Performance Measure Properties and Incentive System Design

MICHAEL J. GIBBS, KENNETH A. MERCHANT,
WIM A. VAN DER STEDE, and MARK E. VARGUS*

We analyze effects of performance measure properties (controllable and uncontrollable risk, distortion, and manipulation) on incentive plan design, using data from auto dealership manager incentive systems. Dealerships put the most weight on measures that are "better" with respect to these properties. Additional measures are more likely to be used for a second or third bonus if they can mitigate distortion or manipulation in the first performance measure. Implicit incentives are used to provide ex post evaluation, to motivate the employee to use controllable risk on behalf of the firm, and to deter manipulation of performance measures. Overall, our results indicate that firms use incentive systems of multiple performance measures, incentive instruments, and implicit evaluation and rewards as a response to weaknesses in available performance measures.

Introduction

PERFORMANCE MEASUREMENT IS PERHAPS THE MOST DIFFICULT CHALLENGE in the design and implementation of incentive systems. Since explicit measures are affected by factors outside the employee's control, they impose risk on the employee. The firm may narrow the focus of evaluation to reduce risk (e.g.,

* The authors' affiliations are, respectively, Graduate School of Business, University of Chicago, Chicago, IL 60637 and Institute for the Study of Labor (IZA); Marshall School of Business, University of Southern California, Los Angeles, CA 90089; London School of Economics and Political Science, London WC2A 2AE; School of Management, University of Texas–Dallas, Dallas, TX 75083. E-mail: *mike.gibbs@chicagogsb.edu*. We are grateful to an unnamed consulting firm for giving us access to their data and clients, and for discussions that helped us understand the auto dealership business and clarify the data. For comments on various drafts, we thank Trond Petersen (the editor), anonymous referees, Jan Bouwens, Mark Bradshaw, Jim Brickley, Jed DeVaro, Leslie Eldenburg, Eva Labro, Joan Luft, Margaret Meyer, Kevin J. Murphy, Walter Oi, Lorenzo Patelli, Canice Prendergast, Michael Raith, Edward Reidl, Bernard Salanié, Marcel van Rinsum, and Sally Widener; seminar participants at the Harvard Business School, London School of Economics, RSM Erasmus University, Tilburg University, Universidad de Navarra, Universitat Pompeu Fabra, University of Aarhus, University of Arizona, University of Rochester, University of Southern California; and conference participants of AAA, BMAS, CAED, CEPR, and Society of Labor Economists. Liu Zheng provided helpful research assistance.

INDUSTRIAL RELATIONS, Vol. 48, No. 2 (April 2009).
Published by Blackwell Publishing, Inc., 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford, OX4 2DQ, UK.

use accounting numbers instead of stock price to evaluate a CEO), but that often results in distorted incentives. In addition, the employee may be able to use private knowledge to manipulate the measure to increase pay without improving firm value. In response to these problems, the firm may add subjectivity to the incentive system, by using explicit measures as inputs into implicit incentives (such as promotion decisions), or by using subjective evaluations as a substitute for explicit measures. However, discretion raises its own concerns, such as the potential for favoritism and bias.

Consistent with their importance in practice, performance measure problems have received increasing attention in agency theory. The original models (e.g., Banker and Datar 1989; Holmstrom 1979) emphasized uncontrollable risk (noise). Later models incorporated multitask incentives (Holmstrom and Milgrom 1991), which motivated formal consideration of distortions and manipulation (Baker 1992; Demski, Frimor, and Sappington 2004; Feltham and Xie 1994). Recent work has emphasized the extent to which the agent can or cannot respond to risk (Baker 2002; Prendergast 2002; Raith 2008). In accounting, the empirical literature analyzing performance measure properties focuses largely on risk or distortion (Bushman, Indjejikian, and Smith 1996; Ittner, Larcker, and Rajan 1997; Ittner and Larcker 2002; Van Praag and Cools 2001). There is also a large literature on manipulation at the level of corporate earnings, and a smaller literature on manipulation at lower levels of the organization (e.g., Holthausen, Larcker, and Sloan 1995). Finally, a smaller literature studies subjectivity in evaluation and rewards (Campbell 2008; Gibbs et al. 2004; Hayes and Schaefer 2000; MacLeod and Parent 1999; Murphy and Oyer 2003). Despite the importance of performance measurement, the empirical literature on performance evaluation is surprisingly small.

This paper contributes to this literature on performance measurement by providing analysis of several parts of the puzzle together. We constructed a unique data set on the entire incentive system for a set of managers in auto dealerships. This allows direct measurement and study of three major performance measure properties: risk (both uncontrollable and controllable), distortions, and manipulation. We show how these properties affect both explicit and implicit incentives. We then study how different incentive instruments are related to each other, a question that has received little attention. Finally, the data provide evidence on how incentive system design takes into account firm strategic variables (degree of competition and emphasis on customer satisfaction). Putting all of this together provides a more comprehensive view of incentive system design than has previously been possible.

Our findings are briefly summarized as follows. First, dealerships put the most weight on measures that have the “best” properties in terms of risk, distortion, and manipulation among those available. This reinforces the existing empirical literature on performance measure properties.

Second, firms use additional bonuses in part to adjust for weaknesses in the performance measure given the most weight. Many dealerships offer a second or third bonus based on different measures. We find that the magnitude of additional bonuses is a function of its performance measure properties (such as distortion) relative to those of the performance measure used for the largest bonus. Thus, multiple bonuses appear to be used to rebalance multitask incentives.

Third, we provide some of the first empirical evidence on the distinction between controllable and uncontrollable risk. Performance measures with more uncontrollable risk are given less weight for incentives, a finding that has been elusive in prior research. In addition, our evidence suggests that incentive system design accounts for the employee's private information or controllable risk in two ways. One is to encourage employees to respond productively to changes in their environment. The other is to reduce incentives to use such information to manipulate performance measures. These are both done in part through implicit rewards granted based on ex post judgments of performance.

Put together, these results suggest two conclusions: performance measure properties are important to both the strength and balancing of incentives, and incentive plans are a system of interrelated instruments, explicit and implicit, that are designed to work together.

Predictions

In this section we develop our predictions. We begin with standard predictions from the theoretical literature on properties of a single performance measure. We then present several other predictions that are either new or little studied in the existing literature. Those predictions arise from our core idea: when a performance measure is flawed in some way, and no better single measure is available, the firm may move to a *system* of multiple instruments to provide better overall incentives. We consider two ways in which a system of incentives can improve on a flawed performance measure. The firm might use additional bonuses on other performance measures, or ex post settling up through implicit incentives or discretionary bonuses.

We use the following terminology. *Performance measure* refers to a quantitative indicator such as accounting profits or number of cars sold. *Formula bonus* (FB) refers to a bonus that is calculated using a mathematical formula based on a performance measure. In our setting, we distinguish up to three FBs, each using only a single performance measure. Both the measure and the formula are set ex ante. Formula bonuses are distinguished from *discretionary bonuses*, which are determined by supervisor judgment. *Implicit incentives* refer to rewards other than discretionary bonuses that are awarded using judgment. Discretionary bonuses and implicit incentives may use numeric performance measures as inputs, but the supervisor may also use qualitative performance information, and may also use judgment in the weights applied to measures. Implicit incentives include the manager's autonomy, raises, promotions, and possible termination. In contrast to FBs, discretion in incentive systems requires ex post judgment.

Predictions Based on Properties of a Single Performance Measure. The literature on key properties of a single performance measure is well known. Performance measures may have uncontrollable risk (noise), which raises costs since agents are risk averse (Banker and Datar 1989; Holmstrom 1979). Measures may also be distorted because their weight misallocates the agent's efforts on different tasks (Baker 1992, 2002; Feltham and Xie 1994; Holmstrom and Milgrom 1991; Van Praag and Cools 2001). The standard predictions are that incentives should be weaker, the greater the noise or more distorted the measure. Several studies analyze the effects of noise on incentives (e.g., Ittner et al. 1997; Ittner and Larcker 2002; see the survey by Prendergast 1999), but this literature has mixed findings. A much smaller literature has examined the effects of distortion on incentives (e.g., Bouwens and van Lent 2006).

Prendergast (2002) suggests that the mixed findings about the effect of noise on incentive intensity stem from failure to consider an additional performance measure property: the employee's knowledge that arises while performing the job. The literature has not yet settled on a term for this concept. Jensen and Meckling (1992) and Raith (2008) call it "specific knowledge"; Baker and Jorgensen (2003) use the term "volatility"; and Shi (2008) refers to "respondable risk." We suggest the terms (long used in the behavioral accounting literature) "controllable risk" and "uncontrollable risk." The advantage of these terms is to highlight two types of risk. Uncontrollable risk is one that the agent cannot react to (noise). Controllable risk is environmental uncertainty that the agent can react to (after observing a signal about the state of the world). To the extent that the employee has such knowledge, incentives should be stronger to motivate the employee to use that knowledge to increase firm value. For example, if gasoline prices rise unexpectedly, the new car sales department might change its emphasis toward selling more fuel efficient cars. Recent empirical evidence using data that distinguish between controllable and uncontrollable risk, which most prior work has been unable to, is consistent with Prendergast's prediction (e.g., Bouwens and van Lent 2008; DeVaro and Kurtulus 2006).

A final performance measure property that has received less attention is manipulability (Courty and Marschke 2004; 2008; Demski et al. 2004; Healy 1985). Manipulation occurs if the agent "games" the incentive plan to increase the reward without increasing (or at the expense of) firm value. The effects of manipulation are similar to the effects of distortion, except that with manipulation the employee uses his or her specific knowledge to increase measured performance in ways that are not consistent with firm value. This distinction is useful because a firm may use different methods to address distortion and manipulation. We return to this point below.

Summing up the discussion above, standard agency theory leads to the following predictions:

1. The incentive intensity on a bonus will be decreasing in the performance measure's noise, distortion, and manipulability. It will be increasing in the measure's controllable risk.

Predictions about Systems of Incentives. Firms often use a system of multiple incentives. An agent may be offered more than one bonus on different measures. Sometimes firms offer bonuses based on discretionary performance evaluation. Firms also use implicit incentives, such as promotion or threat of termination. If a performance measure has no flaws, why use additional incentive instruments or performance measures? Therefore, we argue that additional incentives could be used to mitigate weaknesses in a single incentive based on a single performance measure.

Implicit Incentives. Implicit incentives differ from explicit incentives in an important way: they are based on the principal's ex post evaluation of performance (Gibbs et al. 2004). Supervisors often observe their own signals about the agent's performance. Typically these will be based more on outputs than inputs, so they are not the same information that the agent observes (the agent's controllable risk). These signals can improve the evaluation along the line of Holmstrom's (1979) informativeness principle. They allow the principal to revise incentives based on information that arose after the contract was set with the aim to improve overall incentives. Such ex post settling up is important, because if anticipated it affects the agent's ex ante incentives and perceived risk (Baker, Gibbons, and Murphy 1994). For example, the principal might use ex post information to filter out some of the noise from the performance measure, such as by rewarding the agent more if there was bad luck. Implicit evaluations can also be used to reduce distortion (by rebalancing ex ante incentives when employees anticipate an ex post adjustment). They can also be used to reduce manipulation, as we describe below.

Of course, such signals may not be formally contractable, in part because they are not also observed by the agent. In such cases the firm must use relational contracting, in which the employee must trust the evaluator to be reasonably fair in reporting the evaluation. This is a drawback to this approach to evaluation. The fact that most jobs appear to use discretion in evaluations and implicit tying of rewards to evaluations suggests that the benefits of implicit incentives often outweigh the costs.

Specifically, we examine two possible roles of ex post evaluation, focusing on implicit incentives. First, in addition to filtering out effects of uncontrollable risk, discretion might be used to encourage the employee to respond to controllable risk to improve firm value. For example, the supervisor might evaluate the extent to which the employee took initiative quickly and effectively reacting to events as they unfolded in performing the job. This would be impossible to foresee ex ante. Therefore, we predict that:

2. Implicit rewards will be more strongly related to a performance measure the more important is controllable risk in the measure.

Second, and similarly, the principal might also use ex post evaluation to mitigate manipulation. Manipulation is caused by the employee's knowledge arising while performing the job, and thus, arises from information ex post to setting the contract. Therefore, we predict that:

3. Implicit rewards will be more strongly related to a performance measure, the more manipulable is the measure.

Summarizing these two predictions and the distinction between them, implicit rewards are expected to be used to reward the employee for exploiting controllable risk to improve firm value, or to punish manipulation if it is detected ex post.

Multiple Bonuses. The second way in which a firm might improve incentives based on an imperfect performance measure is to add additional bonuses based on other measures with different properties (Feltham and Xie 1994; Hemmer 1996). Additional measures can reduce risk to the extent that they are not perfectly correlated with the first measure. They can reduce distortion if one measure gives relatively strong emphasis to one dimension of performance and another gives relatively less. Baker (2002) shows that when a second measure is used in an incentive system, the weight is a decreasing function of uncontrollable risk and distortion *relative to* the other measure. For example, if one performance measure does not give enough emphasis to cooperation, the firm might give a second bonus based on a different performance measure that is relatively better at rewarding cooperation. More generally, the idea is that the added measures reduce noise, distortion, or manipulation. The incentive systems that we study often use more than one bonus. We predict that:

4. If the firm uses multiple bonuses, the additional measures will be given greater weight if their properties are *relatively* better.

To our knowledge, our second, third, and fourth predictions have never been tested. We now describe the data used in this study. The data set is new, uses survey data, and is unusually comprehensive. For these reasons, we provide more description than is typical. The descriptive part is designed to provide information on the entire incentive system, something about which little has been previously published.

Data

Survey, Features, and Limitations. A boutique auto dealership consulting firm allowed us to design and implement a survey on incentive practices of their clients. We thus had the opportunity to collect data on variables that are usually not available to academics. Our survey methodology has positive and negative features. To our knowledge, it provides the most detailed information ever collected on the system of incentives, explicit and implicit, used within firms. However, survey data have down sides (Bertrand and Mullainathan 2001). They tend to be noisy; by nature, much of the information is perceptual and difficult to quantify. This may lead to attenuation bias in coefficient estimates. Such data can, however, shed light on questions that are otherwise difficult or impossible to study with more traditional, publicly disclosed data sets.

Before developing the survey, we spent a day at a large dealership interviewing the owner and department managers. This acquainted us with the business, job designs, incentive issues, and language they use. In addition, the consulting firm surveyed its clients on incentive practices several years before the project.[1] We used these sources to develop our survey. The initial version was discussed with the firm's professionals. A revised version was pilot-tested at twenty-four dealerships before the survey was finalized.

[1] The older surveys were not used to consult with dealerships on incentive plan design. The company does not recommend organizational practices to clients. It provides benchmarking studies that assess a dealership against others.

We developed surveys for the owner, general manager, and managers of the service, new car sales, and used car sales departments. The owner survey asked about ownership, bonus payments, and demographics. The general manager survey asked about the dealership's competitive environment, strategy, and management practices. The department surveys were largely identical except for relevant word substitutions. The most important section of those surveys asked detailed questions about salary, bonuses, performance measures, bonus formulas, and subjective evaluations. Outside the compensation section, the surveys principally contained five-point Likert scales. Of these, we use two multi-item scales to assess the degree of competition and emphasis on customer service (see Appendix).

We mailed the final set of five surveys to 1203 dealerships, along with our cover letter and one from the consulting firm stating their support for the study. We sent a reminder letter to nonparticipants after 4 weeks. Six weeks after that, we did a telephone follow-up to dealerships from which we had received at least one survey.[2] We received 1057 surveys, or 18 percent of those mailed. A few were not useful, most commonly because they had substantial missing data. Of the 185 new car department respondents, 39 combined new and used car sales in the same department. We have at least 1 survey from 326 different dealerships, or 27 percent.[3] We found no evidence of sample selection bias on the basis of performance, size, geography, or manufacturer.

Our study follows the recent trend toward industry studies (e.g., Ichniowski, Shaw, and Prennushi 1997). Industry studies have several virtues. Because we had good knowledge of the jobs respondents worked in, we were able to write questions that fit the context. Furthermore, by holding industry constant, much variation is controlled for. In this industry, all firms have essentially the same organizational structures (except that some combine new and used car sales into one department), with essentially the same job designs for general and department managers across dealerships. Our main focus, performance measurement, is similar for all firms sampled. These features of the sample should reduce measurement error, which is particularly important with survey data. Of course, a weakness of industry studies, including this one, is that it is difficult to gauge how generalizable the findings are.

[2] The response rate is probably lower for department managers for two reasons. First, we sent the package of surveys to the general manager. In some cases a survey may not have been passed to a department manager. Second, a few managers may have worried (incorrectly) that their responses would be seen by the general manager or owner.

[3] As some surveys were partially incomplete, sample sizes vary slightly across various tables.

A potential weakness of this study is that we use cross-sectional rather than panel data. It is possible that some of our findings are driven by unobservable heterogeneity across firms. As noted above, a virtue of an industry study is that many variables that might drive such heterogeneity simply do not vary much here because the firms are so similar. Nevertheless, because of this concern we analyzed whether any of our results might be driven by variables other than those included in the tables below, including the region, name plate of car, and whether the dealer sold luxury or economy cars. We found no evidence that these factors had any effect. In addition, we analyzed whether survey variables might be correlated with personal characteristics of managers, but found no evidence of this. These validity checks give us reasonable assurance that our findings are not primarily driven by unobserved heterogeneity.

An interesting question is what kinds of unobservable heterogeneity might drive differences in performance measures, and performance measure properties, across dealerships. The literature on performance measurement provides little guidance. Presumably the job design for the manager is one factor, including the type and size of the department. Controls for those were included in all regressions. The quality of the manager's staff could matter as well. Unfortunately we have no information on this. The competitive environment could be a factor as well: whether the dealership is in a city, suburb, or rural area; number of other dealerships nearby (especially those that sell similar cars); demographics of potential customers, etc. We controlled for several of these, where we had data, including a measure of local competition for the dealership. For implicit incentives, the experience of the supervisor (general manager) and department manager might be relevant. We included controls for the experience of the department manager, but did not have information beyond that.

Descriptive Statistics. Compensation plans for managers are set by dealership owners, not auto manufacturers, generally once per year. Table 1 provides summary statistics. Since these are all privately held firms, managers in our sample are not compensated through the use of stock or options. Pay systems in this industry have three major components: salary, FBs, and discretionary bonuses. Salary averages a bit less than half of the total pay. In the two types of sales departments, roughly 10 percent of managers are paid zero base salary. Compared to most industries, pay for performance is a very large part of compensation for managers in this industry.

The most important component of pay for performance is FBs. In our sample, managers were eligible for up to three bonuses calculated as explicit functions of specific performance measures. We defined these as FBs 1, 2,

TABLE 1

SUMMARY STATISTICS

			General manager	Department managers: New	Used	Service
a. Department characteristics						
GMs who are owners			26%	–	–	–
New/Used combined			–	24%	–	–
# direct reports			22.5	17.0	11.0	29.2
Years of industry experience			20.9	15.6	17.1	23.2
N			250	194	127	205
b. Manager's compensation						
Total compensation			$191,749	$81,892	$81,149	$65,755
% receiving	Salary		98%	88%	89%	94%
	Formula bonus	1	65%	58%	59%	64%
		2	10%	25%	25%	24%
		3	4%	11%	10%	10%
	Discretionary bonus		20%	24%	24%	20%
% eligible	Formula bonus	1	72%	85%	81%	85%
		2	14%	36%	33%	39%
		3	4%	19%	16%	19%
$ if received	Salary		$80,672	$33,555	$34,050	$33,247
	Formula bonus	1	130,893	53,635	47,715	37,462
		2	31,629	20,070	21,050	9866
		3	48,633	9197	12,099	6579
	Discretionary bonus		36,449	20,135	13,295	10,728

NOTES: Means for components of compensation calculated only for managers receiving a positive amount. % receiving is less than % eligible because managers did not receive a bonus when performance was too low. "New" statistics include departments that combine new and used car sales.

and 3, in the order in which they were listed by the respondent. In all cases, respondents listed their largest bonus first, their next largest second, and their smallest last. Thus, this ranking corresponds to the economic importance of the FBs.

Most managers were eligible for at least one FB, though if performance was too low, some managers received no FB even if eligible. If awarded, the typical first FB was larger than the manager's salary, suggesting that incentives from this bonus are quite strong. By contrast, the incidence and magnitudes of the second and third FBs were much smaller, with roughly 10 percent eligible for up to three such bonuses.

The third major component of pay is discretionary bonuses. Because they are discretionary, all managers were eligible to receive such an award at the end of the fiscal year. In practice, roughly one in four managers received such a bonus. When awarded, these bonuses were similar in magnitude to

TABLE 2

CORRELATIONS OF PAY INSTRUMENTS

		Salary	Formula bonus 1	Formula bonus 2	Formula bonus 3
Formula bonus	1	0.15			
	2	–0.07	0.07		
	3	–0.03	0.02	0.56	
Discretionary bonus		0.02	0.02	0.02	0.05

NOTES: Correlations of dollar values of pay instruments, calculated in each case across all available observation pairs with nonmissing values.

the second FB, or roughly a half to a third of formula bonus 1. Thus, they are also likely to be an important source of incentives, but not as important as the FBs as a whole.

The fourth source of pay for managers is "spiffs," idiosyncratic reward programs sponsored by auto manufacturers. For example, Ford might offer a free trip to Hawaii based on meeting certain sales targets. These incentive plans are essentially out of the control of auto dealerships (except that they might have some control over who is eligible to participate). They are a relatively small part of pay, and they are hard to standardize. For these reasons, we ignore spiffs.

One immediate question about the various components of pay is whether they are substitutes or complements for each other. For example, some dealerships might pay low base salaries but high expected bonuses so that overall pay is similar to that of other dealerships. Similarly, some dealerships might provide discretionary rewards that are de facto tied closely to specific performance measures, so that they act very much like explicit FBs. Table 2 provides correlations of pay components to investigate this question. The correlations are almost all very close to zero, with no apparent pattern in positive and negative signs. This suggests that the pay instruments are *not* simply substitutes for each other, and that they may play different roles in the compensation system. The one large correlation is between the second and third FBs: 0.56. This may be an anomaly, or it may suggest that the second and third FBs play similar roles. We provide evidence for this below.

Table 3 describes the formulas used to calculate the FBs. All are piecewise linear contracts. All are convex (or straight linear) in performance, consistent with declining marginal utility of income, and increasing marginal disutility of effort. Less than a handful of formulas involve penalties (these are for

TABLE 3

STRUCTURE OF FORMULA BONUSES

		Formula bonus		
		1	2	3
% with	Floor	6	27	38
	Cap	2	19	12
	Neither	94	72	60
Maximum # of segments		5	6	4
% with lump sums		2	23	24
N		633	186	42

NOTES: Bonuses have a floor if the performance measure must exceed a positive threshold before any bonus is paid; and a cap if no bonus is paid for performance above some threshold.

inventory performance measures such as the number of cars in stock over 30 days).

Consider first the formula for the first bonus, FB1. Only 6 percent have an explicit floor (minimum performance level needed to earn any bonus) above zero. Almost none (2 percent) have a cap, or limit on the magnitude of the bonus that can be earned. Only 2 percent involve any lump sum payout, while 98 percent are simple linear commissions on a performance measure.

Now consider the formulas for the second and third bonuses, FB2 and FB3. These are strikingly different in form from FB1, but similar to each other. Both are much more likely to have floors and caps. Twenty-seven percent of FB2 and 38 percent of FB3 have a floor, while 19 percent and 12 percent, respectively, have caps. Even more interesting is that roughly one fourth of FB2 and FB3 involve lump sum payouts, which are almost never used for FB1. It is not clear why the second and third FBs have different structures than the first bonus. For now, we note that this similarity in structure may explain the correlation between FB2 and FB3 in Table 2. This is consistent with the idea that the second and third FBs play similar roles in the incentive system, and that they are not simply substitutes for FB1.

Other Variables. We now describe the variables. These fall into three categories: performance measures (and most importantly, their properties); explicit and implicit incentives; and controls.

Performance Measures. Most of the measures observed are variants on gross profit (revenue less the cost of goods sold) or net profit (gross profit less other costs). Because the cost of goods sold is the manufacturer's invoice price, it is beyond the manager's control. Thus, gross profit is similar

TABLE 4
PERFORMANCE MEASURE SCOPE

		Performance measure		
		1	2	3
Organizational unit (%)	Above unit	18.2	19.4	26.2
	At unit	73.8	48.4	38.1
	Within unit	7.9	25.8	26.2
	Different unit	0.2	6.5	9.5
	Total	100	100	100
Type (%)	Net profit	54.3	40.3	42.9
	Gross profit or revenue	44.7	29.6	23.8
	Units sold or in inventory	1.0	25.3	23.8
	Customer financing	0.0	4.8	9.5
	Total	100	100	100

NOTES: For performance measures for formula bonuses 1–3, show % measured at each level of organizational unit (top panel), and % of each type (bottom panel). Thus, percentages sum to 100 for each performance measure, in each panel. A measure is "at unit" if it is measured at the level of the manager's department (or the dealership for a GM). A measure is "above unit" if it is measured at the dealership level, for a department manager (not a GM). A measure is "within unit" if the measure covers a proper subset of the manager's department (e.g., parts sales for a service department manager; new car gross profit for a GM). A measure is "different unit" if it measures performance of a different department; these are always either a measure of the used car department, for a new car department manager, or vice versa.

to revenue, though it motivates consideration of profit margin. A very small number of contracts used units of sales or cars in inventory as the measure. Virtually none of the contracts in our sample used nonfinancial performance measures, such as indicators of customer satisfaction.

Table 4 shows the organizational unit at which these variables were measured (first panel), and the type of performance measure (second panel). "At unit" means that performance is measured at the level of the manager's department (the entire dealership for general managers). "Above unit" means that performance is measured at a broader level than the manager's own department. For general managers, this is of course impossible. For department managers, this usually means performance measured at the level of the dealership. The very small number of exceptions refers to cases where performance is measured for combined new and used car departments, but the manager runs only the new or used car department. "Within unit" means that performance is measured for a subset of the manager's unit. A typical example is the performance measure "gross profit, body parts" for a service department manager. This is only one part of the service department's business, which includes repairs and other activities. Another example is use of a performance measure for either new or used sales only, for a manager of a combined new-used car department. Finally, for general managers this would include any measure below the level of the overall dealership.

"Different unit" is the small number of cases where the manager of the new (used) car department is given a bonus based on a statistic from the used (new) car department.

Not surprisingly, almost three out of four measures for FB1 are at the level of the manager's department. This corresponds closely to the job design, since most of what they can control is at their department. It also should not distort much, compared to "within unit" measures, which may be too narrowly focused. At the same time, measures that are "at" or "within" the manager's unit provide little or no incentive to cooperate with other departments. If cooperation is important, then an option would be to use a measure that is broader ("above unit") or even of a "different" unit. Almost all performance measures for FB1 (PM1) are based on gross or net profit or revenue. Net measures are "broader," since they include both revenue and cost. Over half use net profit.

We saw above that the structures of FB2 and FB3 are similar to each other, but different from that of FB1. The same observation applies to performance measure choice, in both organizational unit and type of measure. PM2 and PM3 are less likely to be measured at the level of the manager's organizational unit. Instead, they are more likely to be narrower, measured "within" the unit. This is especially true for service department managers, where financial measures for components of revenue or costs (service, body parts, or labor) are sometimes used. The second and third performance measures also are more likely to be measured at a level above the manager's department, or in a "different" department altogether. These are likely attempts to improve cooperation between the manager's department and another department. In such cases, FB2 and FB3 are used to complement (fix weaknesses in) FB1.

Along the same lines, the second and third bonuses are where "nonstandard" performance measures are used—number of cars sold or in inventory, or measures of customer financing (car loans). These measures are almost never used for FB1. Note that the effects of inventory and customer financing on firm value are probably not adequately measured in short-term department revenue or profit. For example, a high inventory level implies a high opportunity cost to the dealership from tying up capital, but this would not usually be included in a department's accounting costs. Customer financing also is typically not included in the sales department's revenue, which is based solely on car sales. In both cases we see again that the second and third FBs are apparently used as complements to, or to address weaknesses in, the first FB.

Properties of Performance Measures. The survey included questions to assess five properties of each performance measure, recorded on a scale from 1 (not at all) to 5 (very high):

"To what extent does this measure:

1. reflect factors outside your control?
2. reflect your overall performance?
3. cause you to focus on short-term goals?
4. encourage cooperation with other departments?
5. motivate manipulating the measure to meet the performance target?"

The first of these properties (factors outside your control) is a good proxy for uncontrollable risk, whereas the second property (reflects overall performance) is a less ideal proxy for controllable risk. The recent literature on controllable risk was not circulating when we wrote the survey, so we will be careful to not over-interpret the evidence on the importance of controllable risk, due to the potential weakness of our measure to capture this concept.

The third and fourth properties (causes focus on short-term goals; encourages cooperation with other departments) measure two common distortions caused by accounting measures. In auto dealerships, some co-operation is needed between all three departments. New car sales frequently go to customers who also wish to sell their old car. Therefore, the departments may have new business leads for each other. In addition, developing a good relationship with a customer may improve the other department's ability to sell to that customer. Similar interdependencies arise between the service department and the sales departments. Both new and used cars require service, so both sales departments can encourage customers to use the dealership for service and repairs. Similarly, a satisfied customer of the service department is more likely to come to the dealership when they wish to buy or sell a car.

The final property is the extent to which the performance measure is manipulable. It might be expected that managers would be reluctant to admit that they manipulate their performance measures. However, in this sample there is roughly the same variation in responses to this question as for the other four questions about performance measure properties. The surveys were filled out by managers privately, handled with complete confidentiality, and sent directly to us (not the consulting firm), which may explain the willingness of managers to answer this question. Furthermore, industry experts indicated to us that manipulation is simply an accepted cost of imperfect performance measurement in such a sales-oriented industry. In any case, reluctance to report manipulation would bias down coefficients on this

TABLE 5

PERFORMANCE MEASURE PROPERTIES AS A FUNCTION OF SCOPE

		Scope of PM 1-3			
		Organizational unit			
		Above unit	At unit	Within unit	Different unit
Properties of PM 1–3	Reflects factors outside mgr.'s control (reverse coded)	3.11	3.27	3.06	3.33
	Reflects overall performance	3.53	3.67	3.28	3.00
	Causes short-term focus (reverse coded)	2.50	2.83	2.84	3.08
	Encourages cooperation	3.75	3.74	3.40	4.08
	Motivates manipulating the measure (reverse coded)	3.08	3.35	3.02	1.73

NOTES: Mean values of responses to questions about performance properties, scaled as: 1, not at all; 2, low; 3, medium; 4, high; 5, very high. Three of the five properties were then reverse coded; see the text.

variable (Bertrand and Mullainathan 2001), giving us some additional confidence in any significant results on manipulation that we are able to uncover.

The first, third, and fifth properties (uncontrollable risk, short-term focus, and manipulability) take larger values if the measure is "worse," while the second and fourth properties (controllable risk and cooperation) take larger values if the measure is "better." To make the presentation of results easier to interpret, the first, third, and fifth properties are reverse coded in all analyses. In other words, all performance measure properties are scaled so that a larger value indicates a better performance measure.

While not reported, we analyzed whether the five performance measure properties, and the four measures of their use for implicit incentives, varied with manager demographics. This is important for interpreting these variables, because they are based on perceptions. We found no evidence for differences in these variables across any manager characteristics, including age, education, and experience. This provides reasonable confidence that Bertrand and Mullainathan's (2001) concern about using survey data as dependent variables is not a significant threat to our analyses.

Table 5 presents summary statistics on these properties as a function of the organizational unit at which performance is measured. The patterns generally accord well with what would be expected. For example, the second property is the extent to which the manager reports that the performance measure reflects his overall performance. This is reported to be highest at the department level, and lower for measures that are either "within" or "above" the unit. It is lower still for measures based on a "different" department.

A performance measure is most likely to encourage cooperation if it is for a different department or the dealership as a whole. It is least likely to motivate cooperation if the measure is "within" the department. Similarly, manipulation is more difficult if the measure represents performance of a different department, and easier at the department level than at the level of the dealership as a whole. The one performance measure property that does not always have expected patterns across organizational units is the extent to which the measure reflects factors outside the manager's control. This is reported to be highest (least reflecting factors outside the manager's control) when it measures another department. One interpretation, however, is that a performance measure for a different department is chosen precisely in those cases where there are the greatest opportunities for cooperation between those two departments.

Explicit Incentives. A potential measure of incentive strength is the commission rate on the bonus plan. However, there are practical difficulties. Contracts use different measures that are not comparable across departments or dealerships. These measures may be on different scales (especially when considering the marginal effect of extra effort on the measure). Even when dealerships use the same nominal measure, there is variation in accounting methods across dealerships. Contracts may have multiple piecewise linear segments with different commission rates, and it is not clear which segment is relevant for incentives in a particular situation. Finally, contracts may use lump-sum bonuses, which are not in the same form as linear commissions and for which the correct measure of incentive intensity is not clear. Effort, and thus expected performance, should be positively related to the intensity of incentives. Thus, total received bonus is a proxy for the strength of the incentive that has the virtue of being comparable across different dealerships, departments, bonus formulas, and performance measures. The bonus regressions are tobits because some managers were eligible for a bonus but did not receive one if performance was too low. Proxying incentive intensity with realized bonus is, of course, imperfect. The bonus will be larger or smaller because of variation in the performance measure that is not due to the employee's effort. This imparts some error-in-variables to our measure of incentives.

Implicit Incentives. A feature of the survey is that it provides information on implicit incentives that have been rarely studied in economics or accounting. For each measure the survey asked:

"If you fail to achieve target performance for this measure, to what extent do you believe that the following will be adversely affected:

1. operating autonomy;
2. pay raise;
3. promotion prospects;
4. continued employment."

Responses were recorded on a scale from 1 (not at all) to 5 (very high). Respondents also reported the size of their discretionary bonus when applicable. While dealership managers have substantial pay for performance through their bonus plans, implicit incentives also are important. Salary is a large component of total pay. These jobs are highly paid, so threat of termination may drive incentives as well. Even promotion incentives may matter for these managers. Department managers might be promoted to general manager, and GMs earn approximately 2.5 times higher average pay than department managers in this sample. Furthermore, many dealerships are part of a network of shops, so department managers and GMs also may have the potential to be promoted to a better location or larger dealership.

Controls. The regressions include a variety of controls:

Service Department Dummy; Emphasis on Customer Service. When the job is more complex and intangible it may be harder to measure performance on some tasks accurately, leading to muted overall explicit incentives (Holmstrom and Milgrom 1991; Slade 1996). Instead, the firm might substitute greater emphasis on implicit incentives. For this reason, we predict that indicators that the job is more complex will have negative effects on incentive intensity. We use two such measures. Most regressions include dummy variables for whether a department manager is a service department manager. Service department jobs are more complex and involve more tasks for which performance is difficult to quantify. Our second indicator for a job with more intangible components is the emphasis placed on customer service (this variable was derived using factor analysis; see Appendix). Customer service has many dimensions compared to number of cars sold, and most are intangible.

Perceived Degree of Competition. We include a measure of the degree of competition (see Appendix). If the competitive environment is stochastic, the firm may want to provide incentives for the manager to respond to competition (Raith 2003). Therefore, we expect that employees will be given stronger incentives in more competitive environments. Evidence for this effect would favor the idea that greater controllable risk implies stronger incentives.

Number of Employees; Experience; General Manager Dummy. Finally, agency theory usually predicts that incentives should be stronger, the larger is the marginal product of effort. We include the number of employees reporting to the manager (a measure of resources under the manager's control), the manager's experience in the position (a measure of human capital), and a dummy variable for general managers. We predict that these will be positively related to the strength of incentives.

Findings

Table 6 presents analysis of the first prediction, that the incentive intensity for explicit incentives should be decreasing in noise, distortion, and manipulation of the measure; and increasing in controllable risk. The tobits assess the magnitude of formula-based bonuses for the full sample, and for general managers and department managers separately. They include the five performance measure properties as well as the controls described above.[4]

Since the performance measure properties are scaled so that a higher value means a "better" performance measure along that dimension, these variables are predicted to have positive coefficients. In most cases, the estimated coefficients are positive, and they are often statistically significant. The economic significance of the coefficients is straightforward to interpret (and similarly in Table 8 below). The standard deviation of the five performance measure properties is typically about 1.0. This means that the coefficient on the tobits in Tables 6 and 8 represents approximately the marginal effect of increasing or decreasing a performance measure property by one standard deviation. For example, a standard deviation improvement of 1 in the extent to which a performance measure encourages cooperation increases the average bonus by $11,257 overall, $27,357 for GMs, and $4480 for department managers. Similar magnitudes are found for the other properties. Those estimates constitute increases of 10 percent in the first FB, and even more for the second and third bonuses. These numbers are large economically. Thus, Table 6 provides strong evidence that performance measure properties have important economic effects on the magnitude of incentives.

The first two properties are our attempts to proxy for controllable and uncontrollable risk. The first is a relatively good proxy for uncontrollable risk. With the inclusion of the first factor, the second is a less perfect proxy for controllable risk. Despite this caveat, the coefficients for both are always positive

[4] Because the data include multiple observations from the same dealership, we ran all relevant analyses with Huber-White standard errors as a check. There were no important differences in significance. In fact, there is variety in incentive contracts (performance measures and formulas) for managers in the same dealership, perhaps because they run different types of departments.

TABLE 6

DETERMINANTS OF BONUS WEIGHTS

		Pred. sign	All Coef.	All SE	General managers Coef.	General managers SE	Dept. managers Coef.	Dept. managers SE
Intercept			–144,818	48,075***	–400,231	146,317***	–28,879	21,126*
Performance measure properties (PM1, 2, or 3)	Reflects factors outside mgr.'s control (reverse coded)	+	8151	4633**	11,238	11,736	4233	2125**
	Reflects overall performance	+	12,612	4578***	33,179	14,928***	2991	2039*
	Causes short term focus (reverse coded)	+	–4600	3821	3257	9027	–4836	1643
	Encourages cooperation	+	11,257	4172***	27,357	14,928**	4797	1726***
	Motivates manipulation (reverse coded)	+	8795	3214***	16,009	9027**	4480	1431***
Job and manager characteristics	# of employees	+	128	215	23	428	390	139***
	Degree of competition	+	15,512	6193***	71,583	23,383***	3482	2504*
	Emphasis on customer service	–	–16,857	8,324**	–55,588	23,812***	–4928	3682*
	Experience	+	3026	783***	6824	2465***	831	337***
	General manager	+	64,150	11,240***				
	Service department manager	–	–8396	12,185			–12,858	4853***
N			722		205		517	
% Bonus > 0			72%		81%		68%	

NOTES: Tobits predicting the magnitude of formula bonuses 1, 2, or 3. SE, standard error. *** significant at 1 percent; ** at 5 percent; * at 10 percent. Predicted signs of coefficients are shown after variable names; one-tailed tests in those cases. The first five variables are responses to survey questions (1–5 scale) asking about properties of performance measures. The variables "degree of competition" and "emphasis on customer service" are constructed from several survey questions using factor analysis (see Appendix).

and usually significant. Thus the evidence is consistent with Prendergast's (2002) analysis of risk and incentives. This is one of the few empirical studies to find a positive relationship between strength of incentives and degree of performance measure precision, after controlling for a measure of controllable risk (see DeVaro and Kurtulus (2006) for an earlier and more thorough empirical analysis of this question).

The next two properties measure whether the metric distorts incentives in two common ways, toward short-term results, and toward lack of cooperation. The results show that a performance measure that does not cause a short-term emphasis is *not* given stronger incentives in auto dealerships. In fact, in two of three regressions the coefficient is the opposite of predicted. One explanation is that auto dealerships desire their managers to emphasize short-term financial results, perhaps because of the terms of their contracts with manufacturers. However, that is speculation. Our prediction about the short-term focus of the performance measure is rejected. On the other hand, measures that encourage cooperation are indeed given greater weight for incentives, in all three specifications.

The final performance measure property is the extent to which it is unlikely to be manipulated to improve measured performance. Once again, in all three regressions this property has a significant effect on the strength of incentives, in the predicted direction. This provides evidence that managers do manipulate their performance measures, and that this affects the incentive plan's design. For this to be possible, managers must have some specific knowledge in performing their jobs that they can use to manipulate the measure. Thus, our evidence that manipulation occurs and is factored into incentives is additional evidence for Prendergast's view that agents have asymmetric information about how they perform their jobs, and that this has important effects on incentive system design.

The second half of the table includes controls for job design and the manager's human capital. Number of employees supervised (span of control) is a measure of the manager's marginal product of effort. This appears to have little effect on incentives once other controls are included. However, a dummy for general manager does have a positive sign. Experience is a proxy for the manager's human capital. Greater human capital may imply a larger marginal product of effort. The positive coefficients on experience suggest that this is the case in auto dealerships.

Degree of competition is another proxy for controllable risk (Raith 2003). Competitive actions by other dealerships are a kind of risk that managers can respond to with their own actions. We find a positive coefficient on our measure of competition in all three regressions. The effect is largest for general managers. This can be expected because they set overall policy and strategy for the dealership, and thus should control the dealership's response to competition.

Our proxies for job complexity and importance of intangibles show mixed results. The dummy variable for service departments is insignificant. However, the measures of emphasis on customer service are significant and positive in all three models as predicted. In summary, Table 6 provides good evidence that performance measure properties—controllable and uncontrollable risk, distortions, manipulation, and inability to capture intangibles—do matter for their use in incentive systems.

Table 7 examines the second and third predictions about the effects of performance measure properties on implicit incentives. These predictions involve the idea that implicit incentives allow the principal to use ex post evaluation to improve incentives. Specifically, implicit incentives can be used to punish the employee for failure to exploit controllable risk to improve firm value, or to punish manipulation if it is detected ex post. These two hypotheses are reflected in the predicted signs for the coefficients on the second and fifth performance measure properties in Table 7.

The dependent variables in this table are survey responses to questions that asked, "If you fail to achieve target performance for this measure, to what extent do you believe that [an implicit reward] will be adversely affected?" In other words, the questions asked whether a low value for a performance measure might be *punished* implicitly through promotions, raises, etc. Since these answers are on a 0–5 scale, ordered probits were estimated.

A concern in Table 7 is that the dependent variables are subjective answers to survey questions. Bertrand and Mullainathan (2001) conclude that, while survey data can be useful independent variables (as in Tables 5 and 8), they are more problematic as dependent variables. Specifically, suppose that GMs and department managers have different attitudes about how their evaluation affects their promotion prospects. Then coefficients on the GM dummy variable in Table 7 would reflect the difference in attitudes, as well as any difference in actual evaluation practices for GMs compared to department managers. As stated above, we found no significant differences in perceived performance measure properties across manager demographic groups. Nevertheless, interpretation of coefficients should be handled carefully when the dependent variable is subjective. We present Table 7 with this qualification in mind, and in the spirit of trying to see whether survey data provide useful insights into incentive practices. The main conclusions that we draw from the table are consistent with the predictions as well as with the inferences in the rest of the paper, however, and so we interpret them as reinforcing those conclusions and providing useful suggestions for future research.

TABLE 7

Effects of Performance Measure Properties on Implicit Incentives

			Ordered Probits							
		Pred. sign	a. Operating autonomy		b. Pay raise		c. Promotion prospects		d. Continued employment	
			Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Perf. measure 1	Reflects factors outside mgr.'s ctl. (reverse coded)		–0.147	0.048	–0.091	0.048	–0.006	0.048	–0.121	0.048
(PM1) properties	Reflects manager's overall performance	+	0.151	0.049***	0.111	0.049***	0.082	0.049**	0.164	0.049***
	Causes short-term focus (reverse coded)		–0.034	0.043	–0.002	0.043	–0.097	0.044***	–0.048	0.043
	Encourages cooperation		0.004	0.043	0.004	0.042	0.060	0.043	–0.016	0.042
	Motivates manipulation (reverse coded)	–	–0.124	0.034***	–0.078	0.034***	–0.115	0.034***	–0.103	0.034***
General manager			–0.272	0.114***	–0.366	0.112***	–0.581	0.115***	–0.459	0.114***
Service department manager			0.128	0.110	–0.026	0.109	–0.268	0.110**	–0.063	0.109
Cutoffs	1		–1.374	0.293	–1.123	0.291	–1.130	0.295	–1.167	0.294
	2		–0.551	0.290	–0.507	0.289	–0.443	0.293	–0.399	0.291
	3		0.476	0.290	0.106	0.288	0.393	0.292	0.482	0.292
	4		1.259	0.298	0.882	0.290	1.218	0.299	1.059	0.297
N			580		587		583		588	
Likelihood ratio			58.2		33.2		67.8		62.2	
Prob. > χ^2			0.00		0.00		0.00		0.00	

Notes: Ordered probits predicting responses to: "If you fail to achieve target performance for this measure, to what extent do you believe that the following will be adversely affected?" Survey responses scaled 1–5: 1, not at all; 2, low; 3, medium; 4, high; 5, very high. SE, standard error. *** significant at 1 percent; ** at 5 percent; * at 10 percent. Predicted signs of coefficients are shown after variable names; one-tailed tests in those cases.

TABLE 8

Effects of Performance Measure Properties on Other Formula Bonuses

		Pred. sign	Formula bonus 2 or 3	
Intercept			4027	2094**
Property of PM2	Reflects factors outside mgr.'s control (reverse coded)	+	201	1522
or PM3 Minus	Reflects overall performance	+	1633	1632
Property of PM1	Causes short-term focus (reverse coded)	+	86	1436
	Encourages cooperation	+	4046	1401***
	Motivates manipulation (reverse coded)	+	2527	1628**
General manager			2844	4225
Service department manager			–6979	3120***
N			315	
% Bonus (#2 or 3) > 0			60%	

Notes: Tobit predicting magnitude of formula bonuses 2–3. SE, standard error. *** significant at 1 percent; ** at 5 percent; * at 10 percent. Predicted signs are shown after variable names; one-tailed tests in those cases.

The results in Table 7 are consistent with the predictions. Roughly speaking, a one-unit change in either the second or fifth performance measure property increases the mean value of the dependent variable by about one quarter unit—increasing the likelihood that the manager's implicit incentive will be adversely affected. The more that a measure reflects overall performance, the more likely it is that a low value of that measure will be punished implicitly. We have interpreted this property as a potential proxy for controllable risk, but with qualification, so we will not put much weight on this finding. The most interesting result in the table is that if a measure is less likely to motivate manipulation, it is less likely that poor performance will be punished implicitly. Put in reverse, if performance is low even though the measure might be manipulated, it must be quite poor performance indeed, and it is punished. This finding is interesting, because it is evidence for our notion that manipulation makes use of the employee's specific knowledge in performing the job, and so must be deterred through ex post punishment. Distorted incentives, on the other hand, are predictable in advance, since the performance measure's balance (or lack) across different tasks is known in advance. Thus, distortions are less likely to require ex post punishment for deterrence.

Table 8 tests the fourth prediction, that bonuses on additional performance measures can be used to rebalance incentives from the first performance measure. We measured the five performance measure properties of the second or third measure relative to the value of that property for the first measure, by subtracting the value for the first measure. A larger value means that the second or third measure is reported to be relatively better along that dimension than is the first measure. To the extent that this is true, we predict that the new measure will be given greater weight in the evaluation—especially for the measures of distortion (short-term focus or cooperation) and manipulation, since those are most easily "reversed" by use of a second performance measure. Risk is less likely to be "reversible" with a second measure, since the measure would have to have risk properties that are negatively correlated with those of the first measure. The regressions in Table 8 are tobits predicting the magnitude of the second or third bonus.

The results in Table 8 suggest that an additional performance measure is given greater weight for incentives if it improves the manager's incentives for cooperation, or if it is less subject to manipulation. These effects are both statistically and economically significant. As in Table 6, the standard deviation of the key independent variables—in this case, differences in performance measure properties—is approximately equal to 1.0. Therefore, coefficients can be interpreted as approximately the marginal effect of raising or lowering the difference in performance measure property by 1 standard deviation. For example, a 1 standard deviation improvement in the relative extent to which an additional performance measure improves cooperation compared to the primary performance measure results in an average increase in bonus 2 or 3 of about $4046. That is a large effect compared to the average size of bonuses 2 or 3. The effect of such an improvement in the relative extent to which an additional measure does not motivate manipulation is about $2527, also a large effect. Recalling that we found no evidence that short-term focus was an important performance measure property in our sample, these findings do suggest that additional measures are chosen, at least in part, to improve the overall evaluation of the manager's performance compared to the first performance measure.

Conclusions

In this paper we use data from a survey that we designed and collected to study the effects of performance measure properties on incentive system design. We develop direct measures of performance measure properties: controllable and uncontrollable risk, distortion, and manipulation, and analyze their effects on incentive plan design. We explore the premise that a firm uses a *system* of interrelated measures and incentives—explicit and implicit—because of weaknesses in available performance measures.

The performance measure properties that we analyze are the measure's noise, controllable risk, distortion, and manipulability. We find that all of these properties are important to incentive plan design. The more that a measure is flawed along any of these dimensions, the less weight is given to that measure for explicit incentives. We find some evidence that a second measure can mitigate distortions or manipulation arising from the first performance measure. This indicates that the firm may pick a set of performance measures based on how their properties are related to each other.

Prior empirical research on the trade-off between risk and incentives has often failed to find the predicted relationship. We do find such a relationship, and present evidence supporting the distinction between controllable and uncontrollable risk. We also present evidence on the importance of distortions and manipulation, two topics that have received relatively less attention in economics. Our results on the existence and deterrence of manipulation, and on the effects of competition, are additional evidence for the relevance of controllable risk to incentive plan design.

Finally, we explore a relatively under-studied issue, implicit rewards. One of the most important reasons for implicit incentives is to, in effect, turn a numeric performance measure into a subjective evaluation (or similarly, to make the weight on the measure subjective). This flexibility allows the supervisor to use ex post information to "fix" problems in the numeric measure, improving the overall incentive. Our results indicate that this is particularly useful for deterring manipulation, and may also be used to motivate the employee to exploit controllable risk on behalf of the firm.

Several important caveats apply to this research. Our data are cross-sectional. We have made every attempt to control for possible unobserved heterogeneity, and the sample is from a single industry, but panel data would be preferred. Our data are also survey based, and survey data are more noisy. However, it is worth noting that they can be less noisy than proxying for hard to measure concepts using traditional archival data. Once again the industry study design mitigates but does not eliminate this concern. The fact that we have some statistically significant findings despite the potential for attenuation bias is encouraging. An additional concern of survey data is unobserved heterogeneity driving correlations between dependent and independent variables. We find no evidence that manager demographic characteristics drive our findings. However, we cannot be certain, and this concern may be higher with survey data. One purpose of our study is to explore the potential for survey data to provide new insights into incentive plan design. Survey data have advantages in addition to weaknesses, notably in that it allows for the study of important questions that cannot be easily addressed with more typical data sets. Therefore we view our findings as suggesting interesting directions for future research with other data sources—and perhaps for future new theoretical insights.

References

Baker, George. 1992. "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100(3):598–614.

———. 2002. "Distortion and Risk in Optimal Incentive Contracts." *Journal of Human Resources* 37(4):728–52.

———, and Bjorn Jorgensen. 2003. "Volatility, Noise and Incentives." Working Paper, Harvard Business School.

———, Robert Gibbons, and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109(4):1125–56.

Banker, Rajiv D. and Srikant M. Datar. 1989. "Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation." *Journal of Accounting Research* 27(Spring):21–39.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review Papers and Proceedings* 67–72.

Bouwens, Jan, and Laurence van Lent. 2006. "Performance Measure Properties and the Effect of Incentive Contracts." *Journal of Management Accounting Research* 18(1):55–75.

———, and ———. 2008. "Assessing the Performance of Business Unit Managers." *Journal of Accounting Research* 45(4):667–97.

Bushman, Robert, Raffi Indjejikian, and Abbie Smith. 1996. "CEO Compensation: The Role of Individual Performance Evaluation." *Journal of Accounting and Economics* 21(April):161–93.

Campbell, Dennis. 2008. "Nonfinancial Performance Measures and Promotion-Based Incentives." *Journal of Accounting Research* 46(2):297–332.

Courty, Pascal, and Gerald Marschke. 2004. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics* 22(1):23–56.

———, and ———. 2008. "A General Test of Gaming." *Review of Economics & Statistics* 90(3):428–41.

Demski, Joel S., Hans Frimor, and David E. M. Sappington. 2004. "Efficient Manipulation in a Repeated Setting." *Journal of Accounting Research* 42(1):31–49.

DeVaro, Jed, and Fidan Ana Kurtulus. 2006. "An Empirical Analysis of Risk, Incentives, and the Delegation of Worker Authority." Working Paper, Cornell University.

Feltham, Gerald A., and Jim Xie. 1994. "Performance Measure Congruity and Diversity in Multi-task Principal/Agent Relations." *The Accounting Review* 69(3):429–53.

Gibbs, Michael, Kenneth A. Merchant, Wim A. Van der Stede, and Mark E. Vargus. 2004. "Determinants and Effects of Subjectivity in Incentives." *The Accounting Review* 79(2):409–36.

Hayes, Rachel M., and Scott Schaefer. 2000. "Implicit Contracts and the Explanatory Power of Top Executive Compensation for Future Performance." *RAND Journal of Economics* 31(2):273–93.

Healy, Paul M. 1985. "The Effect of Bonus Schemes on Accounting Decisions." *Journal of Accounting and Economics* 7:85–107.

Hemmer, Thomas. 1996. "On the Design and Choice of 'Modern' Management Accounting Measures." *Journal of Management Accounting Research* (8):87–116.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10:74–91.

———, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7:24–52.

Holthausen, Robert, David Larcker, and Richard Sloan. 1995. "Annual Bonus Schemes and the Manipulation of Earnings." *Journal of Accounting and Economics*, 19:29–74.

Ichniowski, Casey, Kathryn Shaw, and Giovanni Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *American Economic Review* 87(3):291–313.

Ittner, Christopher D., and David F. Larcker. 2002. "Determinants of Performance Measure Choices in Worker Incentive Plans." *Journal of Labor Economics* 20(2):S58–91.

———, ———, and Madhav V. Rajan. 1997. "The Choice of Performance Measures in Annual Bonus Contracts." *The Accounting Review* 72(2):231–55.

Jensen, Michael, and William Meckling. 1992. "Specific and General Knowledge and Organizational Structure." In *Contract Economics*, edited by Lars Werin and Hans Wijkander. Oxford: Blackwell.

MacLeod, Bentley, and Daniel Parent. 1999. "Job Characteristics and the Form of Compensation." *Research in Labor Economics* 18:177–242.

Murphy, Kevin J., and Paul Oyer. 2003. "Discretion in Executive Incentive Contracts." Working Paper, USC.

Prendergast, Canice. 1999. "The Provision of Incentives Within Firms." *Journal of Economic Literature* 37(1):7–63.

———. 2002. "The Tenuous Tradeoff between Incentives and Risk." *Journal of Political Economy* 110(5):1071–102.

Raith, Michael. 2003. "Competition, Risk, and Managerial Incentives." *American Economic Review* 93:1425–36.

———. 2008. "Specific Knowledge and Performance Measurement." Working Paper, University of Rochester.

Shi, Lan. 2008. "Respondable Risk and Incentives for CEOs: The Role of Information-collection and Decision-making." Working Paper, University of Washington.

Slade, Margaret. 1996. "Multitask Agency and Contract Choice." *International Economic Review* 37(2):465–86.

Van Praag, Mirjam, and Kees Cools. 2001. "Performance Measure Selection: Aligning the Principal's Objective and the Agent's Effort." Working Paper, University of Amsterdam.

APPENDIX

Description of Factor Variables

Survey questions used to construct factors		Factor loadings (Cronbach's α)
Perceived degree of competition		($\alpha = 0.72$)
In your trading area, how much competition does your dealership face?		0.87
How intense is competition for good employees in the car dealer business?		0.70
How intense is price competition for new cars?		0.81
Emphasis on customer service (general managers)		($\alpha = 0.84$)
To what extent do you . . .	Evaluate department managers on customer service performance?	–0.82
	Review customer service issues in meetings with department managers?	0.78
	Consider customer service to be a way to increase profits?	0.77
	Find customer service important relative to financial performance?	0.68
	Provide feedback to dept. mgrs. about customer service performance?	0.67
	Provide training to employees to increase customer service awareness?	0.43
Emphasis on customer service (department managers)		($\alpha = 0.92$)
To what extent do you . . .	Involve personnel in customer service improvement?	0.78
	Hold personnel responsible for customer service?	0.77
	Discuss customer service in personnel meetings?	0.80
	Consider customer service a way to increase profits?	0.73
	Make customer service data available to personnel?	0.78
	Use customer service data to evaluate your personnel?	0.77
	Display customer service data at employee workstations?	0.59
	Give employees feedback on customer service performance?	0.82
	Have employees participate in customer service improvement decisions?	0.73
	Build ongoing awareness about customer service among employees?	0.84

Notes: Factor analysis with principal component extraction and oblique rotation ($\delta = 0$). The Kaiser–Meyer–Olkin measure of sampling adequacy is adequately high (0.80). The Bartlett test of sphericity yielded highly significant χ^2 ($p = 0.00$). The Cronbach's α's are highly adequate ($\alpha > 0.70$).