Contracting on Contemporaneous vs. Forward-Looking Measures: An Experimental Investigation

Anne M. Farrell
Assistant Professor
University of Illinois at Urbana-Champaign
Champaign, Illinois 61820
amf@uiuc.edu

Kathryn Kadous
Associate Professor
Emory University
Atlanta, Georgia 30322
kkadous@emory.edu

Kristy L. Towry
Assistant Professor
Emory University
Atlanta, Georgia 30322
Kristy_Towry@bus.emory.edu

October 2005

This paper has benefited from comments by Ramji Balakrishnan, Jonathan Grenier, Lisa Koonce, Liesel Mitchell, Mark Peecher, Hongmei Zhao, workshop participants at Emory University, and two anonymous reviewers for the 2005 American Accounting Association ABO Section Conference. We thank the Research Computing staff in the Goizueta Business School at Emory University (especially Lisa Harris) for development of the program used in the experiment. We thank the Lost Dog Cafe in Arlington, VA and Zingerman's in Ann Arbor, MI for allowing us to use their creative and extensive sandwich menus in our experimental task.

Contracting on Contemporaneous vs. Forward-Looking Measures: An Experimental Investigation

Abstract

We experimentally examine how employees' employment horizons and the performance measures in their incentive contracts affect employee effort and performance. Economic models suggest that contracts that incorporate leading measures of firm performance mitigate the shortsighted efforts of employees whose employment horizons are not aligned with the firm's profitability horizon, but have less influence on effort choices as employment horizons approach the firm's profitability horizon. We argue that incorporating forward-looking measures in incentive contracts influences employee effort allocation *regardless* of employment horizon because these measures have both decision-influencing and decision-facilitating benefits. Results show that employees with short employment horizons exert more farsighted effort when their incentive contracts incorporate forward-looking measures than when they incorporate only contemporaneous measures, consistent with decision-influencing benefits. However, employees with long employment horizons not only exert more farsighted effort but are also more efficient in task execution with contracts that incorporate forward-looking measures because of decision-influencing benefits.

Keywords: Performance Measurement, Incentives, Effort Allocation, Employment Horizon

Data Availability: Contact the authors.

Contracting on Contemporaneous vs. Forward-Looking Measures: An Experimental Investigation

I. Introduction

The design and implementation of incentive contracts is one of the most difficult challenges facing firm owners and management. When compensation is linked to measures of contemporaneous performance, such as profit or ROI, firms can suffer from shortsighted efforts by employees whose employment horizons are shorter than the profitability horizon of the firm. Suggested remedies for this goal incongruity problem include integrating contract renewal bonuses and stock options into contracts (Indjejikian 1999; Murphy 1999; Jensen, Murphy and Wruck 2004).

The judicious choice of non-traditional accounting performance measures in incentive contracts can also mitigate this problem. Specifically, performance measures such as customer satisfaction and quality are believed to be leading indicators of future financial performance. Recent analytic work relies on the *decision-influencing* role (Demski and Feltham 1976) of accounting information to suggest that optimal contracting weights for forward-looking measures depend on employment horizon. That is, firms can mitigate goal incongruity problems with short-horizon employees by placing weight on forward-looking measures in incentive contracts, because doing so encourages such employees to allocate current-period efforts to actions that increase firm profitability in future periods. However, as employment horizon increases, the optimal contract places less weight on forward-looking measures; because the goals of long-horizon employees are already aligned with those of the firm's owners, contracting on contemporaneous measures is sufficient to ensure that these employees allocate effort to actions that increase future firm profitability (Dikolli 2001).

We argue that placing weight on forward-looking measures in contracts can also play a

decision-facilitating role by providing a partial solution to a *different* problem. Specifically, even if an employee plans to stay at the firm long-term and thus wants to maximize the firm's long-term performance, s/he may not know what day-to-day task strategies will be most effective in doing so. Practitioner literature takes the view that the provision of forward-looking measures helps individuals understand the strategy necessary to obtain long-term performance (Kaplan and Norton 1996, 2001). We extend this view by arguing that linking pay to forward-looking measures increases task strategy understanding to a greater extent than does simply providing the measures (e.g., Cheng and Luckett 2003). Specifically, we expect that incorporating forward-looking measures into the incentive contracts of long-horizon employees more explicitly informs them of the appropriate allocation of efforts across multiple periods and multiple activities, simplifying the cognitive decision task. As a result, cognitive resources that would have been devoted to developing a task strategy can instead be devoted to development of efficiencies in task execution.

In this paper we merge prior literature on the decision-influencing and decision-facilitating roles of forward-looking performance measures to examine the following research questions: (1) how does the employee's employment horizon (short versus long) affect the extent to which the incorporation of forward-looking measures in incentive contracts increases the employee's farsighted efforts?, and (2) for employees with long employment horizons, does the incorporation of forward-looking measures increase the clarity of task strategy and the efficiency with which that strategy is executed?

We conduct an experimental investigation in which the primary independent variables are the employees' employment horizon (long or short) and the incentive contract type (forward-looking or contemporaneous). We examine the effects of employment horizon and contract type

on efforts that increase future performance, efforts that increase contemporaneous performance, and task strategy clarity.

Results suggest that consistent with the decision-influencing role, use of the forward-looking contract as opposed to the contemporaneous contract has a greater effect on the farsighted efforts of short-horizon employees than on those of long-horizon employees. Results are also consistent with the decision-facilitating role. Specifically, results suggest that the forward-looking contract leads to a greater clarity with which long-horizon employees discern the optimal balance between farsighted and contemporaneous efforts in their task strategy. Further, for these long-horizon employees, the forward-looking contract leads to both higher levels of farsighted efforts *and* comparable levels of contemporaneous efforts. In other words, for long-horizon employees, a higher level of farsighted efforts appears to result from gains in efficiency because we do not observe concurrent decreases in contemporaneous efforts. We obtain this result even though employees compensated under both contracts receive identical information about the relationship between forward-looking performance measures and future firm performance, and identical performance reports that include the forward-looking measures.

This research is important to both academic researchers and practitioners. First, analytic research suggests that the effect of including forward-looking performance measures in incentive contracts is dependent on employee employment horizon, but empirical tests of this proposition are limited. We provide such a test and find that contracting on forward-looking measures can add to firm value regardless of employment horizon. Our use of the experimental method allows strong causal inferences regarding the effects of different incentive contracts on performance. For example, prior archival research has examined the factors that influence the contracting weights firms place on different performance measures (e.g., Ittner and Larcker 1995, 2002;

Bushman, Indjejikian and Smith 1996; Ittner, Larcker and Rajan 1997). However, because the contracting weights are endogenous, it is difficult to determine whether differences in firm performance are due to employee effort levels chosen in response to the weights or to the firm-level factors that led to the choice of weights (Ittner and Larcker 2001). In our study, firm characteristics are held constant across experimental conditions, allowing for stronger inferences about the effects of performance measure weights on employee performance.[1]

Second, prior experimental research in psychology, economics, and accounting has examined the effects of various incentive contracts on employee performance (see Bonner and Sprinkle 2002 for a review). However, these studies have generally focused on how different contracts affect the level of effort directed to one distinct action and the subsequent effect on performance.[2] We, on the other hand, provide experimental evidence on how incentive contracts affect the allocation of limited resources to two different actions.

Third, our analysis provides an enhanced understanding of the circumstances under which incentive contracts with forward-looking measures add value. Firms can use this information to better balance the costs of collecting and reporting such measures against the benefits derived from their decision roles. Fourth, the literature advocating the use of strategic performance measurement systems claims that forward-looking performance measures make the actions that maximize firm value more evident to employees, but this literature has wavered on whether there are benefits to linking pay to these measures (e.g., Kaplan and Norton 1996, 2001). We provide evidence that linking pay to these measures can make the optimal effort choices

[1] Other studies examine how psychological factors affect the weights managers place on measures when making performance evaluation judgments of employees after they have exerted effort (e.g., Lipe and Salterio 2000; Ittner, Larcker and Meyer 2003); in contrast, our study examines how weights in incentive contracts affect employees' choice of effort.

[2] An exception is Kelly (2003), who experimentally examines the allocation of resources to capital expenditures versus research and development activities, and demonstrates that the effects of placing weight on non-financial measures depends on the degree to which the firm utilizes tangible versus intangible assets.

clearer to employees than simply providing the measures, and we demonstrate the effects of this clarity on employee performance.

This remainder of this paper is organized as follows. Section II explains the relevant theories and develops our hypotheses. Sections III and IV describe the research design and the results of hypothesis tests. We summarize and conclude in Section V.

II. Theory and Hypothesis Development

Both academic and practitioner literature suggests that traditional financial performance measures (e.g., accounting profits) provide contemporaneous information about employee actions that affect the firm's current financial performance, but little information about actions taken that affect future performance. Therefore, the use of such contemporaneous measures in incentive contracts may motivate employees to take actions that benefit the short-term to the detriment of the long-term. Alternatively, other performance measures (e.g., quality) are believed to be leading indicators of future firm performance. Incorporating such forward-looking measures in incentive contracts can motivate employees to take actions that increase firm performance in the future rather than only the current period (Kaplan and Norton 1996, 2001; Ittner, Larcker and Meyer 2003). In the often-quoted vernacular of Demski and Feltham (1976), this view suggests that forward-looking measures play a decision-influencing role in firms.

Our investigation begins with the intuition provided by Dikolli (2001), who develops an analytic model based on the decision-influencing role of forward-looking measures. The analytic literature on performance measurement has focused largely on the relationship between the sensitivity and precision of performance measures and their weights in incentive contracts (e.g., Holmstrom 1979; Banker and Datar 1989; Feltham and Xie 1994; Hemmer 1996; Lambert

2001), but has generally not examined explicitly how the use of contemporaneous and forward-looking measures influence employees' action choices. Dikolli (2001) extends this literature by formalizing the intuition that the effect of incorporating forward-looking measures (as opposed to only contemporaneous measures) in incentive contracts depends on employees' employment horizon at the firm.

For example, assume that for a particular firm there is a link between product quality in the current period and firm profitability in the future, such that consumers are willing to pay increased prices in the future if the firm offers high quality products today, but prices will decrease in the future if low quality products are offered today. From the firm's perspective, employees who have limited resources maximize long-term performance when they achieve an optimal balance of actions that increase quality and actions that increase quantity. From the employees' perspective, they must choose to allocate effort between two dimensions of their production task, quality of output and quantity of output. Specifically, an employee must decide whether to make more products of lower quality that are nonetheless saleable in the current period (a choice that increases contemporaneous firm financial performance at the expense of future performance), or to make fewer but higher-quality products that will allow the firm to increase prices in the future (a choice that increases future firm financial performance at the expense of contemporaneous performance).[3] Dikolli's (2001) model puts forth propositions about how firm owners can influence an employee's effort choices through the use of forward-looking and contemporaneous measures in incentive contracts, with the weights on each depending on the alignment between the employee's employment horizon and the profitability

[3] For exposition purposes and operational ease, we assume a zero discount rate. This is a reasonable assumption in the laboratory, where all payments are made in cash at the end of the experimental session. Further, this assumption has the inferential advantage of providing an unambiguous economic benchmark against which to compare results (Evans and Moser 2004).

horizon of the firm.

A widely-accepted premise of much of the contracting literature is that the goals of employees and the goals of the firm's owners are misaligned: employees want to maximize short-term compensation, even at the cost of long-term firm performance, while firm owners want to maximize the firm's long-term performance. Dikolli's (2001) model proposes that the judicious use of forward-looking measures in incentive contracts can influence the effort choices of such employees. Specifically, employees with short employment horizons will allocate more effort to actions that increase future firm performance when their incentive contract incorporates forward-looking performance measures than when their contract includes only contemporaneous measures. This is the basis for our first hypothesis (illustrated in Figure 1).

H1: Employees with short employment horizons will allocate more effort to actions that increase future firm performance when the incentive contract incorporates forward-looking performance measures than when it includes only contemporaneous measures (i.e., in Figure 1, (2) > (1)).

Dikolli's (2001) model also proposes that the decision-influencing role of forward-looking performance measures lessens as an employee's employment horizon approaches the profitability horizon of the firm: as the employment horizon of employees increases, their goals more closely match those of the firm's owners. Thus, it is not necessary to explicitly reward long-horizon employees for actions that increase long-term firm performance, because they will be willing to take such actions even if doing so results in lower pay in the short-term. In other words, the effort choices of employees with long employment horizons do not need to be directed via the use of forward-looking performance measures in incentive contracts, because (assuming the contract does include contemporaneous performance measures) employee and firm goals are aligned and employees will take the appropriate actions to meet those goals. This leads to our second hypothesis, which captures the notion that the longer the employment

horizon, the smaller the effect of incorporating forward-looking measures into incentive contracts (illustrated in Figure 1).

H2: The effect of incorporating forward-looking performance measures (as opposed to only contemporaneous measures) into incentive contracts on the allocation of efforts to actions that increase future firm performance will be greater for employees with short employment horizons than for those with long employment horizons (i.e., in Figure 1, (2)-(1) > (4)-(3)).

Together, H1 and H2 predict that incentive contract type interacts with employee employment horizon, such that the incorporation of forward-looking measures in incentive contracts will have a greater effect on the farsighted efforts of short-horizon employees than on the farsighted efforts of long-horizon employees. Prior research on the decision-influencing role of accounting information does not provide us with theory to make an unambiguous prediction about whether the effect of forward-looking contracts on long-horizon employees' farsighted efforts will be positive or simply non-negative. To make this prediction, we rely on the decision-facilitating role of accounting information. In other words, the problem firm owners face with long-horizon employees is not one of goal misalignment, as it is with short-horizon employees. Rather, the problem with long-horizon employees is that while their goal may be to maximize long-term performance, determining what current-period actions will do so is a cognitively difficult task.

We hypothesize that long-horizon employees compensated with contracts that incorporate forward-looking performance measures will devote more effort to actions that benefit future firm performance than will those compensated with contracts that include only contemporaneous measures. Our expectation is that the incorporation of forward-looking measures in incentive contracts can simplify the cognitive processing required to make effort allocation decisions by converting a multi-period problem into a single-period problem, and by making appropriate effort allocation in a given period more salient.

First, it is widely acknowledged that decision-makers often exhibit bounded rationality, particularly when making intertemporal choices (Loewenstein and Elster 1992). Prior theoretical and empirical research finds that in the absence of a clear understanding of the optimal strategy needed to maximize overall performance in a multi-period decision task, individuals tend to over-invest in efforts that maximize current period performance to the detriment of future performance (Herrnstein 1990; Herrnstein and Prelec 1990, 1991; Herrnstein et al. 1993; Kachelmeier and Granof 1993; Mainwaring 1997; Antonides and Maital 2002; Tunney and Shanks 2002).

In our setting, long-horizon employees want to maximize long-term firm performance, and economics-based theoretical research suggests they will try to appropriately allocate effort between different actions to do so (Dikolli 2001). However, maximizing long-term performance is a cognitively difficult task that involves the analysis of the intertemporal costs and benefits of different actions. Basing incentive pay on forward-looking measures rather than solely contemporaneous measures essentially converts the intertemporal decision of how to maximize performance over multiple periods into a series of single-period decisions. That is, assuming that the incentive contract places appropriate weights on forward-looking measures, employees are able to maximize performance across all periods by simply maximizing each current period's compensation. In this way, incorporating forward-looking measures in incentive contracts simplifies the task, thus facilitating the decision-making of employees who wish to maximize long-term performance.

This cognitive task simplification might occur if employees are compensated based on contemporaneous measures and also provided forward-looking measures in accounting reports, even if these forward-looking measures are not the basis of compensation. However, we argue

that the benefits will be greater when pay is linked to forward-looking measures, because the task is simplified even further through the signals about current-period task strategy provided in such an incentive contract. While practitioner literature argues that simply providing forward-looking measures to employees can help them understand the day-to-day actions necessary to achieve higher long-term firm performance (Kaplan and Norton 1996, 2001), several scholars suggest that well-structured incentive plans also inform employees of the appropriate allocation of efforts across multiple activities (Holmstrom and Milgrom 1991; Merchant 1998; Prendergast 1999; Bonner and Sprinkle 2002). This is particularly relevant for employees with long employment horizons, who, as noted, have a more cognitively difficult decision (how to maximize firm performance and thus their pay across multiple periods) than those with short-employment horizons (how to maximize current period pay).

The incorporation of forward-looking measures in incentive contracts makes task strategy more salient because it provides an explicit signal that allocating current-period efforts to actions that improve the forward-looking measure will maximize long-term pay and performance. On the contrary, incentive contracts that include solely contemporaneous measures provide a signal that efforts should be directed to actions that improve contemporaneous performance, which conflicts with long-horizon employees' (and firm owners') goals. Further, providing forward-looking measures without linking them to pay provides only implicit signals about effort allocation; employees are likely to overlook these in favor of explicit signals provided in their incentive contracts. In other words, the use of forward-looking measures in incentive contracts increases the clarity about the direction of efforts employees should choose each period to maximize long-term pay and performance. This frees up cognitive resources that employees would have used for task strategy selection, and allows them to divert these resources to strategy

execution and refinement.

In summary, we expect that for employees with long employment horizons, the incorporation of forward-looking performance measures into incentive contracts plays a decision-facilitating role by converting a multi-period problem into a series of single-period problems, and by easing identification of the task strategy needed to maximize performance over the long horizon. This leads to two predictions, both of which assume that forward-looking performance measures are available to employees regardless of whether the contract incorporates these measures. First, employees with long employment horizons will allocate more effort to actions that increase future performance when their incentive contract incorporates forward-looking measures than when their contract includes only contemporaneous measures (illustrated in Figure 1); and second, they will do so because the actions they must take to maximize future performance are clearer.

H3: Employees with long employment horizons will allocate more effort to actions that increase future firm performance when their incentive contract incorporates forward-looking performance measures than when it includes only contemporaneous measures (i.e., in Figure 1, (4) > (3)).

H4: For employees with long employment horizons, the strategy needed to execute their task will be clearer when their incentive contract incorporates forward-looking performance measures than when it includes only contemporaneous measures.

Finally, for long-horizon employees, if task strategy is indeed clearer when the incentive contract incorporates forward-looking measures, and if these employees do in fact devote less cognitive resources to strategy selection and more to strategy execution and improvement, then it follows that this redirection in cognitive resources should be manifested in higher efficiency in task execution for these employees than for employees whose incentive contract includes only contemporaneous measures. In other words, incorporating forward-looking measures will facilitate decision making by not only directing employee efforts in the optimal direction, but

also enabling employees to increase the strength of those efforts.

Recall our example of a firm in which there is a link between product quality in the current period and firm profitability in the future, such that consumers are willing to pay increased prices in the future if the firm offers high quality products today, but prices will decrease in the future if low quality products are offered today. Assuming that the effect of current quality on future profitability is sufficiently large, the task strategy that optimizes long-term firm performance is one in which employees first direct efforts to actions that optimize quality, and then direct efforts to actions that increase quantity. If this task strategy is clearer to participants, then they will redirect cognitive resources from strategy discovery towards the refinement of its execution. This leads to our last hypothesis.

H5: For employees with long employment horizons, those whose incentive contracts incorporate forward-looking performance measures will be more efficient in the execution of task strategy than will those whose contracts include only contemporaneous measures.

III. Research Design

Participants and Experimental Design

Eighty undergraduate students enrolled in upper-level accounting courses at a large university participated in the experiment. Seventy-four percent of the participants had taken three or more economics courses, and 81% had taken three or more accounting courses; males comprised 57% of the participant pool. Participants were paid based on their performance, as described below. On average, participants earned $25.55.

Participants acted as employees making sandwiches to order for a virtual sandwich shop. The experiment used a 2 x 2 (between-subjects) x 12 (within-subjects) design. The two between-subjects independent variables were employment horizon (*long* or *short*) and incentive contract type (*forward-looking* or *contemporaneous*). The third variable, *work period*, represented the

number of four-minute work periods in the task. Details of the experimental manipulations for the between-subjects variables are in Exhibit 1. Because within-subject differences are not central to this study's research questions, we include supplemental analysis of these effects in Section IV.

For the employment horizon independent variable, we informed each participant in the *long* condition that s/he worked as a sandwich maker for the same sandwich shop through all work periods, and each participant in the *short* condition that s/he for a different sandwich shop in each work period. While all participants worked for 12 four-minute periods, in order to prevent end-of-task gaming, we never told participants the number of work periods.

For the contract type independent variable, each participant in the *contemporaneous* condition was paid 5% of the sandwich shop's revenue generated from the sale of all sandwiches s/he produced in each work period.[4] Each participant in the *forward-looking* condition was paid 5% of the sandwich shop's revenue generated only from the sale of perfect-quality sandwiches s/he produced in each work period (where a 'perfect-quality' sandwich had no mistakes as compared to a customer's order). Note that while the *forward-looking* participants received pay based on a single measure (revenue from perfect sandwiches), this single measure comprises both contemporaneous (revenue) and forward-looking (quality) components. All participants were informed that sandwich shop revenue for each period was computed as the number of saleable sandwiches made (where 'saleable' was defined as a sandwich with fewer than four errors as compared to a customer's order) times that period's selling price per sandwich, and that the selling price per sandwich in the first period was $5.

We informed participants in the *long* employment horizon condition that the per-

[4] We use revenue instead of profitability as the contemporaneous performance measure to simplify the experimental setting. This choice is reasonable if we assume that all costs vary proportionally to revenue.

sandwich price in subsequent work periods was dependent on the average number of mistakes per sandwich in preceding periods. Specifically, if in a given work period the average number of mistakes was exactly zero per sandwich, the sandwich price in the next period would be 10% higher; if the average was greater than zero but less than two mistakes per sandwich, the sandwich price in the next period would remain the same; and if the average was two or more mistakes per sandwich, the sandwich price in the next period would be 10% lower. This pricing function was designed such that the optimal task strategy for maximizing future sandwich shop revenue was to allocate effort primarily to quality, making as many perfect sandwiches as possible, rather than making as many saleable sandwiches as possible. The pricing function parameters were based on an analysis of the task performance of 43 pilot test participants. [5]

In contrast, we informed each participant in the *short* employment horizon condition that the selling price in subsequent periods had been determined in advance and was based on customer demand at the sandwich shop for which s/he worked in a given period. In fact, the prices provided to each participant in this condition were those generated from the work of a participant in the *long* employment horizon condition under the same contract type. This "matched pairs" design equates prices across the *short*- and *long*-horizon conditions, ensuring that price, which is exogenously set in the *short* horizon condition, does not create any inferential difficulties.

Materials and Procedures

We conducted the experiment in a controlled laboratory environment using custom-

[5] Pilot participants did not participate in the main study. To develop the pricing function, we regressed the number of sandwiches completed each period on the number of errors per sandwich, in order to quantify the tradeoff between quantity and quality. We used the resulting estimates to simulate the firm's revenue each period, comparing a high quality versus high quantity strategy. Pricing parameters were set to ensure that the high quality strategy resulted in higher firm revenue. Note that a participant using the high quality strategy would optimally switch to a high quantity strategy when there were only three periods remaining if the number of periods were known; we report analyses to ensure that our results are not driven by participants' anticipation of the end of the experimental task in the results section.

designed software on stand-alone computers. Each computer ran either the *contemporaneous* or the *forward-looking* contract type condition. For both contract types, the software automatically alternated between the *long* and the *short* employment horizon conditions (i.e., the first participant using a particular computer was in the *long* condition, the second in *short*, the third in *long*, etc.). Prices provided to participants in the *short* employment horizon conditions were those generated by the immediately preceding participant in the *long* employment horizon condition, in accordance with the matched pairs design previously described. As participants arrived, we randomly assigned them to seats at the computers and thus randomly assigned them to experimental conditions.

On the first screen of the computerized task, we instructed participants to assume they made their living as sandwich makers, and as such their task was to make sandwiches ordered by a sandwich shop's customers. Customer orders would be transmitted to them via a computerized ordering system.[6] Detailed instructions about the ordering system and the sandwich-making task followed this introduction, and then participants practiced the task during a four-minute work period. A sample task screen is in Exhibit 2.

The sandwich-making task began when a customer's sandwich order appeared in the "Order" window on the computer screen. The participant used a drop-down "Menu" box to find that sandwich from among all those served at the sandwich shop, and then reviewed the sandwich's ingredient list. The ingredient list disappeared when a participant moved the computer's mouse away from the "Menu" box, although participants could review the ingredient list repeatedly. The participant assembled the sandwich by selecting ingredients from five drop-down menus (breads, meats, cheese, vegetables, and condiments), and as s/he did so, images of

[6] For each customer order, the computer program selected from a pool of 51 pre-programmed sandwiches. Sandwich sequences in each work period were pre-determined and the same for each participant.

the selected ingredients appeared in a production space in the center of the screen. Participants could remove ingredients from the production space by using a "Remove ingredient" button. When the participant completed the sandwich, s/he clicked on the "Finished" button. The computer program checked the assembled sandwich against its ingredient list and added it to the participant's production tally. The production space cleared, and the next customer order appeared in the "Order" window.

After the instructional and practice periods, participants read details of their employment relationships with sandwich shops (the employment horizon manipulation), and how their pay would be computed (the contract type manipulation). In all experimental conditions, we instructed participants that they would work as a sandwich maker for several work periods, that they would receive their pay in cash at the end of the session, and that all sandwich shops required that any sandwich with four or more mistakes be thrown away. Such sandwiches did not produce revenue but were also not included in computations of the average number of mistakes per sandwich.

Participants then took a quiz to ensure they understood how prices were determined and how revenue and pay were computed. Participants could not begin the sandwich-making task until they answered all quiz questions correctly, and read reinforcement explanations of the answers even if they answered questions correctly. Successful completion of the quiz and subsequent completion of the task provided assurance that participants understood their employment relationship with the sandwich shops and how their pay would be computed (i.e., that our manipulations were successful). Quiz questions and answers are in Exhibit 3.

Participants then worked for 12 four-minute work periods. At the end of each period, a feedback screen displayed participant pay, sandwich shop revenue, the number of sandwiches

made with zero, one, two, three, or four or more mistakes for the current period, and the per-sandwich price in the upcoming period. Note that across all experimental conditions, we provided identical production quality and quantity information at the end of each work period; thus, even participants in the *contemporaneous* contract type condition had forward-looking performance measures available, which biases against finding our predictions in Hypotheses 3, 4, and 5. After the twelfth work period, participants completed a post-experimental questionnaire.

Dependent Variables

For tests of Hypotheses 1 through 3, we use two proxies for employee efforts to increase future firm performance, which in our setting were actions that resulted in higher sandwich quality. Our primary measure of quality efforts is *average errors per sandwich*, computed as the simple average of the total number of errors made each period (exclusive of sandwiches thrown away) divided by the number of saleable sandwiches made; lower values for *average errors per sandwich* suggest higher levels of quality efforts.[7] This is a comprehensive measure of quality effort because it captures all the various actions employees could take to increase quality. For example, an action participants could have taken to produce high quality sandwiches was to spend more time viewing the sandwich menu in order to make sure that s/he correctly encoded the customer's order. Participants could also have concentrated harder while constructing the sandwiches in order to prevent inadvertent errors. The *average errors per sandwich* measure captures these and other types of quality effort.

Our second dependent measure of quality efforts is *menu time per sandwich*, computed as the simple average of the number of seconds spent viewing the sandwich menu each period divided by the number of saleable sandwiches completed; higher values for *menu time per sandwich* suggest higher levels of quality efforts. This is a more direct but less comprehensive

[7] The dependent measures for Hypotheses 1, 2, 3 and 5 were automatically captured by the software used in the task.

measure of quality effort, in that it captures only one type of effort by employees – effort directed at correctly encoding the ingredients required for each order.

To test Hypothesis 4, we developed a measure of task strategy clarity using four questions from the post-experimental questionnaire, all of which were answered using a Likert scale ranging from zero (strongly disagree) to ten (strongly agree):

1) The strategy I needed to use to complete the sandwich-making task was very clear and specific; I knew exactly how much to focus on quality vs. quantity to receive the highest possible pay.
2) I had a hard time choosing how much effort to devote to making high quality sandwiches vs. making a large number of sandwiches.
3) I was certain how to perform the sandwich-making task in order to earn the highest possible pay.
4) The strategy I needed to use to do my job was obvious.

In a confirmatory factor analysis, the eigenvalue for the first factor was 2.69 while all other eigenvalues were less than one. The factor loadings for the questions (in the order listed above) were 0.87, -0.65, 0.88, and 0.86. Cronbach's alpha, which measures the reliability of the construct, was 0.82, above the generally accepted cutoff of 0.80.[8] These analyses suggest that the four questions define a single task *strategy clarity* construct. We use the factor score as the dependent measure for Hypothesis 4.

Finally, to test Hypothesis 5, we couple an analysis of *average errors per sandwich* with an analysis of *sandwich quantity*, measured as the average number of saleable sandwiches produced across periods.

IV. Results

Tests of Hypotheses 1 through 3 - Quality Efforts

Recall that we base Hypotheses 1 through 3 on two roles of forward-looking performance

[8] The response to the second question was reverse coded in the calculation of Cronbach's alpha.

measures. In predicting the effort directed toward quality, the decision-influencing view predicts a main effect of contract type when employment horizon is short (H1), as well as an interaction between contract type and employment horizon (H2). The decision-facilitating view predicts a main effect for contract type when employment horizon is long (H3). Descriptive statistics for the dependent variables used to test Hypotheses 1 through 3, and for all other dependent variables, are in Table 1.[9]

Results of the Kolmogorov-Smirnov test show significant departures from normality in the *average errors per sandwich* variable *(*$Z = 2.75$*,* $p < 0.01$*)*, caused by the clustering of data points just to the right of zero. As suggested by Conover and Iman (1981), we address this issue by conducting our analyses using a rank transformation of the variable. Unless otherwise noted, results are inferentially identical with raw and rank values.

Table 2, Panel A provides results of an ANOVA, with contract type and employment horizon as the independent variables and the rank of *average errors per sandwich* as the dependent variable. This analysis shows a marginally significant interaction of employment horizon and contract type *(*$F_{1,76} = 3.34$, $p = 0.07$); when the raw *average errors per sandwich* measure is used as the dependent variable, this interaction is highly significant ($F_{1,76} = 25.14$, $p < 0.01$). This result suggests that the effect of the forward-looking contract on quality efforts depends on employment horizon, which provides support for H2. Simple effects analysis (Table 2, Panel B) indicates that when employment horizon is short, *average errors per sandwich* is significantly lower with the forward-looking contract than with the contemporaneous contract *(*$F_{1, 76} = 21.36$, $p < 0.01$), providing support for H1. Further, when employment horizon is long, *average errors per sandwich* is significantly lower with the forward-looking contract than with

[9] Recall that lower values for *average errors per sandwich* suggest higher levels of quality efforts, while higher values of *menu time per sandwich* suggest higher levels of quality efforts.

the contemporaneous contract ($F_{1,76} = 4.14$, $p = 0.02$), providing support for H3; when the raw *average errors per sandwich* measure is used rather than its rank, this effect is marginally significant ($F_{1,76} = 2.08$, $p < 0.06$).

This pattern of results suggests that the use of forward-looking measures in incentive contracts has both a decision-influencing and a decision-facilitating effect in our setting. That is, we find that the effect of compensating participants with the forward-looking contract (in contrast to the contemporaneous contract) on quality efforts is greater for short-horizon participants than for long-horizon participants, providing evidence of the interaction predicted by the decision-influencing role. Further, even for long-horizon participants, the forward-looking contract results in significantly higher quality efforts, providing evidence of the decision-facilitating role. Therefore, we find that for all participants, contracting on quality creates a shift in effort toward high quality task strategies.[10]

Table 3, Panel A provides results of an ANOVA with *menu time per sandwich* as the dependent variable. Results show a significant interaction between employment horizon and contract type ($F_{1\ 76} = 8.13$, $p < 0.01$), which indicates that the effect of the forward-looking contract on quality efforts depends on employment horizon, providing support for H2. Simple effects analysis (Table 3, Panel B) shows that when employment horizon is short, *menu time per sandwich* is significantly higher with the forward-looking contract than with the contemporaneous contract ($F_{1,76} = 12.69$, $p < 0.01$), providing support for H1. However, when the employment horizon is long, *menu time per sandwich* does not differ under the forward-

[10] Given the parameters used in our experiment, a participant using a high-quality strategy would optimally switch to a high-quantity strategy when there were three work periods remaining. However, participants were not told the number of work periods and so could not make this determination. Further, we found no evidence that the number of errors per sandwich increased in later periods, which suggests that participants did not switch strategies in anticipation of the final period. Lastly, we conducted each of our analyses using only periods 1 through 2, 1 through 3, . . . , and 1 through 11; except where otherwise noted, results were inferentially identical to those reported.

looking and contemporaneous contracts ($F_{1\ 76} = 0.22$, $p = 0.68$), which does not provide support for H3.

When viewed in isolation, the results using *menu time per sandwich* seem to suggest the absence of a decision-facilitating role for contracting on forward-looking measures in incentive contracts. However, we argue that another interpretation is more plausible. Recall that *menu time per sandwich* is a less comprehensive measure of quality efforts than *average errors per sandwich*; the former captures only the time participants spent on the order encoding process, while the latter captures a broad range of actions that participants could have taken to increase quality. This suggests that for short-horizon participants, the higher level of quality efforts resulting from the forward-looking contract resulted at least in part by participants taking more time in the order-encoding process.[11] For long-horizon participants, however, time spent on the order-encoding process is not significantly different with the forward-looking and contemporaneous contracts, so the significantly-higher level of a broad range of quality efforts must have resulted from other actions, such as concentrating harder while constructing the sandwiches. This interpretation is consistent with the fact that with the contemporaneous contract, long-horizon participants produced relatively high-quality sandwiches (particularly when compared to participants with short horizons compensated with the same contract). To achieve still-higher quality than these participants, long-horizon participants compensated with the forward-looking contract would have to employ a broader range of quality efforts.

Tests of Hypotheses 4 and 5 - Task Strategy Clarity and Efficiency

Hypothesis 4 predicts that when employment horizon is long, task strategy clarity will be higher with the forward-looking contract than with the contemporaneous contract. Recall that participants in all experimental conditions received feedback on quality after each work period;

[11] Formal mediation analysis (not tabulated) confirms this intuition.

the difference between our contemporaneous and forward-looking contract conditions was whether participants received this feedback for informational purposes only or also received pay based on the quality measure.

To test H4, we compare our *strategy clarity* measure across contract types for participants in the long employment horizon. Simple effects analysis (Table 4, Panel B) indicates that *strategy clarity* is significantly higher in the forward-looking contract condition than in the contemporaneous contract condition *($F_{1,72}$ = 2.65, p = 0.05)*, providing support for H4.

For informational purposes, we report the results of an ANOVA with employment horizon and contract type as the independent variables and *strategy clarity* as the dependent variable. The interaction of employment horizon and contract type is marginally significant ($F_{1,72}$ = 3.31, p = 0.08), which indicates that the effect of the forward-looking contract on *strategy clarity* depends on employment horizon. Further simple effects analysis (Table 4, Panel B) shows that for short-horizon participants, *strategy clarity* does not differ across contract type *($F_{1,72}$ = 0.79, p = 0.80)*; strategy clarity is relatively high for short-horizon participants regardless of contract type (see Table 1). Interestingly, further analysis shows that with the contemporaneous contract, *strategy clarity* is significantly higher for participants in the short horizon condition than for those in the long horizon condition *($F_{1,72}$ = 14.54, p < 0.01)*. Taken together, these results confirm our reasoning in the development of Hypotheses 3 and 4 that long-horizon employees are less clear about what task strategies will maximize their pay than are short-horizon employees.

Hypothesis 5 predicts that for long-horizon participants, the efficiency with which participants execute the sandwich-making task will be higher with the forward-looking contract than with the contemporaneous contract. To test H5, we couple our prior results for *average*

errors per sandwich, our comprehensive measure of quality efforts, with an analysis of *sandwich quantity*.

First, recall that our theory and design assumed that individuals have limited resources to allocate across tasks, so allocating efforts to actions that increase long-term firm performance (i.e., quality) must come at the expense of actions that increase contemporaneous performance (i.e., quantity). We gain some insight into the validity of this assumption by comparing the quality and quantity efforts of short-horizon employees, to see if, in this simpler single-period maximization task, a tradeoff between quality and quantity exists. Table 5, Panel A shows results of an ANOVA with employment horizon and contract type as independent variables and the rank transformation of *sandwich quantity* as the dependent variable. The interaction is significant ($F_{1,76} = 8.67$, $p < 0.01$), indicating that the effect of the forward-looking contract on *sandwich quantity* depends on employment horizon.[12] Simple effects analysis (Table 5, Panel B) demonstrates that for short-horizon participants, *sandwich quantity* is lower with the forward-looking contract than with the contemporaneous contract ($F_{1,76} = 15.30$, $p < 0.01$). Further, recall that our results for H1 indicated that for short-horizon participants, quality efforts are higher with the forward-looking contract than with the contemporaneous contract. While we acknowledge that our experimental manipulations played a large part in these differences, it also appears that participants who adopted a high-quality task strategy were not able to reach the output levels of those who adopted a high-quantity strategy. Thus, differences in performance across contract types in the short-horizon condition may result, at least in part, by a tradeoff between types of effort – participants allocated *more effort to quality* and *less to quantity* under the forward-looking contract, and vice versa for the contemporaneous contract. This provides

[12] The Kolmogorov-Smirnov test finds significant departures from normality for the sandwich quantity variable ($Z = 1.95$, $p < 0.01$), so we use a rank-transformation of sandwich quantity in our analysis; results are inferentially identical when the analysis is conducted using the raw variable.

some support for our assumption.

Second, for long-horizon employees, we can infer efficiency gains for participants in the forward-looking contract condition if we show that their (1) quality efforts (as measured by *average errors per sandwich*) are higher than, and (2) *sandwich quantity* is higher than or equal to, those of participants in the contemporaneous contract condition. Our tests of Hypothesis 3 confirmed that long-horizon participants in the forward-looking contract condition allocated more effort to quality than did those in the contemporaneous contract condition, providing support for the first requirement. Simple effects analysis for *sandwich quantity* (Table 5, Panel B) for long-horizon participants indicates that quantity does not differ across the forward-looking and contemporaneous contract conditions ($F_{1,76} = 0.06$, $p = 0.40$), providing support for the second requirement. Taken together, these analyses provide support for H5, in that participants in the long horizon condition compensated with the forward-looking contract not only achieved higher quality than those compensated with the contemporaneous contract; they also simultaneously maintained output levels.

Supplemental Analysis – Trend Analysis

In this section, we provide analysis of the period-by-period trends in each of our dependent variables. As noted earlier, these within-subjects effects are not central to the study's research question, but an examination of these trends may provide additional insights into the effects of our experimental manipulations on participant performance. Figure 2 illustrates the trends in *average errors per sandwich*, *menu time per sandwich*, and *sandwich quantity*.[13] We have identified four possible reasons for differences in performance across work periods:

[13] As described earlier, *average errors per sandwich* and *sandwich quantity* exhibited significant departures from normality so we performed our hypothesis tests using rank-transformed variables. Because it is not possible to conduct trend analysis using rank-transformed variables and results of hypothesis tests were inferentially the same when raw variables were used instead, we rely on raw variables for this trend analysis.

differences in task difficulty, end-of-game effects, learning (task), and learning (strategy). We discuss each of these in turn.

Differences in task difficulty occur across periods because the program generated a different set of sandwich orders each period. For example, in one period, the first sandwich order might require only three ingredients, whereas in another period, the first order might require seven ingredients. These period-by-period differences would not result in any trend, but in seemingly random variations. In fact, these variations are not random, because the sandwich sequences in each work period were pre-determined and the same for each participant. Thus, variations due to differences in task difficulty across periods should be similar across conditions. Inspection of Figure 2 confirms this assumption. For example, note that in period seven across all four conditions, there is an upward variation in *menu time per sandwich*, and a downward variation in *quantity*. This pattern suggests that the sequence of orders that period included more difficult sandwiches than in other periods.

Period-by-period differences could also occur if participants changed their behavior in anticipation of the end of the session. As described earlier, we attempted to prevent such behavior by not informing participants of the number of work periods. However, participants' beliefs about the number of periods could have still potentially affected behavior. End-of-game behavior would be evidenced by changes in trends over time. The patterns of results in Figure 2 show no obvious changes in trends over time, and thus, we conclude that end-of-game behavior is not a significant driver of performance.

Two other reasons for period-by-period differences relate to learning. First, with practice, participants may have become more adept at performing the sandwich-making task, leading to fewer errors and/or greater quantity in later periods. Second, through experimentation,

participants may have refined their task strategy choice, and depending on whether they ultimately choose a high quality or a high quantity strategy, this refinement could have led to higher or lower quality and lower or higher quantity in later periods. It is difficult, if not impossible, to disentangle the effects of these two types of learning. Therefore, we restrict our analysis to a general exploration of the trends in the four experimental conditions. For this analysis, we used the statistical tool of orthogonal polynomial contrasts, with the assumption that any trends were linear in nature. Specifically, for each dependent variable we conduct a three-way ANOVA (results not tabulated), with two between-subjects variables (contract type and employment horizon) and one within-subjects variable (work period). We use the orthogonal polynomial contrast to test the null hypothesis that the dependent variables exhibited no linear trend over the 12 work periods against the alternate hypothesis that either a positive or negative trend existed. This method also allows us to test whether the trend differed by experimental condition.

The three-way ANOVA using *average errors per sandwich,* our primary proxy for quality effort, as the dependent variable finds a significant three-way interaction *($F_{1,76}$ = 7.62, p < 0.01)*, meaning that significant differences in trend occurred across the four experimental conditions. Further analysis shows that when employment horizon is long, no linear trend exists in *average errors per sandwich* ($F_{1,76}$ = 0.009, p = 0.92), and this finding did not vary across the two contract types ($F_{1,76}$ = 0.010, p = 0.90). However, for short-horizon participants, the linear trend in *average errors per sandwich* depends on contract type ($F_{1,76}$ = 17.87, p < 0.01): there is no linear trend for the forward-looking contract ($F_{1,76}$ = 0.009, p = 0.92), but the trend is significantly positive for the contemporaneous contract ($F_{1,76}$ = 31.25, p < 0.01). We interpret this trend as providing evidence that participants in this condition were continuing to "push the

boundary" in terms of the quality of their output. Recall that sandwiches with four or more mistakes were thrown out. Note that participants in the short horizon contemporaneous condition made an average of only 1.3 errors per sandwich in the first work period, and eventually increased to 2.2 by the final period. Thus, it appears that this trend was consistent with task strategy refinement.

The three-way ANOVA for *menu time per sandwich*, our second proxy for quality effort, provides evidence of a negative linear trend ($F_{1,76} = 37.93$, $p < 0.01$), suggesting that participants put less effort into quality over the course of the experiment. Neither of the other two variables interacted with *period*, meaning that the trend does not differ by experimental condition. In the short-horizon − contemporaneous contract condition, this trend may have been driven by a task strategy refinement process, as these participants became more convinced that their best strategy was to focus on quantity. However, the fact that the trend does not differ by condition suggests that it was primarily the result of participants becoming more adept at the sandwich-making task over time.

The three-way ANOVA for the *quantity* variable shows a significant three-way interaction, suggesting that significant differences in trend occur across the four experimental conditions ($F_{1,76} = 13.79$, $p < 0.01$). Follow-up analysis finds that for long-horizon participants, the trend in *quantity* is positive ($F_{1,76} = 41.08$, $p < 0.01$) and does not vary by contract type ($F_{1,76} = 0.54$, $p = 0.47$). For short-horizon participants, while the trend in *quantity* is positive for participants working under both contemporaneous ($F_{1,76} = 21.34$, $p < 0.01$) and forward-looking ($F_{1,76} = 27.24$, $p < 0.01$) contracts, a significant period by contract type interaction ($F_{1,76} = 35.36$, $p < 0.01$) indicates that the trend is significantly more positive for those working under a contemporaneous contract. We interpret the positive *quantity* trend across all periods as

indicative that participants became more adept at the sandwich-making task over the course of the experiment. For short-horizon participants, we interpret the incremental positive trend for those working under a contemporaneous contract as evidence of task strategy refinement, consistent with our conclusion regarding trends in *average errors per sandwich*.

V. Conclusions

This study provides evidence that the effect of contracting on forward-looking measures will depend on the employment horizon. For short-horizon employees, forward-looking measures play a decision-influencing role, in that incorporating these measures in incentive contracts aligns the employee's goals with those of the firm's owners and directs their efforts to actions that will increase long-term firm performance, thus mitigating the shortsightedness that occurs when compensation is based solely on contemporaneous measures. Because the goals of long-horizon employees are already aligned with those of the firm's owners, our results confirm the notion that incorporating forward-looking measures in incentive contracts has a smaller effect on the farsighted efforts of these employees. Importantly, though, incorporating forward-looking measures *does* lead to a significant increase in the farsighted efforts of even long-horizon employees, because forward-looking measures also play a decision-facilitating role. They simplify a multi-period decision and increase the clarity with which employees identify the optimal task strategy. We argue that this increased strategy clarity frees up cognitive resources and thus leads to improvements in efficiency, allowing for increases in quality without concurrent decreases in quantity. Thus, while managerial accountants generally view incentive contracts as playing a decision-influencing role in firms, our results suggest that they can also play a significant decision-facilitating role, improving the performance even of those employees whose goals are more closely aligned with the firm's long-term interests.

While our results provide support for both the decision-influencing and decision-facilitating roles of forward-looking performance measures, our study has several limitations, each of which presents an opportunity for future research. The form of the performance measure might influence results. We operationalized the forward-looking contract by paying participants a percentage of the revenue from all perfect sandwiches. Thus, while the contract includes both contemporaneous (revenue) and forward-looking (quality) features, these features are combined into one measure. Future research could address whether incorporating separate quantity and quality measures would lead to even greater task strategy clarity, as well as addressing the more general question of whether and how the composition of the performance measures in incentive contracts affects the cognitive processes leading to effective strategy development and execution.

Our findings also suggest extensions to basic psychology research into choice over time. While our experimental participants completed 12 periods of production, these periods were compressed into one experimental session. We expect that over a longer period of time results would be similar but hypotheses would be even more strongly supported. Future work could test this expectation.

Finally, our experimental examination occurs in a world without uncertainty. Firm owners know with certainty quality's effect on future prices, so they can determine the optimal task strategy and create an incentive system to induce employees to engage in that strategy; employees know the length of their relationship with the firm and the exact basis of their compensation for the entire term of that relationship. These abstractions are important for experimental control, but each presents a potential moderating variable for the effects presented here. We encourage future research on the moderating effects of various types of uncertainty on the relationships among performance measurement, incentive compensation, and performance.

References

Antonides, G. and S. Maital. 2002. Effects of feedback and educational training on maximization in choice tasks: Experimental-game evidence. *Journal of Socio-Economics* 31(2): 155-165.

Banker, R. and S. Datar. 1989. Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research* 27(1): 21-39.

Bonner, S. and G. Sprinkle. 2002. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society* 27(4/5): 303-345.

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

Cheng, M. and P. Luckett. 2003. The use of multiple measures in performance reports and task-related performance. Working Paper. The University of New South Wales, Sydney, Australia.

Conover, W. and R. Iman. 1981. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician* 35(3): 124-129.

Demski, J. and G. Feltham. 1976. *Cost Determination: A Conceptual Approach*. Ames, IA: The Iowa State University Press.

Dikolli, S. 2001. Agent employment horizons and the contracting demand for forward-looking performance measures. *Journal of Accounting Research* 39(3): 481-494.

Evans, J. and D. Moser. 2004. Agency theory and incentive contracting experiments. Working paper. University of Pittsburgh, Pittsburgh, PA.

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

Hemmer, T. 1996. On the design and choice of "modern" management accounting measures. *Journal of Management Accounting Research* 8: 87-116.

Herrnstein, R. 1990. Rational choice theory: Necessary but not sufficient. *American Psychologist* 45(3): 356-367.

Herrnstein, R., G. Loewenstein, D. Prelec, and W. Vaughan. 1993. Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making* 6(3): 149-185.

Herrnstein, R. and D. Prelec. 1990. Melioration. In G. F. Loewenstein and J. Elster (Eds.),

Choice Over Time, 235-263. New York, NY: Russell Sage Foundation.

Herrnstein, R. and D. Prelec. 1991. Melioration: A theory of distributed choice. *The Journal of Economic Perspectives* 5(3): 137-156.

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

Holmstrom, B. and P. Milgrom. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization* 7: 24-52.

Indjejikian, R. 1999. Performance evaluation and compensation research: An agency perspective. *Accounting Horizons* 13(2): 147-157.

Ittner, C. and D. Larcker. 1995. Total quality management and the choice of information and reward systems. *Journal of Accounting Research* 33(Supplement): 1-34.

Ittner, C. and D. Larcker. 2001. Assessing empirical research in managerial accounting: A value-based management perspective. *Journal of Accounting and Economics* 32(1-3): 349-410.

Ittner, C. and D. Larcker. 2002. Determinants of performance measure choices in worker incentive plans. *Journal of Labor Economics* 20(2): S58-S90.

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

Ittner, C., D. Larcker, and M. Rajan. 1997. The choice of performance measures in annual bonus contracts. *The Accounting Review* 72(2): 231-255.

Jensen, M., K. Murphy, and E. Wruck. 2004. Remuneration: Where we've been, how we got to here, what are the problems, and how to fix them. Harvard Business School Negotiation, Organizations and Markets Working Paper No. 04-28. Harvard University, Boston, MA.

Kachelmeier, S. and M. Granof. 1993. Depreciation and capital investment decisions: Experimental evidence in a governmental setting. *Journal of Accounting and Public Policy* 12(4): 291-323.

Kaplan, R. and D. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Kaplan, R. and D. Norton. 2001. *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment.* Boston, MA: Harvard Business School Press.

Kelly, K. 2003. Does adding nonfinancial value drivers to a summary financial measure improve

the learning and performance of managers? Working Paper. Nanyang Business School, Nanyang Technological University, Singapore.

Lambert, R. 2001. Contracting theory and accounting. Working Paper. University of Pennsylvania, Philadelphia, PA.

Lipe, M. and S. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75(3): 283-298.

Loewenstein, G. and J. Elster. 1992. *Choice Over Time*. New York, NY: Russell Sage Foundation.

Mainwaring, L. 1997. Maximisation and melioration as alternative forms of firm behavior. *Journal of Economic Behavior and Organization* 32(3): 395-411.

Merchant, K. 1998. *Modern Management Control Systems*. Upper Saddle River, NJ: Prentice Hall.

Murphy, K. 1999. Executive compensation. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Vol. 3*, Chapter 38. Amsterdam, The Netherlands: Elsevier Science B.V.

Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37(1): 7-63.

Tunney, R. and D. Shanks. 2002. A re-examination of melioration and rational choice. *Journal of Behavioral Decision Making* 15(4): 291.

EXHIBIT 1
Experimental Materials (Included Text for Either the *Long* or *Short* Employment Horizon and Either the *Contemporaneous* or *Forward-Looking* Contract)

Now that you have experience with the sandwich-making job and the computerized ordering system, you will work as a sandwich maker for several work periods.

Long Employment Horizon: You have signed a contractual agreement to work with the same sandwich shop for **all of these work periods.** Your pay each period is based on the revenue the sandwich shop earns.

Short Employment Horizon: You have signed a series of contractual agreements, and you will **work with a different and unrelated sandwich shop for each of these work periods.** In other words, you will work with a given sandwich shop for one and only one work period, after which you will work with a different shop. You will never work with the same sandwich shop twice, and all contractual agreements have been signed and cannot be changed or cancelled for any reason. Your pay each period is based on the revenue the sandwich shop earns.

Contemporaneous Contract: Specifically, your pay will equal 5% of the total revenue you generate for the sandwich shop in each work period. **You will receive this pay in real cash at the end of today's session!**

Forward-looking Contract: Specifically, your pay will equal 5% of the total revenue that the shop earns from perfect sandwiches (i.e., sandwiches with exactly zero (0) mistakes) that you produce. **You will receive this pay in real cash at the end of today's session!**

Long Employment Horizon: The sandwich shop's revenue for a given period is computed as the number of saleable sandwiches you make times the selling price for that period. The selling price at this sandwich shop is $5.00 in the first work period. In each subsequent period, the price at this shop depends on the average number of mistakes for all sandwiches produced in the immediately-preceding period, because this affects customer demand for the sandwich shop's sandwiches. Specifically, if in a given work period there is an average of:

- exactly zero (0) mistakes per sandwich, the sandwich price in the **next** period will be *10% higher*;
- greater than zero (0) but less than two (2) mistakes per sandwich, the sandwich price in the **next** period will *remain the same*;
- two (2) or more mistakes per sandwich, the sandwich price in the **next** period will be *10% lower*.

Before the start of each work period, you will be shown the new per-sandwich price for the sandwich shop.

Short Employment Horizon: A sandwich shop's revenue for a given period is computed as the number of saleable sandwiches you make times the selling price for that period. The selling price at your first sandwich shop is $5.00. The selling price at each of the subsequent sandwich shops you will work with has been determined in advance, and is based on customer demand for that sandwich shop's sandwiches. Before the start of each work period, you will be given the per-sandwich price for the sandwich shop you have contracted with in that period.

Finally, sandwich shops in this market require that any sandwich with four (4) mistakes or more when compared to a customer order be thrown away. Therefore, sandwiches with four (4) or more mistakes do not produce any revenue. The mistakes in sandwiches thrown away are not included in the computation of average number of mistakes.

EXHIBIT 2
Sample Computer Screen from Experimental Task

Each participant, acting as sandwich maker, completed the experimental task on screens like the one shown. A customer's sandwich order (selected from a population of 51 sandwiches) appeared in the "Order" box at the top right of the screen. The participant used the "Menu" drop-down box to find the sandwich from among all those served at the sandwich shop, and then reviewed the sandwich's ingredient list (which disappeared when the computer's mouse was moved away from the "Menu" box). As the participant assembled the sandwich using the drop-down ingredient menus on the left side of the screen, images of the ingredients appeared in the production space in the center; ingredients could be removed using the "Remove ingredient" button at the top center of the screen. When the sandwich was complete, the participant clicked on the "Finished" button at the bottom center of the screen; the sandwich was automatically checked by the computer program and added to the participant's production tally, the production space was cleared, and the next customer order appeared in the "Order" box.

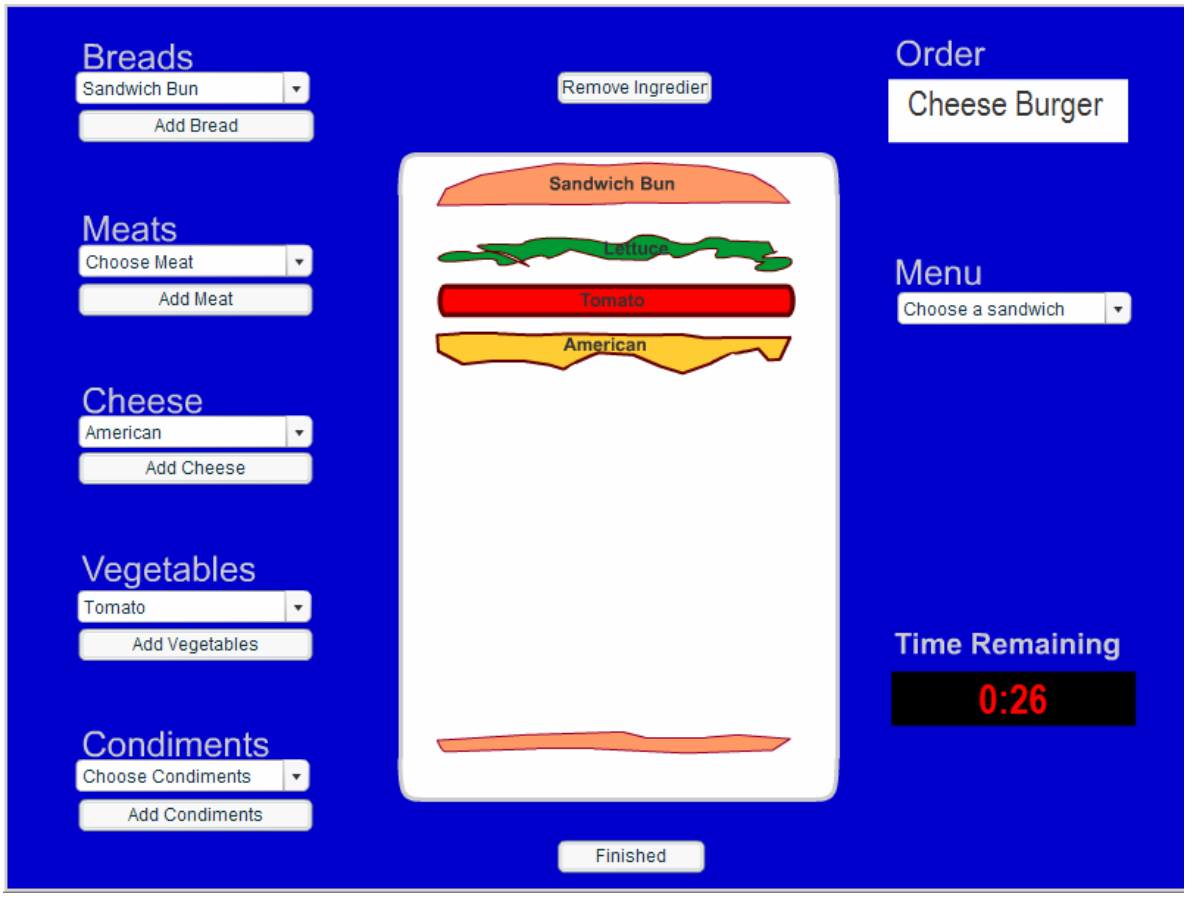


EXHIBIT 3
Quiz and Answers Used in Experimental Materials

Participants had to answer all quiz questions correctly before they could begin the sandwich-making task, and had to read reinforcement explanations of correct answers even if they had answered questions correctly.

Panel A – Quiz for All Experimental Conditions

Please answer the following questions before going on. You may review the material on your screen to help you answer the questions. You must get these answers correct before moving on.

1. If the selling price of sandwiches is $5.00 in the first work period, and you make 6 sandwiches with an average of 0 mistakes per sandwich (that is, no mistakes at all in any sandwiches) in the first period, what is the selling price of sandwiches in the second period?
   a. $4.50
   b. $5.00
   c. $5.50
   d. The price cannot be determined from this information.

2. If the selling price of sandwiches is $5.00 in the first work period, and you make 2 sandwiches with an average of 0.5 mistakes per sandwich in the first period, what is the selling price of sandwiches in the second period?
   a. $4.50
   b. $5.00
   c. $5.50
   d. The price cannot be determined from this information.

3. If the selling price of sandwiches is $5.00 in the first work period, and you make 3 sandwiches with an average of 1.67 mistakes per sandwich in the first period, how much revenue does the sandwich shop earn in that period?
   a. $0.75
   b. $1.75
   c. $15.00
   d. $16.50

4. If the selling price of sandwiches is $5.00 in the first work period, and you make 7 sandwiches with an average of 2.1 mistakes per sandwich in the first period, how much do you earn in that period? (Of the sandwiches, 2 are perfect and the other 5 each have 3 mistakes.)
   a. $0.50
   b. $1.58
   c. $1.75
   d. $35.00

EXHIBIT 3 (continued)

Panel B - Answers to Quiz for Each Experimental Condition

Experimental Condition		**Correct Answer to Question No.:**			
Employment Horizon	**Contract Type**	**1**	**2**	**3**	**4**
short	contemporaneous	d	d	c	c
short	forward-looking	d	d	c	a
long	contemporaneous	c	b	c	c
long	forward-looking	c	b	c	a

TABLE 1
Descriptive Statistics for All Dependent Variables

PANEL A - Means (Standard Deviations) of Dependent Variables by Experimental Condition

Dependent Variable	Employment Horizon			
	Long horizon		Short horizon	
	Contemporaneous Contract	Forward-looking Contract	Contemporaneous Contract	Forward-looking Contract
Average Errors Per Sandwich	0.37 *(0.60)*	0.10 *(0.09)*	1.82 *(1.01)*	0.21 *(0.18)*
Menu Time Per Sandwich	8.73 *(3.85)*	8.27 *(3.09)*	4.64 *(2.50)*	8.18 *(2.96)*
Strategy Clarity (factor score)	-0.55 *(1.05)*	-0.13 *(0.99)*	0.53 *(0.82)*	0.26 *(0.77)*
Sandwich Quantity	5.76 *(1.89)*	5.26 *(1.02)*	9.35 *(3.43)*	5.45 *(1.11)*
N	20	20	20	20

Average Errors Per Sandwich is computed as the simple average of the total number of errors made each period (exclusive of sandwiches thrown away) divided by the number of saleable sandwiches made in that period.

Menu Time Per Sandwich is computed as the simple average of the number of seconds spent viewing the sandwich menu each period divided by the number of saleable sandwiches made in that period.

Strategy clarity is the composite factor score of the following questions from the post-experimental questionnaire, all of which were answered using a Likert scale of zero (strongly disagree) to ten (strongly agree):

1) The strategy I needed to use to complete the sandwich-making task was very clear and specific; I knew exactly how much to focus on quality vs. quantity to receive the highest possible pay.
2) I had a hard time choosing how much effort to devote to making high quality sandwiches vs. making a large number of sandwiches.
3) I was certain how to perform the sandwich-making task in order to earn the highest possible pay.
4) The strategy I needed to use to do my job was obvious.

Sandwich Quantity is computed as the average number of saleable sandwiches produced across periods.

TABLE 2
ANOVA Results – Rank of Average Errors per Sandwich

Average errors per sandwich is computed as the simple average of the total number of errors made each period (exclusive of sandwiches thrown away) divided by the number of saleable sandwiches made in that period. All observations are then rank-ordered, and the rank-ordering is used for tests.

Because participants completed the task for multiple work periods, work period was a within-subjects variable. However, because differences across work periods are not of primary interest in this study, results are aggregated across periods and only between-subjects effects are reported in these tables.

Because hypotheses include directional predictions for contract type, *p*-values for *Contract Type* are reported on a one-tailed basis; all other *p*-values are two-tailed.

PANEL A – Results of ANOVA

Source	**df**	**Mean Square**	***F***	***p***
Employment Horizon	1	0.31	0.00	0.96
Contract Type	1	2365.31	22.16	< 0.01
Employment Horizon X Contract Type	1	357.01	3.34	0.07
Error	76	106.76		

PANEL B – Simple Effects of Contract Type on Average Errors per Sandwich

Source	**df**	**Mean Square**	***F***	***p***
Long employment horizon:				
Contract Type	1	442.23	4.14	0.02
Error	76	106.76		
Short employment horizon:				
Contract Type	1	2280.10	21.36	< 0.01
Error	76	106.76		

TABLE 3
ANOVA Results – Menu Time per Sandwich

Menu time per sandwich is computed as the simple average of the number of seconds spent viewing the sandwich menu each period divided by the number of saleable sandwiches made in that period.

Because participants completed the task for multiple work periods, work period was a within-subjects variable. However, because differences across work periods are not of primary interest in this study, results are aggregated across periods and only between-subjects effects are reported in these tables.

Because hypotheses include directional predictions for contract type, *p*-values for *Contract Type* are reported on a one-tailed basis; all other *p*-values are two-tailed.

PANEL A – Results of ANOVA

Source	df	Mean Square	*F*	*p*
Employment Horizon	1	87.65	8.89	< 0.01
Contract Type	1	47.19	4.79	0.03
Employment Horizon X Contract Type	1	80.12	8.13	< 0.01
Error	76	9.86		

PANEL B – Simple Effects of Contract Type on Menu Time per Sandwich

Source	df	Mean Square	*F*	*p*
Long employment horizon:				
Contract Type	1	2.17	0.22	0.68
Error	76	9.86		
Short employment horizon:				
Contract Type	1	125.15	12.69	< 0.01
Error	76	9.86		

TABLE 4
ANOVA Results – Strategy Clarity

Strategy clarity is the composite factor score of the following questions from the post-experimental questionnaire, all of which were answered using a Likert scale of zero (strongly disagree) to ten (strongly agree):

1) The strategy I needed to use to complete the sandwich-making task was very clear and specific; I knew exactly how much to focus on quality vs. quantity to receive the highest possible pay.
2) I had a hard time choosing how much effort to devote to making high quality sandwiches vs. making a large number of sandwiches.
3) I was certain how to perform the sandwich-making task in order to earn the highest possible pay.
4) The strategy I needed to use to do my job was obvious.

Because hypotheses include directional predictions for contract type, *p*-values for *Contract Type* are reported on a one-tailed basis; all other *p*-values are two-tailed.

PANEL A – Results of ANOVA

Source	**df**	**Mean Square**	***F***	***p***
Employment Horizon	1	11.13	13.16	< 0.01
Contract Type	1	0.21	0.25	0.31
Employment Horizon X Contract Type	1	2.65	3.31	0.08
Error	72	0.85		

PANEL B – Simple Effects of Contract Type on Strategy Clarity

Source	**df**	**Mean Square**	***F***	***p***
Long employment horizon:				
Contract Type	1	2.24	2.65	0.05
Error	72	0.85		
Short employment horizon:				
Contract Type	1	0.66	0.79	0.80
Error	72	0.85		

TABLE 5
ANOVA Results – Rank of Sandwich Quantity

Sandwich quantity is computed as the average number of saleable sandwiches produced across periods. All observations are then rank-ordered, and the rank-ordering is used for tests.

Because participants completed the task for multiple work periods, work period was a within-subjects variable. However, because differences across work periods are not of primary interest in this study, results are aggregated across periods and only between-subjects effects are reported in these tables.

Because hypotheses include directional predictions for contract type, *p*-values for *Contract Type* are reported on a one-tailed basis; all other *p*-values are two-tailed.

PANEL A – Results of ANOVA

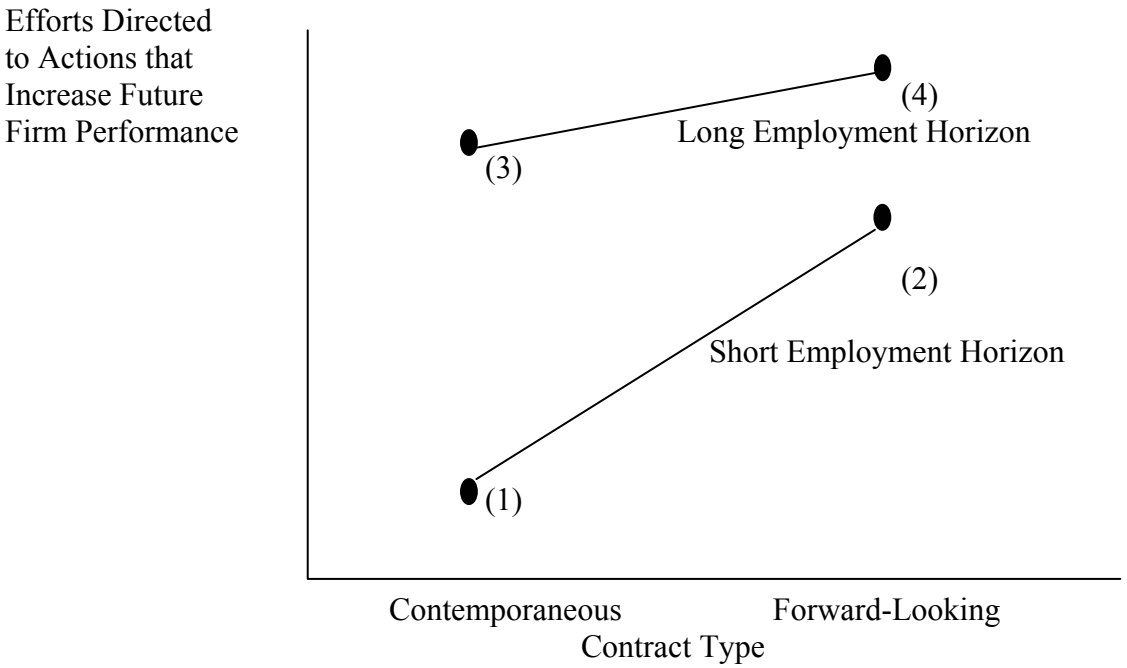
Source	**df**	**Mean Square**	***F***	***p***
Employment Horizon	1	10.88	0.09	0.76
Contract Type	1	803.28	6.69	0.01
Employment Horizon X Contract Type	1	1040.40	8.67	< 0.01
Error	76	119.99		

PANEL B – Simple Effects of Contract Type on Sandwich Quantity

Source	**df**	**Mean Square**	***F***	***p***
Long employment horizon:				
Contract Type	1	7.66	0.06	0.40
Error	76	119.99		
Short employment horizon:				
Contract Type	1	1836.03	15.30	< 0.01
Error	76	119.99		

FIGURE 1
Predictions for Hypotheses 1 Through 3

PANEL A – Graph of Predictions in Hypotheses 1 through 3



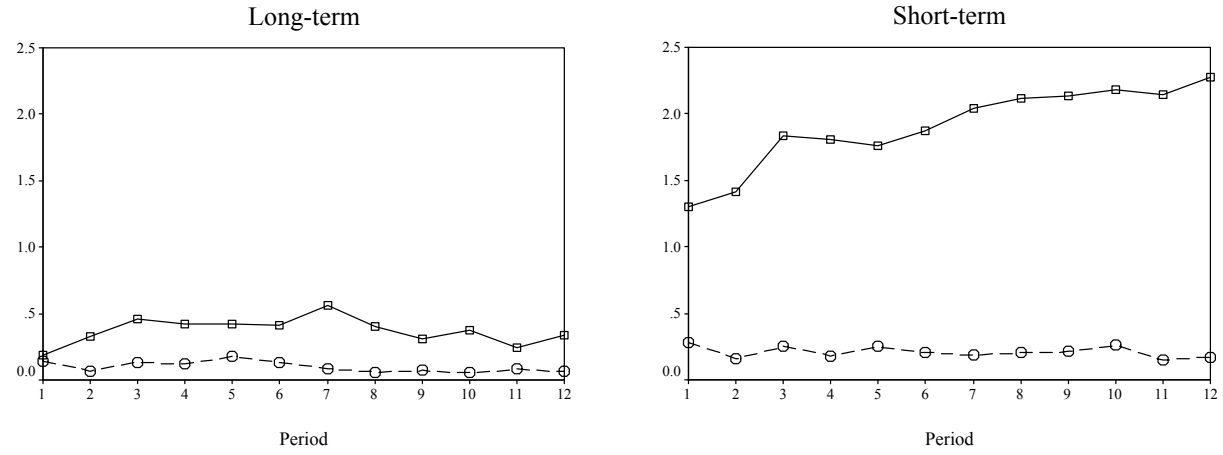
Efforts directed to actions that increase future firm performance are measured by *average errors per sandwich* and *menu time per sandwich*. *Average errors per sandwich* is computed as the simple average of the total number of errors made each period (exclusive of sandwiches thrown away) divided by the number of saleable sandwiches made in that period. *Menu time per sandwich* is computed as the simple average of the number of seconds spent viewing the sandwich menu each period divided by the number of saleable sandwiches made in that period.

Predictions and results for Hypotheses 1 through 3 are as follows:

Hypothesis	Prediction	Results For:	
		Average Errors per Sandwich	Menu Time per Sandwich
H1	(2) > (1)	supported	supported
H2	(2) - (1) > (4) - (3)	supported	supported
H3	(4) > (3)	supported	not supported

FIGURE 2
Work Period Trends

Panel A – Average Errors per Sandwich



Panel B – Menu Time per Sandwich

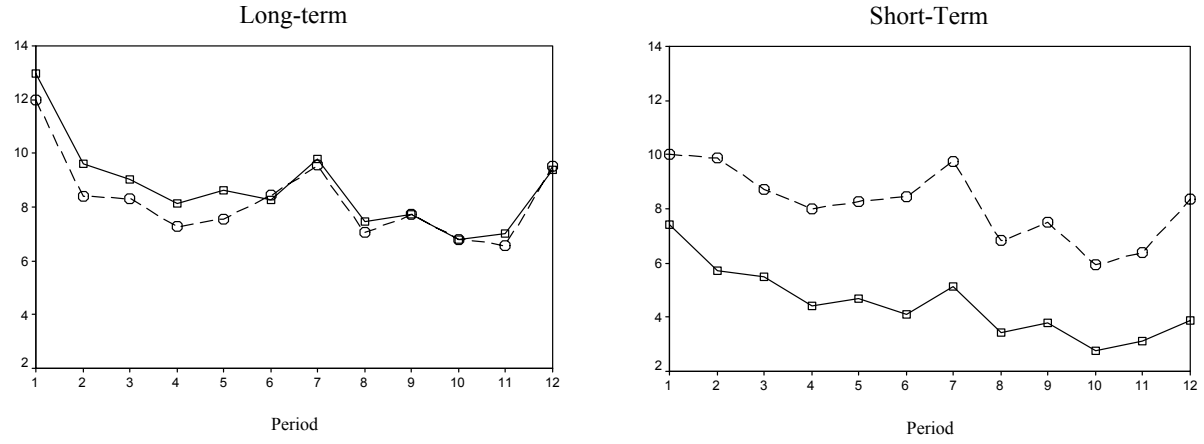
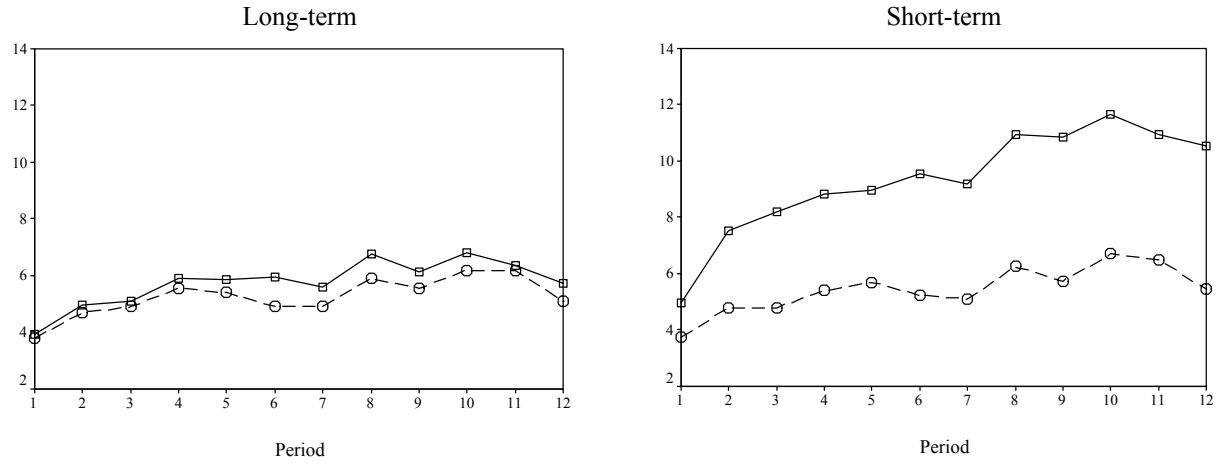


FIGURE 2 (continued)

Panel C – Sandwich Quantity



_____ Contemporaneous contract type condition
------- Forward-looking contract type condition

Average errors per sandwich is computed as the simple average of the total number of errors made each period (exclusive of sandwiches thrown away) divided by the number of saleable sandwiches made in that period.

Menu time per sandwich is computed as the simple average of the number of seconds spent viewing the sandwich menu each period divided by the number of saleable sandwiches made in that period.

Sandwich quantity is computed as the average number of saleable sandwiches produced across periods.
