Center for
Effective
Organizations

PERFORMANCE MEASURE PROPERTIES AND INCENTIVES

CEO PUBLICATION
T 07-2 (512)

MICHAEL GIBBS
UNIVERSITY OF CHICAGO

KENNETH A. MERCHANT
UNIVERSITY OF SOUTHERN CALIFORNIA

WIM A. VAN DER STEDE
LONDON SCHOOL OF ECONOMICS & POLITICAL SCIENCES

MARK E. VARGUS
UNIVERSITY OF TEXAS – DALLAS

April 2007

Center for Effective Organizations - Marshall School of Business
University of Southern California - Los Angeles, CA 90089–0871
(213) 740-9814 FAX (213) 740-4354
http://www.marshall.usc.edu/ceo

PERFORMANCE MEASURE PROPERTIES AND INCENTIVES*

Michael Gibbs,[a] **Kenneth A. Merchant**,[b] **Wim A. Van der Stede**,[c] **and Mark E. Vargus**[d]

September 28, 2006

We provide a comprehensive empirical analysis of incentive design, focusing on effects of performance measure properties (controllable and uncontrollable risk, distortion, and manipulation). The data are from auto dealership manager incentive systems. We find that dealerships put the most weight on measures that are "better" with respect to these properties, for both explicit incentives (bonuses) and implicit incentives (e.g., promotions or autonomy). In addition, a measure's properties relative to those of other measures also affects its weight for incentives. Implicit incentives, which are awarded ex post, are used to deter manipulation of the performance measure. Our results are consistent with a setting where employees have multitask jobs, and face controllable risk. Firms use multiple bonuses and implicit rewards to balance multitask incentives. Controllable risk affects incentive design in two ways: it motivates using the information to improve firm value, and deters manipulation of the performance measure.

Keywords: performance measures, incentives, risk, specific knowledge, distortion, manipulation, multitasking.

[a] Graduate School of Business, University of Chicago, Chicago, IL 60637; Institute for the Study of Labor (IZA).

[b] Marshall School of Business, University of Southern California, Los Angeles, CA 90089.

[c] London School of Economics & Political Science, London WC2A 2AE.

[d] School of Management, University of Texas–Dallas, Dallas, TX 75083.

* We are grateful to an unnamed consulting firm for giving us access to their data and clients, and for numerous discussions that helped us understand the auto dealership business and clarify the data. For comments on the various drafts of this project, we thank Jan Bouwens, Mark Bradshaw, Jim Brickley, Jed DeVaro, Leslie Eldenburg, Joan Luft, Margaret Meyer, Kevin J. Murphy, Walter Oi, Canice Prendergast, Michael Raith, Edward Reidl, Bernard Salanié, and Sally Widener; seminar participants at Arizona, Harvard, Rochester, University of Aarhus, University of Tilburg, Universidad de Navarra, Universitat Pompeu Fabra, USC; and conference participants of AAA, BMAS, CAED, CEPR, and Society of Labor Economists. Liu Zheng provided helpful research assistance.

PERFORMANCE MEASURE PROPERTIES AND INCENTIVES

1. INTRODUCTION

Performance measurement is perhaps the most difficult and significant 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 (e.g., use accounting numbers instead of stock price to evaluate a CEO) to reduce risk, but that often results in distorted incentives (e.g., too short-term focused). In addition, the employee may be able to use his 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. Of course, discretion raises its own concerns, such as the potential for favoritism and relational contracting.

Consistent with their importance in practice, performance measure problems have received increasing attention in agency theory. The original models (e.g., Holmstrom 1979; Banker & Datar 1989) emphasized uncontrollable risk (noise). Later models incorporated multitask-incentives (Holmstrom & Milgrom 1991), which motivated formal consideration of distortions and manipulation (Baker 1992; Feltham & Xie 1994; Demski, Frimor & Sappington 2004). Recent work has emphasized controllable risk (Prendergast 2002; Baker 2002; Raith 2005). In accounting, the empirical literature analyzing performance measure properties focuses largely on risk or distortion, with less work on manipulation (Bushman, Indjejikian & Smith 1996; Ittner, Larcker & Rajan 1997; Van Praag & Cools 2001; Ittner & Larcker 2002). Despite the importance of performance measurement in practice, in agency theory, and in accounting, the empirical economics literature on performance evaluation is surprisingly small.

This paper contributes to this small literature on performance measurement by providing analysis of several parts of the puzzle together. We constructed a unique dataset on the entire incentive system for

a set of managers in auto dealerships. This allows 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 designed 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 potential for manipulation) among those available. This finding applies to both explicit incentives (bonuses) and implicit incentives (e.g., promotions or autonomy). Second, the ex post awarding of implicit incentives is used to deter manipulation of the performance measure.

Third, the paper provides some of the first empirical evidence on the distinction between controllable and uncontrollable risk. Performance measure properties that correspond best to 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 in two ways. One is to reduce incentives to use such information to manipulate performance measures. The other is to encourage employees to productively respond to changes in their environment.

Fourth, our evidence indicates that that multiple bonuses are used 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.

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

2. PERFORMANCE MEASURE PROPERTIES: THEORY

Theoretical work on performance measures is scattered, with different models focusing on different issues. We synthesize that literature briefly here, using the following terminology. *Performance measure* refers to a quantitative indicator such as accounting profits or number of cars sold. *Formula bonus* refers to a bonus that is calculated using a mathematical formula based on a performance measure. In our dataset, we distinguish up to three formula bonuses, each using only a single performance measure. 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. These include the manager's autonomy, raises, promotions, and possible termination.

Incentives Based on a Single Measure

Classic agency theory emphasizes the tradeoff between risk and incentives. Imperfect performance measurement implies that pay for performance is risky. When performance measure is imperfect, a stronger relationship between performance and compensation results in greater risk to the employee, and a larger risk premium is required. The firm might spend more resources to improve the accuracy of the measure, but it is also likely to reduce the strength of the incentive (e.g., Holmstrom 1979). This literature also noted that a larger marginal product of effort should imply a stronger incentive, all else equal. In the accounting literature, these two effects are often referred to as performance measure "signal" and "noise" (Banker & Datar 1989).

Later work expands on the question of risk in two ways. One strand notes that firms may adopt more narrowly focused measures (e.g., department profits instead of firm profits) to reduce the effects of noise on the performance measure. However, this may result in performance measure distortion (Feltham & Xie 1994; Baker 1992, 2002). The reason for this is that the more focused measure may filter out not only measurement error, but also the effects of some of the employee's tasks (e.g., cooperation with other departments). More generally, when the worker has a multitask job, a performance measure may give too

much or too little relative weight to some aspects of the job, resulting in distortion (Holmstrom & Milgrom 1991).

A more recent strand distinguishes between uncontrollable and controllable risk or specific knowledge (Prendergast 2002; Baker & Jorgensen 2002; Raith 2005). Classical agency theory considers risk as additive noise (as above), which is a nuisance to performance evaluation. This is a form of what we call uncontrollable risk; that is, factors that affect performance that are beyond the control or knowledge of the employee. There is also a second form of risk: states of the world that are not known when the contract is signed, but are realized and observed by the employee as events unfold. If the employee has such "specific knowledge" or controllable risk, he can use it to improve performance. Greater controllable risk of this type should imply stronger pay for performance, to support greater delegation to the worker.

Another important performance measure property is potential for manipulation by the employee to improve the measure but not firm value, and perhaps even at the expense of firm value (Courty & Marschke 2004, forthcoming).[1] The term manipulation is often used casually in ways that are similar to what we call distortion above (both are often put under the rubric of "gaming"), but they are different ideas. Both involve the employee focusing on incentive compensation, rather than firm value, but for different reasons. Distortion arises because the incentive system's ex ante weights on different aspects of the job are not balanced appropriately. Manipulation arises because the employee exploits private information ex post (after the incentive plan is designed). Thus, manipulation can arise even if the weights on measures are balanced ex ante. Note that manipulation requires controllable risk (employee specific knowledge). Therefore evidence for manipulation would also be evidence that controllable risk plays a role in the incentive system.

[1] Of course, performance measures might well be manipulated by the firm as well as the employee. Since our dataset provides no information on manipulation by the firm, we ignore that possibility.

There are two implications of manipulation. First, since manipulation reduces the ex post correlation between the measure and firm value, a measure should be given less weight for incentives if it is more open to manipulation. Second, if the principal can observe signals about the state of nature, typically ex post (or at least after the contract is written), she can use implicit methods of ex post settling up to reduce (deter) manipulation (Baker, Gibbons & Murphy 1994). Therefore, we predict that implicit incentives will be more important, the greater is the potential for a performance measure to be manipulated by the employee.

Incentives Based on Multiple Measures

Many of the dealers in our sample use more than one performance measure for bonuses. Feltham & Xie (1994) suggest several reasons to do so. 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. Similarly, Baker (2002) shows that when two measures are used in an incentive system, the weight on each is a decreasing function of both the riskiness and the 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.

Hemmer (1996) models the use of two performance measures for bonuses in exactly the empirical setting we study, automobile dealerships. Hemmer analyzes the use of a non-financial performance measure, customer satisfaction, alongside a sales-based measure. Despite the identical setting, Hemmer's model does not apply to our sample. In our data, no dealership used a customer satisfaction measure for any bonus plan. Rather, they used discretionary bonuses for any motivation of attention to customer satisfaction. Discretionary bonuses are not considered in Hemmer's model.

Empirical Implications

Summing up the discussion above, the literature points to several performance measure properties of interest: uncontrollable risk (noise); controllable risk; distortion; and manipulation. These properties

affect the optimal weight that should be placed on a measure for both explicit and implicit incentives. The incentive intensity – for both explicit and implicit incentives – should be decreasing in noise, distortion, and potential manipulation of the measure. The incentive intensity should be increasing in the importance of controllable risk.

Our discussion points to two general ways that weaknesses in performance measures can be addressed. One (discussed here, but not emphasized in the theoretical literature) is to use ex post subjectivity to "fix" the overall evaluation (Gibbs, Merchant, Van der Stede & Vargus 2004). Such an approach is costly (especially because workers often do not trust such an approach), but can in principle mitigate any potential problem with a numeric performance measure. For example, if an explicit incentive based on a "contractible" measure distorts incentives, the manager might take into account qualitative dimensions of performance in awarding a promotion, to rebalance incentives. Such discretion in rewards is likely to be particularly important for deterring manipulation, since manipulation relies on employee knowledge arising after the incentive contract is written. A manager may observe ex post signals suggesting that the employee manipulated the measure, and use those in the awarding of implicit rewards. The expectation of doing so can deter some manipulation. Therefore, implicit rewards should be especially important when performance measures used for formal bonuses are most subject to manipulation.

Finally, a second way to address weaknesses in a performance measure is to choose other measures for additional bonuses. The key prediction is that a second or third performance measure can be chosen to fix weaknesses in a first measure, by reducing uncontrollable risk or overall distortion. If a firm uses multiple bonuses, we expect that the additional performance measures will be given greater weight if they are relatively better (compared to the measure for the first bonus) – less noisy or distorted (and possibly less subject to manipulation).

3. DATA

We now describe the data collected and used for this study. A boutique auto dealership consulting firm contracted us to ask if we were interested in designing and implementing a survey on incentive prac-

tices in the industry. We thus had the opportunity to ask questions and collect data on variables that are usually not measured in datasets available to academics. For this reason, the dataset provides the most comprehensive information on the design of a set of incentive systems that we are aware of. In addition, because it is based on a survey, it includes information on issues that have rarely been studied due to lack of data, such as subjective evaluation and implicit incentives.

Our survey methodology has positive and negative features. To our knowledge, it provides the most detailed information ever collected on the systems of incentives, explicit and implicit, used within firms. However, survey data have downsides. Since it can be difficult getting surveys filled out, our sample size is limited. In addition, survey data tend to be noisy: by nature, much of the information is perceptual and difficult to quantify. Such data can, however, shed light on questions that are otherwise difficult or impossible to study with more traditional, publicly-disclosed datasets.

Our study also follows the recent trend towards industry studies (e.g., Ichniowski, Shaw & 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. In this sense our data are less noisy than those from general surveys of participants in a wide variety of circumstances. 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. A substantial benefit is that our main focus, performance measurement, is very similar for all firms sampled. Similar issues of distortion, manipulation, and so forth should apply to every manager in our dataset. By contrast, studying these particular variables with cross-industry data would be much more difficult. Of course, a weakness of industry studies, including this one, is that it is difficult to gauge how generalizable the findings are.

Before developing the survey, we spent a day at a large dealership interviewing the owner (who owns several dealerships) and operating managers. This acquainted us with the business, job designs, incentive issues, and the language they use. In addition, the consulting firm surveyed its clients on incentive

practices several years before the project. We used these survey questions and responses, and our field interviews, to develop our survey. The initial version was discussed with five of the firm's consultants. A revised pilot-test survey was then mailed to 24 dealerships whose feedback we used to make further changes and finalize the survey.

We developed surveys for the owner (survey A), general manager (B), and the managers of the service (C), new car sales (D), and used car sales (E) departments. The owner survey was relatively short, asking about ownership, bonus payments, and owner demographics. The general manager survey included questions about the dealership's competitive environment, strategy, and management practices. The service, new car, and used car department surveys were largely identical except for relevant word substitutions. The most important section of surveys B-E asked detailed questions about compensation, including salary, bonuses, performance measures, bonus formulas, and subjective evaluations. Outside the compensation section, the surveys principally contained 5-point Likert scales. Of these, we use two multi-item scales to assess the degree of competition and dealership/ department emphasis on customer service (see Appendix).

We mailed the final set of five surveys to 1,203 dealerships, along with both our cover letter and one from the consulting firm stating their support for the study. We sent a reminder letter (without replacement) to non-participants after four 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 1,057 surveys (A:277; B:250; C:205; D:186; E:136), or 18% of those mailed. A few were not useful, most commonly because they had substantial missing data. Of the 185 D-respondents (new car), 39 had a combined new and used car sales department, and hence were instructed to complete survey D only and discard E. We have at least one

[2] Our response rate might have been improved by a second follow-up with replacement (Dillman 1978). The first mailing and one follow-up (reminder letter) yielded 977 replies, leaving 5,038 non-replies. However, the consulting firm did not want us to approach their clients aggressively.

survey from 326 different dealerships, or 27%.[3] We found no evidence of sample selection bias on the basis of performance, size, geography, or manufacturer.

Where feasible, responses were keypunched by a professional data entry firm. Compensation contract data were coded in two steps. First, a doctoral student (who was knowledgeable about the academic literature) coded the contracts literally as written by the respondents; i.e., without interpretation or loss of information. Next, the authors went through each survey to reduce variability in wording of performance measures and to double-check numerical entries. We also called several respondents to ask for clarification. Finally, we conducted several data integrity checks to further ensure coding accuracy.

4. COMPENSATION OF AUTO DEALERSHIP MANAGERS

Compensation plans for managers in automobile dealerships are set by dealership owners, not auto manufacturers. Generally, compensation plans are set each year, with an annual evaluation to determine rewards based on last year's performance, and establishment of incentive plans for the next year. Auto dealers do not tend to use complex methods of evaluation such as MBO (managing by objective) or 360° appraisals. Instead, relatively simple methods are used. Table 1 provides summary statistics for our sample.

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, formula bonuses, and discretionary bonuses. While almost all managers are paid a salary, it is interesting to note that not all are. In the two types of sales departments, in particular (new and used), roughly 10% of managers are paid *only* through pay for performance, with zero base salary. Salary averages a bit less than half of total pay. Compared to most industries, pay for performance is a very large part of compensation for managers in this industry.

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

The most important component of pay for performance is formula bonuses. In our sample, managers were eligible for up to three bonuses calculated as an explicit function of a specified performance measure. We defined these bonuses as Formula Bonuses 1, 2, and 3, in the order in which they were listed by the respondent on the survey instrument. In all cases, respondents listed their largest (in expected or actual value) bonus first, their next largest second, and their smallest last. Thus, this ranking corresponds to the economic importance of the formula bonuses.

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

The third major component of pay is discretionary bonuses. A discretionary bonus is awarded not by explicit formula, but by the supervisor's subjective judgment of whatever factors deemed important. Because they are discretionary, all managers are eligible to receive such an award at the end of the fiscal year. In practice, roughly one in 4 or 5 managers receives such a bonus each year. When awarded, these bonuses are similar in magnitude to the second formula bonus, 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 formula bonuses as a whole.

The fourth source of pay for managers is spiffs, which are small, idiosyncratic reward programs sponsored by auto manufacturers. For example, Ford might offer a free trip to Hawaii to employees of dealerships who meet 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 in both incidence and magnitude. 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 relatively low base salaries, but structure their bonus plans with relatively high expected payout levels, so that overall *expected* pay is

similar to that of other dealerships. Similarly, some dealerships might provide discretionary rewards that are de facto tied very closely to specific numeric performance measures, so that they act very much like explicit formula bonuses. 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 strongly 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 correlation that is large is between the second and third formula bonuses: their correlation is 0.56, much larger than any other correlation. This may be an anomaly, or it may suggest that the second and third formula bonuses play similar roles.

Table 3 describes the structure of the formulas used to calculate the formula bonuses. *All* of the formula bonuses are piecewise linear contracts. In *all* cases, if the slope changes at all, it is steeper for higher levels of performance. This is consistent with declining marginal utility of income, and increasing marginal disutility of effort.

Less than a handful of formula bonuses involve penalties (these are for inventory performance measures such as the number of cars in stock over 30 days). This means that virtually all formula bonuses have a floor of zero (for example, no bonus paid if gross profit equals zero).

Consider first the structure of the first formula bonus, FB1. Only 6% have an explicit floor (minimum performance level needed to earn any bonus) above zero. Similarly, almost none (2%) have a cap, or limit on the magnitude of the bonus that can be earned. Only 2% of these bonuses involve any lump sum payout; 98% are paid as linear commissions on some performance measure. In fact, a large number of FB1 contracts are simply linear schemes that pay a certain fraction of the performance measure as the bonus (say, 2% of gross profit).

Now consider the structure of Formula Bonuses 2 and 3, FB2 and FB3. These are strikingly different in form from FB1. Both are much more likely to have floors and caps in payouts. 27% of FB2 and 38% of FB3 have a floor. Similarly, 19% and 12%, 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 formula bonuses have such different structures than the first formula

bonus. For now, we note that this similarity in formula structure may help explain the large correlation between FB2 and FB3 in Table 2, and zero correlation with FB1. This is consistent with the idea that the second and third formula bonuses may play similar roles to each other in the incentive system, and that they are not simply substitutes for other pay instruments. We return to this question below.

5. VARIABLES

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

Performance Measures. Most of the performance 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 to revenue from the manager's perspective, though it motivates consideration of profit margin. A very small number of contracts used units of sales or cars in inventory as the measure. *None* of the contracts in our sample used non-financial performance measures, such as indicators of customer satisfaction.

Table 4 shows the organizational unit at which these variables were measured for managers' formula bonuses (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 and never occurs. For department managers, this usually means performance measured at the level of the dealership. The very small number of exceptions are 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 of responsibility. A typical example of this is the performance measure "Gross Profit, Body Parts" for a Service Department manager. Auto body parts is only one part of the 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 De-

partment. 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 3 out of 4 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 andPM3 are much less likely to be measured at the level of the manager's organizational unit. Instead, they are much more likely to be narrower, measured "Within" the unit. That 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 "non-standard" performance measures are used –units sold or in inventory, or measures of customer financing (car loans). These measures are almost never used for FB1. Note that units in inventory, and customer financing, are aspects of the department's activities whose effects on long-term firm value are probably not fully measured in department revenue or profit. For example, a high inventory level implies a high opportunity cost to the dealership from tying up capital, but this cost is generally not included in a department's costs. Customer fi-

nancing also generally is not included in the department's revenue calculations, which are simply sales of cars. Thus, in both cases we see again that the second and third formula bonuses are apparently used to address weaknesses in the first formula bonus.

Properties of Performance Measures. The survey included questions to assess five properties of each performance measure:

"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?"

Responses were recorded on a scale from 1 (Not at All) to 5 (Very High). These five properties are the focus of our analyses in this paper.

The first of these properties (factors outside your control) clearly is a proxy for uncontrollable risk, whereas the second property (reflects overall performance) rather is a 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 as recently developed in the theoretical literature.

The third and fourth properties (causes focus on short-term goals; encourages cooperation with other departments) reflect two of the most common distortions caused by accounting performance measures. In fact, we chose to ask explicitly about these, rather than asking generic questions about distorted performance measurement, because our discussions with industry experts and managers indicated that these are common distortions in this particular industry. By their nature, accounting measures emphasize short-term incentives because they focus on a given period, are backward looking, and tend to ignore investments in intangibles. Similarly, the effects of an employee's cooperation may not show up directly in accounting measures, especially ones that are measured at the level of the manager's own department.

The final performance measure property is a proxy for 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 handled with complete confidence, and sent directly to us (not the consulting firm), which may explain the willingness of managers to answer this question. An alternative explanation provided by some of the industry experts that we spoke to is that, particularly in a sales-oriented dealership environment, manipulation is common and taken for granted; it is simply an accepted cost of imperfect performance measurement.

The first, third, and fifth performance measure properties (reflects factors outside the manager's control; emphasizes short term; manipulation) listed above take larger values if the performance measure is "worse." For example, the more that the measure emphasizes short-term goals, the larger should be the respondent's answer for part 3 of this survey question. By contrast, the second and fourth properties (reflects overall performance; encourages cooperation) take larger values if the performance measure is "better." In order to make the presentation of results easier to interpret, the first, third and fifth properties are *reverse coded* in all analyses in this paper. In other words, all performance measure properties are scaled so that a larger value indicates a better performance measure.

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. Recall that these variables are measured on a scale of 1-5, and that a larger number indicates a "better" performance measure. 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" and "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, a measure is very difficult to manipulate if it represents performance of a different department, and manipulation is easiest at the department level than at the level of the dealership as a whole. The one performance measure property that does not seem to have expected pat-

terns across organizational units is the first, the extent to which the measure reflects factors outside the manager's control. This is reported to be highest, or best (least reflecting factors outside the manager's control) when it measures another department, although the magnitude of the difference of this seemingly inconsistent finding is small compared to the differences where this performance measure property does accord with expectations.

Explicit Incentives. One potential measure of incentive strength is the commission rate on the bonus plan. However, there are several practical difficulties. Contracts use different measures that are not comparable across departments or dealerships. Even when dealerships use the same nominal measure, there is variation in accounting methods across dealerships. Contracts also often 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. For these reasons, we use the total bonus as a proxy for the strength of the incentive. Effort, and thus expected performance, should be positively related to the size of the incentive. Thus, total bonus is a proxy for the strength of the incentive and 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. We can infer eligibility for formula bonuses, even if the manager did not earn that bonus because of low performance, from whether or not they answered the survey questions about performance measure properties.

Implicit Incentives. A feature of the survey is that it provides information on implicit incentives that are almost never studied in economics or accounting. For each performance 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). In addition, respondents also reported the size of their discretionary bonus when applicable.

Controls. The regressions include a variety of controls. First, when the job is more complex and intangible it will be harder to measure performance accurately. Placing a strong incentive on a performance measure in such a setting would impose more risk on the manager. For this reason, we predict that indicators that the job is more complex will have negative effects on incentive intensity. We use two measures. First, most regressions include dummy variables for whether a department manager is a service department manager (instead of a new or used car sales department manager), to control for broad differences in job design. Service department jobs are more complex and involve more tasks for which performance is difficult to quantify. This may mute overall explicit incentives, if performance measures do not adequately weight hard-to-measure tasks (Holmstrom & Milgrom 1991; Slade 1996).

Our second indicator for a job with more intangible components is a measure of the emphasis placed on customer service at the department (for department managers) or dealership (for general managers) level (see Appendix). Customer service has many dimensions compared to simple sales units or revenue, and most of those dimensions are intangible. The measure of emphasis on customer service is based on several survey questions that assess the extent of dealership/ department focus on price vs. customer service. A factor analysis of these survey questions resulted in empirical scales that match the customer service concept as expected.

We also include a measure of the degree of competition (see Appendix). If the work environment (marginal product of effort) is stochastic, and the employee observes signals about the state of the world after the contract is written but before choosing actions, the firm will want to provide some incentives for the employee to exploit this specific knowledge to the advantage of the firm (Raith 2003). Therefore, we expect that employees will be given stronger incentives in more competitive environments. Evidence for this effect would be evidence in favor of the idea that greater controllable risk implies stronger incentives.

Finally, theory predicts that incentives should be stronger, the larger is the marginal product of effort. As proxies for this, we include the number of employees reporting to the manager (a measure of re-

sources under the manager's control), the manager's experience in the position (a measure of human capital), and a dummy variable for general managers (instead of department managers). We predict that these will be positively related to the strength of incentives.

6. FINDINGS

Table 6 presents analysis of the 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 explicit (formula-based) bonuses (including cases where the bonus was negative) for the full sample, and also for general managers and department managers separately. The Tobits 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 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, both are always positive, and usually significant. Thus our evidence is consistent with Prendergast's (2002) analysis of risk and incentives, and one of the first since then (see DeVaro and Kurtulus (2006) for an earlier and more thorough analysis of this question) to find a positive relationship between strength of incentives are degree of performance measure precision (or negative relationship between incentive strength and uncontrollable risk).

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

[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 in-

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 emphasis short term financial results (such as sales), perhaps because of the terms of their economic relationship 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. Thus, narrower performance measures do appear to distort incentives away from cooperation, and are weighted accordingly.

The final performance measure property is the extent to which it is unlikely to be manipulated by the manager to improve measured performance. Once again, in all three regressions this property has a significant effect, in the predicted direction, on the strength of incentives. This provides evidence that managers may, in fact, attempt to manipulate their evaluations. For this to be possible, managers must have some specific knowledge (controllable risk) in performing their jobs so that they can use this knowledge to manipulate the measure without perfect knowledge of that by the evaluator. Thus, our evidence that manipulation occurs and is factored into incentives is additional evidence for Prendergast's view that managers have asymmetric information about how they perform their jobs, and that 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. Depending on the production function, greater human capital may imply a larger marginal product of effort. The positive coefficients on experience suggest that this may be the case in auto dealerships.

---

centive contracts (performance measures and formulas) for managers in the same dealership, perhaps because they run different types of departments.

Degree of competition is another proxy for controllable risk, following Raith (2003). Competitive actions by other dealerships are a kind of risk that is in many ways controllable by the firm's managers, since they can respond with their own actions. Thus, we predict a positive coefficient on our measure of competition, which we find in all three regressions.

Emphasis on customer service measures the importance of a key intangible. For a similar reason, a dummy variable is included for whether or not the manager runs the service department. Interactions with customers in sales departments largely focus on fixed car options, and haggling over prices. Customer interactions in service departments are much more complex and likely to involve long-term relationships with the dealership. Such considerations can be difficult to quantify. If intangibles are important to the dealer, then multitask incentive theory (Holmstrom & Milgrom 1991) predicts that numeric measures will be given less weight for explicit incentives, since they almost always under-emphasize intangibles. Indeed, that is what we find in all three regressions for the dealer's emphasis on customer service, and for the service department in the department manager sub-sample. These results provide evidence that an important weakness of performance measures is that they may not proxy adequately for intangible factors that affect firm value. When that is the case, and such factors are important, implicit incentives and subjective performance evaluations are likely to take on more importance in the incentive system.

In summary, Table 6 provides relatively 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 now examines the effect of our five performance measure properties on implicit incentives.

In Table 7, the dependent variables 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.[5] Implicit incentives of this form have received very little analysis in economics (Gibbs, Merchant, Van der Steve & Vargus 2004, Murphy & Oyer 2003).

There are at least two potential roles for implicit incentives such as promotions. One is to reduce risk: to use ex post information to filter out uncontrollable events from the overall evaluation. Numeric performance measures are affected by events beyond the manager's control (and beyond the manager's ability to react). Since numeric measures are set ex ante, they cannot filter out such events with ex post knowledge. A supervisor may need to implicitly remove the effects of "bad luck" on the employee's overall evaluation, or on the employee's overall reward or punishment. We would expect such a role of implicit incentives to be relevant for the first two performance measure properties, which proxy for controllable and uncontrollable risk. If the performance measure is relatively unaffected by uncontrollable risk, or relatively affected by controllable risk, then poor measured performance should be *more* likely to be punished implicitly. For that reason, we predict a positive coefficient on the first two properties.

The second role for implicit incentives is to punish, and thereby deter, distorted efforts or manipulation of metrics. This role fits the wording of the survey questions most directly. We therefore predict that measures that are less likely to distort (toward short term effort, or away from cooperation) are *less* likely to be used to punish poor numeric performance. We also predict that a measure that is less likely to be manipulated is less likely to be punished for low measured performance. In fact, in the theory discussion we noted that the most likely way to deter manipulation is through ex post punishment, so that potential manipulation of a measure is particularly likely to be strongly related to implicit incentives. Thus, the predicted coefficients on the last three performance measure properties are negative.

[5] Table 7 uses observations on properties of the performance measure for the first formula bonus only, to limit the Probits to a single observation for each manager, since the dependent variable has only a single value for each manager. Including the second and third performance measure's properties has no significant effect on conclusions. We also ran the order Probits with the full controls from Table 6, again with no effect on conclusions about the effects of the five performance measure properties.

The results in Table 7 are mixed. The only performance measure factors that have the predicted coefficients and are statistically significant are the second and fifth. The more that a measure reflects the manager's overall performance, the more likely is a low value of that performance measure to be punished implicitly. We have interpreted this property as a potential proxy for controllable risk, but with strong 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 a low value of 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 new 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.

The last table tests the idea that additional performance measures can be used to rebalance incentives from the first performance measure (Feltham & Xie 1994; Baker 2002). To test this, we measured the five performance measure properties of the second or third performance measure *relative to* the value of that property for the first measure, by subtracting the value for the first measure. Thus, a larger value of this relative measure means that the second or third performance measure is reported to be relatively better along that dimension than is the first performance 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 regression 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.

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. Thus, the table provides some preliminary evidence in favor of the idea that performance measures are chosen as a complementary set. However, more work with other datasets would be required to give this conclusion greater confidence.

7. CONCLUSIONS

In this paper we use data from a survey that we designed and collected to provide a relatively comprehensive study of performance measure properties and incentive plan design. Prior empirical work has tended to focus on the tradeoff between risk and incentives, but has often failed to find the predicted relationship. We do find such a relationship, and in the process present evidence supporting the more recent 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. Finally, we study two topics that have been almost ignored in the empirical literature: implicit incentives, and interactions between incentive instruments designing reward schemes.

We studied five properties of performance measures. Two proxy (though imperfectly) for uncontrollable and controllable risk. Two measure classic distortions, short term focus and lack of cooperation. The last measures the metric's tendency to motivate manipulation. We also added a control for the importance of competition to the business environment, in a further attempt to assess the importance of controllable risk on incentives. Finally, we added two measures of the importance of intangibles (customer service, and type of department) to assess their effect on the use of performance metrics for incentives.

Our results are broadly consistent with predictions for the first two properties, the extent to which the measure reflects the manager's overall performance, and reflects factors outside the manager's control. This is evidence for the importance of controllable and uncontrollable risk. This conclusion is bol-

stered by our finding that greater competition leads to stronger incentives, suggesting that managers are then given stronger incentives to react to competitive challenges as they arise.

Furthermore, one of our most consistent findings is that potential manipulation of the measure is a significant factor in the measure's use for incentives. Since manipulation requires private information on the part of the employee, that strongly suggests that dealership managers do have leeway to address controllable risk in performing their jobs.

Our results also suggest that performance measure properties affect not only explicit incentives, but also implicit incentives. Since they are important indicators of success, this should not be surprising. However, implicit incentives have been given relatively little attention in either theoretical or empirical work in economics and accounting. Our results suggest that more attention to such considerations in incentive provision is warranted.

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, thus improving the overall incentive system. Our results indicate that this is particularly useful for deterring manipulation. Our companion paper on subjective evaluations, together with this paper, provides a relatively complete analysis of the entire incentive system used for managers in auto dealerships.

We provide evidence that dealerships choose additional performance measures in part based on their properties relative to the measure that is given greatest weight. This makes sense, since the overall evaluation will be less distorted, less subject to manipulation, or less risky.

Overall, the results presented show that performance measure properties play a crucial and complex role in incentive plan design. Since the empirical literature on this topic is relatively small, much more work might be done with suitable data. Finally, the interrelationships between different bonuses, their performance measures, and explicit and implicit incentives illustrate that an incentive plan is a system that is designed together. Most empirical work focuses on specific parts of an incentive system. Once

more, we expect that a great deal of interesting empirical work can be done on how the parts work together.

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): 728-752.

_____ & Bjorn Jorgensen (2002). "Turbulence, Risk, and Incentives." Working Paper, Harvard Business School.

_____, Robert Gibbons & Kevin J. Murphy (1994). "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109 (4): 1125-1156.

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

Bushman, Robert, Raffi Indjejikian & Abbie Smith (1996). "CEO Compensation: The Role of Individual Performance Evaluation." *Journal of Accounting and Economics* 21 (April): 161-193.

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

_____ (forthcoming). "A General Test of Gaming." *Review of Economics & Statistics.*

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

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

Dillman, Don A. (1978). *Mail and Telephone Surveys: The Total Design Method*. NY: Wiley.

Feltham, Gerald A. & Jim Xie (1994). "Performance Measure Congruity and Diversity in Multi-task Principal/ Agent Relations." *The Accounting Review* 69(3): 429-453.

Gibbs, Michael, Kenneth A. Merchant, Wim A. Van der Stede & Mark E. Vargus (2004). "Determinants and Effects of Subjectivity in Incentives." *The Accounting Review* 79(2): 409-436.

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

Hemmer, T. 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.

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

Ichniowski, Casey, Kathryn Shaw & 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. & David F. Larcker (2002). "Determinants of Performance Measure Choices in Worker Incentive Plans." *Journal of Labor Economics* 20(2): S58-S91.

_____, _____ & Madhav V. Rajan (1997). "The Choice of Performance Measures in Annual Bonus Contracts." *The Accounting Review* 72(2): 231-255.

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

Murphy, Kevin J. & Paul Oyer (2003). "Discretion in Executive Incentive Contracts." Working paper, USC.

Prendergast, Canice (2002). "The Tenuous Tradeoff between Incentives and Risk." *Journal of Political Economy* 110(5): 1071-1102.

Raith, Michael (2003). "Competition, Risk, and Managerial Incentives." *American Economic Review* 93:1425-1436.

_____ (2005). "Specific Knowledge and Performance Measurement." Working paper, Simon School of Management, University of Rochester.

Slade, Margaret (1996). "Multitask Agency and Contract Choice." *International Economic Review* 37(2): 465-486.

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

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%
	Spiffs		8%	16%	10%	32%
% 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	9,866
		3	48,633	9,197	12,099	6,579
	Discretionary Bonus		36,449	20,135	13,295	10,728
	Spiffs		9,174	4,239	2,190	3,427

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.

Table 2.
Correlations of Pay Instruments

		Salary	Formula Bonus			Discretionary Bonus
			1	2	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	
Spiffs		0.04	0.03	0.03	-0.02	0.06

Notes: Correlations of dollar values of pay instruments, calculated in each case across all available observation pairs with non-missing values.

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 threshhold before any bonus is paid; and a cap if no bonus is paid for performance above some threshhold.

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 used for formula bonuses 1-3, shows % 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 measure 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.

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. 3 of the 5 properties were then reverse coded; see the text.

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)	+	8,151	4,633	**	11,238	11,736		4,233	2,125	**
	Reflect overall performance	+	12,612	4,578	***	33,179	14,928	***	2,991	2,039	*
	Causes short term focus (reverse coded)	+	-4,600	3,821		3,257	9,027		-4,836	1,643	
	Encourages cooperation	+	11,257	4,172	***	27,357	14,928	**	4,797	1,726	***
	Motivates manipulation (reverse coded)	+	8,795	3,214	***	16,009	9,027	**	4,480	1,431	***
Job & manager characteristics	# of employees	+	128	215		23	428		390	139	***
	Degree of competition	+	15,512	6,193	***	71,583	23,383	***	3,482	2,504	*
	Emphasis on customer service	–	-16,857	8,324	**	-55,588	23,812	***	-4,928	3,682	*
	Experience	+	3,026	783	***	6,824	2,465	***	831	337	***
	General Manager	+	64,150	11,240	***						
	Service Department manager	–	-8,396	12,185					-12,858	4,853	***
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%; ** = 5%; * = 10%. Predicted signs of coefficients are shown after variable names; 1-tailed tests in those cases. The first 5 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 constructured from several survey questions using factor analysis (see Appendix A).

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 (PM1) properties	Reflects factors outside mgr.'s ctl. (reverse coded)	+	-0.147	0.048	-0.091	0.048	-0.006	0.048	-0.121	0.048
	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. > chi²			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%; ** = 5%; * = 10%. Predicted signs of coefficients are shown after variable names; 1-tailed tests in those cases.

Table 8.
Effects of Performance Measure Properties on Other Formula Bonuses

		Pred. sign	Formula Bonus 2 or 3		
Intercept			4,027	2,094	**
Property of PM2 or PM3 Minus Property of PM1	Reflects factors outside mgr.'s control (reverse coded)	+	201	1,522	
	Reflects overall performance	+	1,633	1,632	
	Causes short term focus (reverse coded)	+	86	1,436	
	Encourages cooperation	+	4,046	1,401	***
	Motivates manipulation (reverse coded)	+	2,527	1,628	**
General Manager			2,844	4,225	
Service Department manager			-6,979	3,120	***
N			315		
% Bonus (#2 or 3) > 0			60%		

Notes: Tobit predicting magnitude of Formula Bonuses 2-3. SE = standard error. *** = significant at 1%; ** = 5%; * = 10%. Predicted signs are shown after variable names; 1-tailed tests in those cases.

Appendix.
Description of Factor Variables

Survey Questions Used to Construct Factors		Factor Loadings (Cronbach α)
Perceived Degree of Competition		(α=.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)		(α=.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)		(α=.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 (δ = 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 Alphas are highly adequate ($\alpha > 0.70$).