Sooner or Later:
The Effects of Timing on Managers’ Discretionary Weighting of Multiple Performance Measures

Stephen Deason
Emory University
stephen.deason@emory.edu

Gary Hecht
University of Wisconsin-Madison
ghecht@bus.wisc.edu

William B. Tayler
Brigham Young University
tayler@byu.edu

Kristy L. Towry
Emory University
kristy_towry@bus.emory.edu

March, 2013

We appreciate helpful comments and suggestions from two anonymous reviewers for the Management Accounting Section meeting, as well as from Jeremy Bentley, Melanie Millar, Todd Thornock, Steve Smith, and workshop participants at Brigham Young University’s Accounting Research Symposium and Emory University’s Behavioral Brownbag Series. We are grateful for the research assistance of Julie Mohamed.
Sooner or Later:  
The Effects of Timing on Managers’ Discretionary Weighting  
of Multiple Performance Measures

Abstract

We investigate whether the weights placed on multiple performance measures for the purpose of determining performance-contingent pay depends on whether weights are determined before or after employees exert effort. Analytic accounting research implicitly assumes that the timing with which discretionary weighting occurs will not affect the weights selected. However, we propose that while the overall objective in determining these weights (to guide employees toward desired actions) is unaltered by timing, the immediate objective of performance measure weighting does vary across time. Specifically, prior to employee effort, the immediate objective is to motivate employee effort in line with firm objectives. After employee effort, the immediate objective is to evaluate employee effort in a way that can be justified to the employee as fair. Thus, we expect, and find, that evaluators weight congruent measures more heavily when weightings are determined ex ante (before employee effort) and weight precise measures more heavily when weightings are determined ex post. This effect is mitigated when ex post measure outcomes indicate relatively favorable outcomes for congruent measures.

Key Words: subjectivity, performance evaluation, discretion, contracting, compensation
I. INTRODUCTION

Accounting scholars have long been interested in performance-contingent pay as a component of the overall management control system. This topic is complex because of the various ways in which performance-contingent pay may be implemented. For example, the analytic literature has traditionally assumed that performance-contingent pay is implemented via *ex ante* contracts that explicitly tie pay to the outcomes of pre-specified performance measures. The focus of this rich literature has been on various measure attributes (e.g., congruence and precision) that determine the weights placed on each measure in the optimal contract (Feltham and Xie 1994). However, contracts are often incomplete (Williamson 1985), requiring performance-contingent pay to be based on management discretion that is applied after the employee chooses his action. In this study, we investigate whether the weights placed on multiple performance measures for the purpose of determining performance-contingent pay depend on when those weights are determined (*ex ante* or *ex post*, relative to the employee’s action choice). Further, we investigate whether measure outcomes (only available *ex post*) also influence these weightings.

Prior research highlights firms’ widespread use of multiple performance measures in performance evaluation and compensation systems, especially given that single performance measures usually do not fully capture an employee’s performance (Baker 1992, 2000). Such systems, however, are often limited by “the difficulty, uncertainty, and interpersonal conflict managers experience in deciding upon performance-measure weights [that] will best motivate employees to achieve a given goal” (Krishnan et al. 2005, 1164). As implied by this literature, many factors can influence managers’ weighting, including individual characteristics (i.e., lack of expertise, as discussed by Ittner et al. (2003), cognitive limitations as investigated by Krishnan
et al. (2005), and the spillover of available information across measures as investigated by Bol and Smith (2011)). In this paper, we consider features of the performance measurement system itself; specifically, at what point in the measurement and evaluation process a manager determines the weighting of multiple performance measures.

The analytic contracting literature implicitly assumes that timing will not affect performance measure weights. Specifically, a relatively new branch of this literature (Ederhof 2010; Ederhof et al. 2010; Rajan and Reichelstein 2006, 2009) examines the use of managerial discretion in determining performance-contingent pay. While acknowledging that in the real world, discretion is generally applied after performance, these papers model discretion as an ex ante process. They assume that prior to the employee’s action choice, the evaluating manager anticipates (in the form of probability distributions) all available performance measures and commits to the weights to be placed on each one. Discretion is represented simply by the presence of “subjective measures,” the weights on which cannot be enforced in a court of law, though principals in these models have no incentives to deviate from their committed weights. Thus, the timing at which the weights are determined is assumed away as irrelevant. In this paper, we challenge that assumption.

We provide theory to suggest that the weights on multiple measures will depend on the time at which those weights are determined relative to the employee’s action choice. This effect stems from timing-driven differences in a manager’s “weighting objective” – the immediate purpose underlying the performance measure weighting task. Regardless of timing, the ultimate purpose of performance-contingent pay is to influence employees to take actions in the firm’s best interest. However, the more immediate weighting objective will depend on when that task is performed. When the weights are determined before the employee takes action, as assumed in
the analytic literature, the weighting objective will be to motivate employees to take preferred actions in the upcoming period. This will increase the salience of congruence as a performance measure attribute, and will lead the manager to place greater weight on measures that are more congruent with the organization’s goals. In contrast, when the weights are determined after the employee’s action, the weighting objective will be to evaluate this prior action in a manner that can be justified to the employee as fair. This will increase the salience of precision as a performance measure attribute, and will lead the manager to place greater weight on more precise measures. In summary, our theory suggests that congruent measures will be more heavily weighted in ex ante weighting decisions than in ex post weighting decisions; conversely, precise measures will be more heavily weighted in ex post weighting decisions than in ex ante weighting decisions.

In practice, the differences between ex ante and ex post weighting decisions go beyond timing. When weighting is performed after the employee’s action, performance measure outcomes are typically known. We predict that these performance measure outcomes will also influence performance measure weightings. Specifically, we predict that the increased weight on precise measures ex post (relative to ex ante weighting) will be mitigated when the outcomes of imprecise measures are more favorable than those of precise measures. This is because managers are unlikely to feel justification pressure against placing weight on imprecise measures if the outcomes on those measures are favorable. Conversely, we expect the increase in weight placed on precise measures ex post, (relative to ex ante weighting) to be exacerbated when the outcomes of imprecise measures are less favorable than those of precise measures. This is because managers are even more likely to feel justification pressure against placing weight on imprecise measures if those measures are unfavorable.
We investigate our research questions using an experiment in which MBA student participants (averaging more than five years of work experience) complete a case in which they assume the role of an evaluating supervisor of a hypothetical quality control manager. The case describes two measures of the quality control manager’s performance, with each measure superior to the other on one of two dimensions: precision and congruity. That is, one measure is more precise, but less congruent, and the other measure is less precise, but more congruent. For expositional ease, we refer to these measures as the precise/incongruent measure and the imprecise/congruent measure. Participants’ task is to determine the weight (i.e., percentage of the quality control manager’s compensation) placed on each measure for performance evaluation and compensation purposes.

We use a double-nested design in which we manipulate the timing of participants’ performance measure weighting decision, the availability of outcome information, and the valence of that information. First, we manipulate the timing of the participants’ performance measure weighting decision relative to the employee’s action (ex ante or ex post). Nested within the ex post condition, we manipulate the availability of outcome information. In the outcomes unknown condition, participants decide on their weights after the employee has chosen an action, but the before learning about any performance measure outcomes. While this condition is likely uncommon in the real world, the experimental setting allows us to examine this condition in order to more effectively test our theory. Next, we generate two separate conditions in which outcomes are known by manipulating the valence of outcome information related to the two performance measures. In the favorable precise/incongruent condition, the precise/incongruent measure reflects better performance than the imprecise/congruent measure. In the favorable
imprecise/congruent condition, the imprecise/congruent measure reflects better performance than the precise/incongruent measure.

The results generally support our hypotheses. Specifically, when assigning weights ex ante participants place greater weight on the imprecise/congruent measure (and by construction, less weight on the precise/incongruent measure) than when they assign weights ex post. An analysis of post-experimental responses to a questionnaire suggests that this effect is driven by justification pressure being greater in the ex post setting. Further, this effect is mitigated when the imprecise/congruent measure outcome is favorable. Contrary to our predictions, however, the effect is not exacerbated when the imprecise/congruent measure is unfavorable.

The remainder of this manuscript is organized as follows. Section II presents background and hypotheses. Section III describes our experimental design. Section IV provides analysis and results. Section V concludes the manuscript.

II. THEORY

Background

Performance-based compensation is a key factor within firms’ management control systems. Accountants’ interest in this topic stems from many sources, including the quality of accounting measures, which serve as a primary component of compensation contracts. A large and growing literature in accounting (e.g., Banker and Datar 1989, Feltham and Xie 1994, Datar et al. 2001) examines labor contracting, specifying how the optimal contract weight placed on a given performance measure depends on attributes of that measure such as precision, congruity, and sensitivity. In this manuscript, we focus on two measure attributes – congruity and precision. We rely on Feltham and Xie’s (1994) definitions of these two attributes, specifically:
“congruity refers to the degree of congruence between the impact of the agent’s action on his performance measure and on the principal’s expected gross payoff, while precision refers to the noise in the performance measure” (Feltham and Xie 1994, 434).¹

In contrast to our growing understanding of optimal performance measure weighting, relatively little is known about actual performance measure weighting. In one of the few examinations of actual weighting decisions, Krishnan et al. (2005) show that managers do not generally assign weights in a manner consistent with the optimal contract. We extend this literature by examining an important ecological factor: the timing with which weights are assigned to performance measures for compensation purposes.

The timing of weight determination varies considerably in practice. The accounting literature has generally assumed that weights are determined ex ante, resulting in a complete contract that specifies “a unique payment for each conceivable outcome” (Al-Najjar and Casadesus-Masanell 2007). An example of this ex ante performance measure weighting is observed when a compensation contract establishes sales personnel bonuses before the accounting period to which the bonus pertains begins (e.g., 60% of the bonus is based on net revenues generated and 40% of the bonus is based on percentage of customers with repeat sales). However, in many settings, environmental factors complicate the problem of developing and implementing ex ante complete contracts (Gibbs et al. 2003). Thus, firms often endow evaluating managers with discretion to determine various compensation-related factors on an ex

¹ Sensitivity measures “the change in the expected value of a measure with changes in the level of effort of the agent” (Banker and Datar 1989, 29). More simply, sensitivity refers to “the extent to which an agent’s action affects a measure” (Krishnan et al. 2005, 1166). Sensitivity is not an attribute of the measure, but rather of the relation between the measure and a given action. Therefore, we do not discuss or model sensitivity in our study. Please refer to Section V for further discussion.
For example, Rajan and Reichelstein (2009, 209) note that “several studies have documented that boards have considerable discretion in setting executive pay even after key variables for that given time period have been realized.”

Despite this acknowledgement by Rajan and Reichelstein (2009), the analytic literature generally ignores the effect of timing (e.g., see Rajan and Reichelstein (2006)). In the next section, we develop theory that suggests that timing influences managers’ weighting of performance measures. Further, we provide theory explaining how measure outcomes influence performance measure weightings when weights are assigned on an \textit{ex post} basis.

**Theory Development**

\textit{Weighting Decision Timing}

We propose that the weights on multiple measures will depend on the time at which those weights are determined: either before or after the employee has exerted effort. In essence, when assigning weights \textit{ex ante} with respect to the employee’s actions, the manager is the “first-mover” in the current accounting period, establishing the basis of employees’ compensation before the employee takes action. With \textit{ex post} weighting, the manager is a second mover. One might initially assume that this timing does not matter, because in a multiple-period world, both parties will understand that their actions have potential long-term consequences, and so the \textit{ex ante} vs. \textit{ex post} distinction becomes meaningless. Indeed, economic theory generally assumes that individuals broadly bracket choices, considering current period choices in light of their

\footnote{This discretion can take many forms, including discretionary bonus pools, discretionary adjustments to objective performance measures or related targets, and discretionary weighting of multiple performance measures (Bol 2008). A growing literature addresses when firms endow managers with discretion (Ederhof 2010; Hoppe and Moers 2011) and how managers use this discretion (Bailey et al. 2011; Bol et al. 2012; Maas et al. 2012), though only the aforementioned Krishnan et al. (2005) study investigates discretionary weighting per se.}
multi-period ramifications. However, recent research on choice bracketing suggests that individuals bracket their choices more narrowly, considering repeated decisions one at a time. As described by Read et al. (1999, p. 172), “Broad bracketing allows people to consider all the hedonic consequences of their actions, and hence promotes utility maximization. Narrow bracketing, on the other hand, is like fighting a war one battle at a time with no overall guiding strategy. . .” In fact, Ding (2012) demonstrates that even in an interactive setting (specifically, a multi-period trust game), both the first mover’s and second mover’s choices are affected by narrow bracketing.

Because of this narrow choice bracketing, managers who determine their subordinates’ performance measure weightings for compensation purposes will have different “weighting objectives” at different points in time. Regardless of timing, the ultimate purpose of performance-contingent pay is to align employees’ incentives with those of the firm. However, the more immediate weighting objective will depend on when the weighting decision is made in relation to employees’ actions. When the weights are determined before the employee takes action, as assumed in the analytic literature, the manager’s weighting objective will be to motivate employees to take preferred actions in the upcoming period. In contrast, when the weights are determined after the employee’s action, the weighting objective will be to evaluate this prior performance.

The weighting objective, as affected by the timing of the weighting decision, influences the salience of different performance measure attributes. When the weighting objective is to motivate future behavior (when weights are determined ex ante), the manager’s focus will be on encouraging those behaviors that are most consistent with the firm’s best interest. This focus will increase the salience of congruence as a performance measure attribute, leading the manager
to place greater weight on measures that are more congruent with the organization’s goals.

When the weighting objective is to evaluate prior performance (when the weights are determined
ex post), the manager’s focus will be on making decisions that can be justified to the employee as fair. The manager will want to avoid rewarding or penalizing the employee for factors outside the employee’s control. This consideration is especially strong when the employee’s perception of the fairness of his/her compensation is threatened (i.e. when the employee might be “penalized” or under-compensated relative to his/her effort). This focus on fairness increases the salience of precision as a performance measure attribute, leading the manager to place greater weight on more precise measures. Thus, managers assigning performance measure weights ex post will assign more weight to more precise measures, whereas managers assigning performance measure weights ex ante will assign more weight to more congruent measures.

Notably, the prediction implied by the previous discussion is also consistent with construal level theory. Construal level theory suggests that psychological distance – as determined by a number of factors – influences the level of detail with which individuals consider the antecedents and consequences of events or actions (e.g., Bar-Anan et al. 2006; Trope et al. 2007). Specifically, the greater (smaller) the psychological distance, the more abstract (concrete) the individual’s representation of the setting. Many factors can influence psychological distance, including temporal proximity (Trope and Liberman 2010). A common example in psychology literature on construal level theory is how individuals think about an upcoming vacation. When the vacation is scheduled well in advance, a more abstract or general representation of the vacation is maintained (e.g., the individual thinks of the vacation as relaxing or romantic). In contrast, when the vacation will take place in the short-term, the
individual’s representation or consideration of the vacation is more detailed (e.g., the individual envisions the logistics of transportation and lodging).

In our setting, this theory would suggest that the timing of the weighting decisions influences the level of detail with which the manager envisions the implications of the weighting decision for an employee’s compensation. When a manager determines multiple measures’ weights *ex ante*, the weighting decision and actual compensation determination are temporally distant. Thus, the manager is likely to think about the connection between the weighting decision and the implications of the weighting decision for the employee’s compensation at a relatively general (i.e., abstract) level. In contrast, when a manager determines weights *ex post*, the implications for the employee’s compensation are more immediate. Thus, the manager is likely to consider the implications of the weighting decision for the employee’s compensation at a much more detailed level. This more detailed level likely leads to a more detailed consideration of an employee’s response to the implications of the weighting decision, thereby making more salient ways in which the employee perceives the weighting decision as unfair. Similar to our earlier argument, this more readily available perception of a potential response by employees leads a manager to focus on precision when assigning weights *ex post*.

This discussion results in the following hypothesis:

H1: Congruent measures will be more heavily weighted in *ex ante* weighting decisions than in *ex post* weighting decisions, while precise measures will be more heavily weighted in *ex post* weighting decisions than in *ex ante* weighting decisions.3

Measure Outcomes

---

3 Implied by this hypothesis is a tradeoff between congruence and precision in the design of performance measures. While not always evident (some measures are both congruent and precise), it is precisely this tradeoff in practice that makes the weighting decision difficult. Measures that dominate other measures on both dimensions should unambiguously be weighted heavily.
Though we have discussed the effect of timing on managers’ weighting of multiple performance measures, we have, thus far, ignored an important feature of the ex post scenario. Specifically, when weights are assigned ex post, employee-action outcomes and performance measure realizations are determined. Thus, when assigning weights to measures ex post, managers know the outcomes. While knowledge of employee’s performance likely has many consequences, the effect of interest in our setting relates to the manager’s anticipation of the employee’s reaction to assigned weights. When a manager assigns weights with knowledge of performance measure outcomes, justification pressure varies depending on measure outcomes and across different possible measure weightings.

To illustrate this effect, suppose two performance measures reflect employee performance, and the employee performs relatively well on measure X, and relatively poorly on measure Y. Following the basic tenets of attribution theory and literature on individuals’ propensity toward self-serving biases (e.g., Miller and Ross 1975), employees will likely attribute successful performance reflected by measure X to their own effort, while attributing poor performance reflected by measure Y to some other factor. Thus, employees are more likely to deem being held accountable for measure Y as unfair. Assuming the manager anticipates this self-serving reaction by the employee, the manager likely perceives greater justification pressure associated with heavily weighting measure Y and less justification pressure associated with heavily weighting measure X. That is, the manager anticipates a greater need to explain basing the employee’s compensation on a measure that reflects poor performance. This differential justification pressure – as driven by their anticipation of employees’ claims regarding the fairness of their compensation – represent costs a manager incurs in assigning higher weights to
measure Y (Bol 2008). *Ceteris paribus*, this effect leads to higher weights on favorable measures than unfavorable measures.

Indeed, Bailey et al. (2011) and Bol et al. (2012) both note that when the outcome of a performance measure is relatively unfavorable, evaluating managers are much more likely to point out the role of non-controllable factors, such as a lack of precision, in driving those results. On the other hand, when the outcome of a measure is relatively favorable, evaluating managers are less likely to point out the role of uncontrollable factors. Neither of these studies directly address differential precision of measures or performance measure weighting, but one interpretation of their results is that precision as a performance measure attribute is more salient when an imprecise measure is unfavorable.

These results have interesting implications for the effects described in H1. Recall that when weights are assigned *ex post*, the manager’s weighting objective is to evaluate performance, and perceived justification pressure leads to increased (decreased) weight on precise (imprecise) measures (relative to the weights assigned *ex ante*). However, when imprecise measures reflect favorable performance and precise measures reflect unfavorable performance, there is not as much justification pressure to avoid heavily weighting imprecise measures. Therefore, the increase in weight assigned to precise measures that follows from *ex post* performance evaluation (relative to *ex ante* evaluation) is mitigated when imprecise measures reflect more favorable performance than precise measures. The opposite is true when precise measures reflect more favorable performance than imprecise measures. That is, the justification pressure to avoid heavily weighting imprecise measures in this setting is even greater. This results in the following two hypotheses:
H2a: The increased weight on precise measures due to \textit{ex post} weighting (relative to \textit{ex ante} weighting) will be mitigated when precise measures are less favorable than imprecise measures.

H2b: The increased weight on precise measures due to \textit{ex post} weighting (relative to \textit{ex ante} weighting) will be exacerbated when precise measures are more favorable than imprecise measures.

III. METHOD

Participants and Experiment Case

We recruited graduate students from a \textit{U.S. News} top-20 MBA program to participate in our study. One-hundred-nine participants with an average of 5.0 years of work experience completed the experiment as part of a graduate-level managerial accounting course. Over 63% of our participants have prior work experience evaluating subordinates. The experiment was administered as an in-class exercise, and uses a rich contextual, hypothetical scenario.\(^4\) Together with the experienced participants, the choice to use a context-rich scenario likely enhances the validity of our study, via memory-coding effects, as described by Haynes and Kachelmeier (1998). They argue (p. 120), “. . . there are many settings in which neither context nor experience influences behavior in general, but the \textit{combination} of contextual richness with experience enables the more experienced decision maker to draw upon knowledge structures encoded in memory, as prompted by the richer context.” We reason that the scenario we examine is one in which the experienced participants likely bring with them to the task an

\(^4\) The presence of employee-participants is not necessary for the specific research question addressed by this experiment. Related to this, a “real effort” task would introduce noise, sacrificing experimental control and thus internal validity. Further, modeling an interactive environment in which some participants have discretion over other participants’ compensation would likely constrain the type of setting we could use in our experiment, as well as our ability to manipulate performance outcomes, a key independent variable of interest in this study.
implicit understanding of the complex tradeoffs associated with performance measure weighting decisions.

The case asked participants to assume the role of an owner and president of a chemical producer whose task was to determine how the quality control manager’s bonus will be determined. Participants received information about potential quality control problems (i.e., raw material defects and mixing process errors). Participants were also informed of two actions the quality control manager could take to mitigate the negative implications of these potential problems. Specifically, the quality control manager could perform raw material inspections and mixing process analyses.

We informed participants that although they could not directly observe the quality control manager’s actions, two measures reflected the quality control manager’s performance. We use a two-measure setting to maintain simplicity in the case materials, thereby maximizing the internal validity of our study. One measure, the percent of the product not returned by customers, was described to participants as congruent with the goals of the firm, but noisy. Specifically, this measure reflected performance relating to both actions the quality control manager could undertake to enhance quality control: raw materials inspections and mixing process analyses. However, this measure was influenced by factors outside the control or influence of the manager, and thus the reported measure could range substantially around the quality control manager’s effort. The other measure, the percent of the raw materials inspected, was described to participants as noiseless, but not entirely congruent. Specifically, this measure was not
influenced by factors outside of the manager’s control. However, the measure reflected only raw materials inspections, and not mixing process analyses.\textsuperscript{5, 6}

After learning about these measures, participants assigned weights to the two measures to be used to determine the quality control manager’s bonus. Specifically, participants were told that the weights they chose “reflect how much of the quality control manager’s annual bonus will be based on each of the measures.” Individual measure weights needed to be between 0\% and 100\%, inclusive. Further, participants verified via a checkbox that the sum of the two measure weights was exactly 100\%.

We then asked participants additional questions eliciting the extent to which considerations of fairness, justifiability, and employee motivation influenced their weighting decision. Participants then wrote a memo to the quality control manager explaining how the bonus is determined and provided additional reasoning (not included in the memo to the quality control manager) behind the compensation decision. The case concluded with a post-experimental questionnaire. Overall, participants spent approximately 15 minutes completing the case.\textsuperscript{7}

\textsuperscript{5} Put another way, maximizing this non-congruent measure entails the quality control manager ignoring an important activity (i.e., analyzing mixing processes) not reflected by this measure. Thus, we have modeled incongruence via an incomplete measure. We discuss alternative characterizations of congruence in Section V.

\textsuperscript{6} We chose to use only two measures – each of which is superior to the other on precision or congruence. This characterization provides the opportunity to directly test our hypotheses, comparing participants’ weighting of multiple measures varying in congruence and provision across experiment conditions. Further, as previously noted, measures that dominate on both dimensions should unambiguously be weighted heavily.

\textsuperscript{7} Participants were not compensated for completion of the experiment. Rather, the experiment was completed as part of a management accounting course. Based on participants’ reasoning statements and consistency within completed materials, it appears that participants exerted sufficient effort in completing the experiment. Discussion of the case, results, and its relation to course topics were presented in a subsequent class meeting.
Experimental Design

Our experiment uses a double-nested, between-subjects design with four conditions (depicted in Figure 1). Each participant was randomly assigned to one of the four conditions. We manipulate the timing of the compensation decision at two levels (\textit{ex ante} versus \textit{ex post}), via our description of the scenario within the case. Participants in the \textit{ex ante} condition were told that the fiscal year for which compensation was being determined was about to begin, while participants in the \textit{ex post} conditions read that the fiscal year for which compensation was being determined was just ended. Within the \textit{ex post} condition, we manipulate the availability of outcome information. In the \textit{outcomes unknown} condition, participants decide on their weights after the employee has chosen an action, but before learning about any performance measure outcomes.

When outcomes were known, we manipulated the valence of the outcome information at two levels. In the \textit{favorable precise/incongruent} condition, the precise/incongruent measure reports an outcome of 70, and the imprecise/congruent measure reports an outcome of 30.\textsuperscript{8} Thus, the precise/incongruent measure reflects better performance than the imprecise/congruent measure. The scenario is opposite in the \textit{favorable imprecise/congruent} condition. In this condition, the imprecise/congruent measure reported an outcome of 70, and the precise/incongruent measure reports an outcome of 30. Thus, the imprecise/congruent measure reflects better performance than the precise/incongruent measure. Ultimately, our design yields four conditions, labeled as follows: \textit{ex ante, outcomes unknown, favorable precise/incongruent, and favorable imprecise/congruent}.

\[\text{[Insert Figure 1 about here]}\]

\textsuperscript{8} Participants were informed that both performance measures can have a minimum value of 0 and a maximum value of 100.
Our dependent variable is based on the weight participants assigned to the two performance measures. Given that the two measures are complements (they sum to 100%), we use the weight assigned to the imprecise/congruent measure as our primary dependent variable. We also use participants’ responses to follow-up questions as variables in additional analyses and manipulation checks.

IV. RESULTS

Manipulation Checks and Overview

Participants responded to post-experiment questions designed to assess their understanding of the case materials and ensure they attended to information constituting our manipulations. Twenty-six participants did not correctly answer questions related to attributes of the two performance measures in the case. We exclude these participants from our analysis, as their responses introduce noise into our key dependent variable and our comparisons across conditions.9 Further, one participant did not complete the case materials. Thus, of the 109 participants, we have 82 usable responses to use in testing our hypotheses.10

Table 1 presents descriptive statistics, and Figure 2 presents results for the primary dependent measure (the weight assigned to the imprecise/congruent measure). To compare our actual pattern to the hypothesized pattern, we use contrast coding (Buckless and Ravenscroft 1990). Contrast weights are as follows: 3 for ex ante, -1 for outcomes unknown, 1 for favorable imprecise/congruent, and -3 for favorable precise/incongruent. As reported in Panel A of Table

---

9 Except where noted, results are inferentially identical if these participants are included.
10 Another question asked the participants if, in choosing the performance measure weights, they were determining the employee’s past or future compensation. Surprisingly, this question was answered incorrectly by almost 50% of the participants. A follow-up discussion during the in-class debriefing session indicated that the question was vulnerable to misinterpretation. Results are inferentially identical if we eliminate those subjects who answered the question incorrectly.
2, the contrast is statistically significant \((t = 2.30, p = 0.01, \text{ one-tailed})\).\textsuperscript{11} The pattern of results presented in Figure 2, as well as the results of our contrast-coded one-way ANOVA, are generally consistent with our hypotheses, explicit tests of which are described in the following subsections.

[Insert Table 1, Figure 2, and Table 2 here]

**Test of Hypothesis 1**

Hypothesis 1 suggests that congruent measures will be more heavily weighted in \textit{ex ante} weighting decisions than in \textit{ex post} weighting decisions. Given that in our setting there are two measures, one that is congruent but imprecise, and one that is incongruent but precise, the operational hypothesis is that the imprecise/congruent measure will be more heavily weighted in \textit{ex ante} weighting decisions than in \textit{ex post} weighting decisions. To test this hypothesis, we use a planned contrast to compare participants’ weight on the imprecise/congruent measure in the \textit{ex ante} condition to the (\textit{ex post} outcomes unknown) condition. The planned contrast allows us to parse out the effect due to timing from the effect of outcome information in the conditions in which outcomes are known. The mean weight assigned to the imprecise/congruent measure is 56.04 in the \textit{ex ante} condition, and is 46.81 in the \textit{outcomes unknown} condition. The difference of 9.23 is significant \((t = 5.26, p = 0.04, \text{ one-tailed})\). These results support H1.

**Tests of Hypotheses 2a and 2b**

Hypotheses 2a and 2b establish our predictions related to the effect of outcome information on managers’ weighting of multiple measures. Specifically, Hypothesis 2a suggests that the effect of assigning weights \textit{ex post} is \textit{mitigated} when the imprecise/congruent measure

\textsuperscript{11} These results are robust to other contrast codes, including 2, -1, 1, -2 and 6, -2, -1, -3 for \textit{ex ante}, \textit{outcomes unknown}, \textit{favorable imprecise/congruent}, and \textit{favorable precise/incongruent} conditions, respectively.
outcome is relatively favorable. In contrast, Hypothesis 2b suggests that the effect of assigning weights ex post is exacerbated when the imprecise/congruent measure outcome is relatively unfavorable. For both of these hypotheses, our tests use conditions in which weights are determined ex post, thereby controlling for the effect of timing.

Panel B of Table 2 presents results of the relevant planned contrasts underlying these hypotheses. H2a suggests that the mean weight assigned to the imprecise/congruent measure is higher when this measure suggests favorable performance, relative to when outcome information is not available (holding ex post weighting constant). To test this hypothesis, we compare the mean weight assigned to the imprecise/congruent measure in the outcomes unknown condition and the favorable imprecise/congruent condition. With respect to H2a, the mean weight assigned to the imprecise/congruent measure is 46.81 in the outcomes unknown condition, and is 56.95 in the favorable imprecise/congruent condition. The difference of 10.14 is in the predicted direction and is significant (t = 5.58, p = 0.04, one-tailed). This result supports H2a.

Related to this finding, the difference between participants’ weight on the imprecise/congruent measure in the favorable imprecise/congruent condition (mean = 56.95) and in the ex ante condition (mean = 56.04) does not differ (p = 0.88, two-tailed). This pattern of results suggests that in our setting, participants’ decreased focus on the imprecise/congruent measure induced by an ex post perspective is entirely mitigated by the presence of a favorable imprecise/congruent measure outcome (and corresponding unfavorable precise/incongruent measure outcome). More specifically, when the imprecise/congruent measure signaled relatively favorable performance, participants put as much weight on this measure as participants assigning weights ex ante.
H2b suggests that the mean weight assigned to the imprecise/congruent measure is lower when this measure suggests unfavorable performance, relative to when outcome information is not available (holding ex post weighting constant). To test this hypothesis, we compare the mean weight assigned to the imprecise/congruent measure in the outcomes unknown condition and the favorable precise/incongruent condition. With respect to H2b, the mean weight assigned to the imprecise/congruent measure is 46.81 in the outcomes unknown condition, and is 46.11 in the favorable precise/incongruent condition. The difference of 0.70 is not significant (t = 5.66, p = 0.45, one-tailed). This result does not support H2b. While participants in the favorable precise/incongruent condition assigned a lower weight to the imprecise/congruent measure than participants in the ex ante condition (p = 0.04, one-tailed), this difference is no greater than the difference between the outcome unknown and the ex ante conditions. Therefore, the favorable outcome information related to the precise/incongruent measure does not exacerbate the effect of timing on performance measure weighting.

H2a (the mitigation effect) is supported, while H2b (the exacerbation effect) is not. While not consistent with our predictions, upon reflection this pattern of results is not entirely surprising. As demonstrated by the support for H1, with ex post timing (even absent any outcome information), the imprecise/congruent measure is weighted less heavily than with ex ante timing. If this effect is driven by the manager’s desire to protect herself from potential justification pressure, such protection is necessary only to the extent that imprecise/congruent measure turns out to be unfavorable. That is, the H1 effect might be driven by downside risk, in terms of the employee’s outcome (and therefore, in terms of the manager’s justification pressure). If this downside risk is not realized (i.e., the imprecise/congruent measure outcome is favorable), the H1 effect is mitigated, as demonstrated is H2a. However, if the downside risk is
realized (i.e., the imprecise/congruent measure is unfavorable), the reduced weight on the imprecise/congruent measure is sufficient to relieve the manager’s justification pressure. Regardless, the effect we document in our test of H1 – precise measures are more heavily weighted \textit{ex post} than \textit{ex ante} – is to such an extent that it is not exacerbated by further considerations of the favorability of more precise measures.

\textbf{Additional Analyses}

In this section, we report supplemental analyses examining the underlying processes responsible for the overall pattern of results. For this, we rely on answers to the post-experimental questionnaire, in which we ask participants what factors they considered when making their performance measure weighting decisions.\footnote{Prior research shows that individuals have poor self-insight into their own judgment processes (Slovic and Lichtenstein 1971). Therefore, these results should be interpreted with caution.}

Recall that our theory is based on the notion that the timing at which the weighting decision is made will influence the weighting objective, or the immediate reason for engaging in the weighting task. When the weighting is performed \textit{ex ante} (relative to the employee’s action), we predict that the manager’s weighting objective will be to motivate employees to take preferred actions in the upcoming period. In contrast, when the weights are determined \textit{ex post} (relative to the employee’s action), the weighting objective will be to evaluate this prior action, increasing the focus on justifying the weights as fair. To capture these two weighting objectives, we asked participants to answer, on a Likert scale, the following three questions:

1. “To what extent was your decision affected by the need to motivate the quality control manager to engage in desired actions?”

2. “To what extent was your decision affected by the need to ensure the quality control manager receives fair compensation?”
3. “To what extent was your decision affected by how you would justify the decision to the quality control manager?”

The first question is aimed at capturing motivation as the primary weighting objective and questions 2 and 3 are aimed at capturing evaluation (and the resulting focus on justification and fairness) as the primary weighting objective. We calculate two measures for each participant to capture participants’ relative focus on motivation vs. fairness (response on question 1 minus response on question 2) and the relative focus on motivation vs. justification (response on question 1 minus response on question 3). We view these as two alternate measures of participants’ weighting objective, with higher numbers indicating motivation as the primary weighting objective, and lower numbers indicating evaluation as a the primary weighting objective. The measures are included as descriptive statistics in Table 1 and are pictured in Figure 3, Panels A and B.

We expect these measures to follow the same pattern of results as our main dependent variable. Therefore, we use the same contrast coding as in our primary hypothesis tests (3 for ex ante, -1 for outcomes unknown, 1 for favorable imprecise/congruent, and -3 for favorable precise/incongruent). As shown in Figure 3, Panel A, results on the first measure (the relative focus on motivation vs. fairness) do not follow precisely the pattern we predict. We expected the measure to be lower in the favorable precise/incongruent condition than in the outcomes unknown condition, as the pressure for justification and fairness is exacerbated when the precise measure is more favorable than the imprecise measure. Therefore, it is not surprising that the

---

13 At the highest level, one might expect that participants would provide higher answers to the first question in the ex ante condition than in the ex post conditions, and that they would provide lower answers on questions 2 and 3 in the ex ante condition than in the ex post conditions. An examination of the descriptive statistics for these question responses (in Table 1) indicates that this is the general pattern. However, the data are extremely noisy and these patterns are generally not statistically significant (all p > 0.06, one-tailed).
(untabulated) planned contrast is insignificant \((p = 0.38\), one-tailed\). As shown in Figure 3, Panel B, results on the second measure (the relative focus on motivation vs. justification do follow the predicted pattern. Further, the (untabulated) planned contrast is significant \((p = 0.05\), one-tailed\). One interpretation of these findings is that while outcome information can mitigate or exacerbate justification pressure, this desire to appease the subordinate (i.e., to *seem* more fair) is not necessarily accompanied by a desire to actually *be* more fair.

**V. CONCLUSION**

We show, using an experiment, that the timing with which measure weighting occurs in a performance evaluation task affects the weighting outcomes. Specifically, we find that individuals place greater weight on more congruent measures when assigning weights *before* employees exert effort, and place greater weight on more precise measures when assigning weights *after* employees exert effort. Further, when imprecise measures are more favorable *ex post* than precise measures, we find that the decreased focus on imprecise measures *ex post* is mitigated. Supplemental analyses suggest that these effects are driven by managers’ justification pressure being greater in the *ex post* setting.

Our study contributes to academics’ understanding of factors that influence discretionary compensation contracts, and, more specifically, managers’ discretionary weighting of multiple performance measures (Krishnan et al. 2005). In this paper, we have challenged the assumption implicitly made in analytic contracting literature that the timing at which managers determine the weights of multiple performance measures is irrelevant. Our findings suggest otherwise, establishing that the timing at which weights are determined influence managers’ consideration of measures’ attributes, ultimately translating into different assigned weights on multiple performance measures. Notably, this effect is not driven solely by outcome knowledge. Rather,
assigning weights *ex post* to employees’ effort – but without knowledge of the outcomes of such effort – induces an evaluative weighting objective, which leads to an increased focus on the measures’ precision (as opposed to congruence, as when managers assign weights *ex ante*). Our findings related to the effect of timing are also important to practitioners responsible for developing and maintaining a firm’s incentive system. Specifically, our study explains how the timing of managers’ weighting decision (i.e., *ex ante* vs. *ex post*) influences the objective most salient to managers when applying discretion, and thus the weights applied. Such decisions are integral to the effectiveness and efficiency of the firm’s performance measurement, evaluation and compensation system.

Our study is subject to some limitations that provide opportunities for future research. For instance, we operationalize managers’ weighting decisions as a percentage of an employee’s compensation pertaining to a measure, where the percentages sum to 100%. Notably, this operationalization is different from some prior literature (i.e., scalars or multipliers on measures as operationalized by Krishnan et al. (2005)). Future research could investigate the effect of this (and other dimensions) of the actual mechanism via which managers determine and implement multiple measures’ weights.

Other limitations and opportunities relate to our design and setting choices. For example, we model congruence as the extent to which the measure completely captures important actions. A lack of congruence could also be modeled as a measure that is potentially affected by both desirable and undesirable actions taken by an employee, unobservable by the firm. Also, we have not explicitly modeled measures’ differential sensitivity to various employee actions. Future research could investigate the influence of these variations of precision and congruity, as well as managers’ consideration of sensitivity, and its interaction with other measure attributes.
Finally, our experiment does not explicitly model the dynamics of performance measure weighting and effort in a multiple-period scenario. Because our participants were experienced and likely brought with them their knowledge of the multiple-period nature of this business setting, we do not view this as a significant limitation. However, future research could extend our setting to one in which managers repeatedly interact with employees, who, in turn, respond to such compensation decisions with subsequent effort choices. Such a study could be used to better understand the interactive and evolutionary nature of managers’ use of discretion over multiple measures’ weights, as well as the influence of other ecological factors that are beyond the scope of the current study.
References


Figure 1
Experimental Design

The figure below depicts the double-nested design of our study, and the four resulting experimental conditions (underlined). We manipulate the timing of participants’ performance measure weighting decision, the availability of outcome information, and the valence of that information. We manipulate the timing of the participants’ performance measure weighting decision relative to the employee’s action (ex ante or ex post). Nested within the ex post condition, we manipulate the availability of outcome information. In the outcomes unknown condition, participants decide on their weights after the employee has chosen an action, but before learning about any performance measure outcomes. We generate two separate conditions in which outcomes are known by manipulating the valence of outcome information related to the two performance measures. In the favorable precise/incongruent condition, the precise/incongruent measure reflects better performance than the imprecise/congruent measure. In the favorable imprecise/congruent condition, the imprecise/congruent measure reflects better performance than the precise/incongruent measure.
Figure 2
Weight on the Imprecise/Congruent Measure

The figure below depicts the weights participants placed on the imprecise/congruent measure in our four experiment conditions.
Figure 2  
Measures of Weighting Objective

Panel A: Relative Focus on Motivation vs. Fairness (Question 1 minus Question 2)

Panel B: Relative Focus on Motivation vs. Justification (Question 1 minus Question 3)
# Table 1
Descriptive Statistics: Variable Means (Std Dev)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Ex Ante</th>
<th>Outcomes Unknown</th>
<th>Favorable Imprecise/Congruent</th>
<th>Favorable Precise/Incongruent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight on Imprecise/Congruent Measure</td>
<td>56.04 (13.91)</td>
<td>46.81 (20.89)</td>
<td>56.95 (17.79)</td>
<td>46.11 (17.70)</td>
</tr>
<tr>
<td>Weight on Precise/Incongruent Measure</td>
<td>43.96 (13.91)</td>
<td>53.19 (20.89)</td>
<td>43.05 (17.79)</td>
<td>53.89 (17.70)</td>
</tr>
<tr>
<td>PEQ Question 1 (Motivation)</td>
<td>74.79 (15.29)</td>
<td>68.81 (22.80)</td>
<td>73.68 (22.10)</td>
<td>71.39 (23.25)</td>
</tr>
<tr>
<td>PEQ Question 2 (Fairness)</td>
<td>65.63 (20.82)</td>
<td>69.05 (24.01)</td>
<td>70.53 (15.36)</td>
<td>69.17 (18.09)</td>
</tr>
<tr>
<td>PEQ Question 3 (Justification)</td>
<td>51.88 (27.46)</td>
<td>56.90 (24.82)</td>
<td>58.16 (20.56)</td>
<td>64.17 (25.80)</td>
</tr>
<tr>
<td>Relative Focus on Motivation vs. Fairness (Question 1 minus Question 2)</td>
<td>9.17 (24.52)</td>
<td>-0.24 (32.88)</td>
<td>3.16 (25.83)</td>
<td>2.22 (28.61)</td>
</tr>
<tr>
<td>Relative Focus on Motivation vs. Justification (Question 1 minus Question 3)</td>
<td>22.92 (27.30)</td>
<td>11.90 (27.86)</td>
<td>15.53 (28.91)</td>
<td>7.22 (34.78)</td>
</tr>
<tr>
<td>Number of Participants</td>
<td>24</td>
<td>21</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>

---

a This question asked “To what extent was your decision affected by the need to motivate the quality control manager to engage in desired actions?”

b This question asked “To what extent was your decision affected by the need to ensure the quality control manager receives fair compensation?”

c This question asked “To what extent was your decision affected by how you would justify the decision to the quality control manager?”
Table 2
Hypothesis Tests

Panel A: Contrast (Overall)\(^a\)

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Contrast</th>
<th>Std Error</th>
<th>t</th>
<th>p-value(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition(^b)</td>
<td>39.9</td>
<td>17.4</td>
<td>2.30</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Panel B: Planned Contrasts

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Difference</th>
<th>t</th>
<th>p-value(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex Ante vs. Outcomes Unknown</td>
<td>9.23</td>
<td>5.26</td>
<td>0.04</td>
</tr>
<tr>
<td>Hypothesis 2a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes Unknown vs. Favorable Imprecise/Congruent</td>
<td>10.14</td>
<td>5.58</td>
<td>0.04</td>
</tr>
<tr>
<td>Hypothesis 2b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes Unknown vs. Favorable Precise/Incongruent</td>
<td>-0.70</td>
<td>5.66</td>
<td>0.45</td>
</tr>
</tbody>
</table>

\(^a\) Overall contrast using the following weights: 3, -1, 1, -3. Results are also robust to contrast weights of 2, -1, 1, -2 and 6, -2, -1, -3.

\(^b\) Condition represents the ex ante, outcomes unknown, favorable precise/incongruent, and favorable imprecise/congruent.

\(^c\) p-values in the table are one-tailed due to directional predictions.