraisal meeting but instead forms the basis of a subsequent discussion among supervisors to establish the relative performance of all subordinates at the same level of the organization. A pre-specified distribution curve (subjectively determined by the senior partners) establishes the eventual proportion of subordinates who will be assigned to each rating: 10-15% per cent will obtain a one rating, 55-65% a two, 20-25% a three, and 10% a four. The meeting between supervisors is designed to calibrate the provisional ratings against the pre-specified distribution curve. Each rating is associated with a pre-specified bonus payment and merit increase expressed as a percentage of base salary. For example, those who receive a rating of 4 might receive neither bonus or merit increases while those who attain a 1 will be awarded the maximum bonus and merit increases.

Further layers of subjectivity are introduced through this calibration process. There is usually disagreement over the ratings of approximately ten per cent of subordinates. The final rating for these subordinates is often determined by a decision to rely on 'hard' targets.

Supervisors also seek to differentiate 'high' from 'low' performing 2s because the large number of subordinates awarded a 2 rating is regarded as insufficiently discriminating for the purposes of identifying those with the most potential.

Subjectivity is also apparent in the eventual allocation of bonuses and merit increases awarded to each subordinate after the calibration meeting. Despite the formal link between the attainment of a specific rating and the allocation of subsequent rewards, the senior partners responsible for approving the rewards make discretionary adjustments both upwards and downwards to some payments based on individual-specific factors.

Case 2: InvestServ

Structure: InvestServ is an international asset management firm with a flat organizational structure. The focus of this study is a subsidiary with several hundred employees. The firm's total bonus pool is drawn from the asset management teams, each of which has its own P&L, and is then allocated to all

teams including the support functions. The head of each team evaluates the team members. Our study excludes the sales area because bonuses in that part of the business are paid on a commission basis. For all areas of the business, bonuses and salaries are in line with other firms in the industry.

How InvestServ incorporates subjectivity: At the end of each financial year the executive committee allocates a bonus pool to each team. The formula behind the allocation is not known to the team managers but is understood to be determined by the prior year's performance of both the firm and each individual team. Each supervisor has complete discretion to decide the basis on which team members are evaluated. Most supervisors choose to use a set of high level goals, but specific measures or targets are uncommon even among the investment teams for which performance is published and benchmarked externally.

InvestServ's appraisal process allows supervisors to decide whether or not they apply a performance rating to each individual; it is not as a critical element of the process. This discretion exists because each team's bonus pool is independent of the others and therefore there is no requirement to calibrate the performance of individuals across the firm. A member of the executive management provides a 'sanity check' on the amounts of bonus allocated by the supervisors to each individual. There are no merit increases. Supervisors exercise discretion in recommending promotions on the basis of overall performance and potential. Supervisors also have complete discretion in how they allocate bonuses among their team members.

Case 3: FoodMan

Structure: FoodMan is the subsidiary of a large multinational parent company. The Company employs several thousand employees. It has three Business Units – one regional and two product-focused, and several functional support units including R&D and HR. Functional specialists operate within a matrix structure with both functional and product-group/regional accountability. Participants in the study were drawn from all subunits and a range of functions.

How FoodMan incorporates subjectivity: Formal performance measurement occurs twice per annum, but ratings are only applied at year end. All managers are eligible for a performance incentive:

- Company financial performance (3 components). Performance on these company-level deliverables determines the size of the bonus pool.
- Business unit performance. Business unit performance on financial deliverables determines the share of the bonus pool for each business unit.
- Individual performance. Individuals have typically 5 goals. These goals are largely quantified. Individuals are assessed as to whether they meet, exceed or fall short of negotiated targets on these individual goals. They are assessed not only on 'what' they have achieved relative to these goals, but also 'how' they have achieved them.

The result of applying the above process is a formulaic performance incentive calculation:
Individual bonus = x[weighted company performance indicators] +y[weighted business unit performance indicators] + z[weighted individual performance indicators].

While managers consider objectivity and quantification of performance, there is considerable subjectivity in the final determination of individual bonuses. There is variation in the weighting placed on company, business unit and individual performance (x, y and z in the above formula) depending on level of management. There is discretion in identifying appropriate business unit-level metrics, setting targets that will form the benchmark for the assessment of actual business unit performance, and evaluating performance relative to actual conditions that prevailed during the performance period. The individual performance component is the most subjective. In many cases the measures themselves might be subjective in nature (eg supporting the launch of a new product). In addition, the choice of metrics, targets set, assessments of performance against the target on a 5-point scale for each metric and the weighting on each of the metrics in the final score are all determined subjectively. For each individual measure, the rating reflects not only what is achieved (which is in some cases quantitative and indisputable) but also how it is achieved. This is in all cases subjectively determined. For example a sales target that is exceeded may be assessed as having been achieved

through single-minded, unco-operative, competitive behavior, which would then be factored into the rating.

Overlaying all of these individual performance ratings is a calibration process in which managers debate the ratings given to subordinates, and in which cross-functional interactions (beyond the supervisor/subordinate dyad) are discussed.

Individual performance, and particularly the rating on 'how' targets have been achieved, are used to determine merit increases. Merit increases are calibrated to the market, but the variation in the index is very small. Supervisors exercise some discretion when someone is near the top or bottom of a salary scale. Underperformers (on individual metrics, especially the "how") get zero merit increase. Merit increases generally range up to about 3% (with the performance-based variation typically only occurring between about 2.5% and 3%).

Promotability is assessed with another layer of subjectivity, in which subordinates are assessed for both overall performance (which comes out of the process above) and agility, or ability to manage the complex demands of new roles. High performers can remain promotable within their function, but high performing agile employees are targeted for promotion into more complex managerial roles.

Case 4: EquipMan

Structure: EquipMan is a global leader in equipment manufacturing. Two segments within one of the company's several business units are the focus of this case study. Each segment in the study has over one billion in revenue and over 3,000 employees worldwide. EquipMan is a matrixed organization with each manager having a solid-line reporting structure to a segment head and a dotted-line reporting structure to a functional head (e.g. finance, quality, engineering, etc.). The segment head (supervisor) evaluates manager (subordinate) performance with input from the functional heads.

How EquipMan incorporates subjectivity: Each subordinate is assessed against 4-8 work plan objectives and 2-4 development objectives, all of which may be unique to the individual and are determined jointly with the supervisor. All objectives should be SMART: Specific, Measureable, Actionable, Realistic and Time bound. Human resources provides lists describing specific behaviors to help supervisors make the less quantifiable objectives easier to measure.

Performance is rated based on a 1, 2, 3 scale with 1 being exceeds expectations, 2 being meets expectations and 3 being fails to meet expectation. Performance is evaluated not only on 'what' objectives were met but also on 'how' they were met. 'How' is framed by the company's five criteria for leadership capability. The weight placed on 'how' is not dictated and increases with the level of the subordinate.

The typical process for determining the 1, 2, 3 rating is for the subordinate to self-assess his/her performance, then the supervisor assesses the manager's performance, then they meet to discuss the assessments. Input from multiple stakeholders is required and is used in the evaluations. Finally, performance ratings are subject to a calibration process. The supervisor brings his/her rating for each subordinate into the calibration meeting. The supervisor and his/her peers as well as one level above the supervisor participate in the calibration meetings. The purpose of the calibration process is to compare how subordinates within each compensation class perform relative to each other. While each supervisor within his/her area uses relative performance evaluation to rate his/her subordinates, the calibration process applies relative performance evaluation across supervisors and subordinates. Discussion typically focuses on the outliers (those ranked 1 and 3) and on those who sit near the boundaries between categories.

Merit increases are based on performance ratings. A merit increase pool is established and changes from year to year depending on the corporate-level budget. All employees are expected to be rated 2's and are paid more or less based on where they are rated compared to a 2. Unofficially, the 2 category is broken down into 2+, 2 and 2-. Supervisors have discretion to give 2+'s more merit increase than 2-'s, as long as they stay within the constraints of the total pool.