On the Use of Customized vs. Standardized Performance Measures*

Anil Arya

Ohio State University

Jonathan Glover

Carnegie Mellon University

Brian Mittendorf

Yale School of Management

Lixin Ye

Ohio State University

May 2005

* We thank Shane Dikolli, Ron Dye, John Fellingham, Joan Luft (Editor), Doug Schroeder, Shyam Sunder, Jacob Thomas, Rick Young, and two anonymous referees for helpful comments. Anil Arya gratefully acknowledges assistance from the John J. Gerlach Chair.

Send correspondence to Anil Arya, 2100 Neil Avenue, Columbus, Ohio 43210.

Telephone: (614) 292-2221. Email address: arya@cob.osu.edu.

On the Use of Customized vs. Standardized Performance Measures

Abstract

Despite the influx of measures which can be customized to the demands of each business unit (e.g., customer satisfaction surveys and quality indices), many firms have been dogged in their reliance on standardized measures (e.g., conventional financial metrics) in performance evaluation. In this paper, we consider one justification: though customized measures may more accurately target the goals of a particular unit, standardized measures may offer more meaningful opportunities for relative performance evaluation. Standardized measures have a commonality in errors which is naturally absent among measures targeted to each circumstance. This commonality allows learning about one measure from another and, thus, the construction of more efficient proxies for unobservable employee inputs. The use of comparative evaluation schemes is not without its challenges, since it may induce unwanted coordination by those being evaluated. Even with such gaming concerns, standardized measures can still be preferred, but the requirements are more stringent.

Keywords: *Incentives; Measurement system; Relative performance evaluation; Tacit collusion*

INTRODUCTION

Recent years have seen a surge in the availability of nontraditional performance measures (customer satisfaction surveys, quality indices, etc.). While traditional accounting metrics are generated using a set of widely-accepted standards, these new state-of-the-art measures promise customization so that incentives can be better targeted to the needs of each business unit. Nonetheless, many firms have been dogged in their maintained reliance on standardized measures, such as conventional financial metrics, in performance evaluation (e.g., Ittner and Larcker 1998).

In this paper, we consider one justification for the seemingly anachronistic mindset of many evaluation schemes: though customized measures may more accurately target the goals of each business unit, a standardized set of measures may offer more meaningful opportunities for relative performance evaluation. Standardized measures naturally share common (systematic) errors. This commonality in errors allows learning about one employee's measure from another and, thus, the construction of more efficient proxies for unobservable employee inputs. The use of comparative evaluation schemes is not without its challenges, since it may induce unwanted coordination by those being evaluated. But, such challenges are surmountable.

Our model considers a simple scenario: a firm with two divisions is unable to costlessly make use of all potential performance measures and must choose a parsimonious set of two performance metrics. If the firm has one measure available for one division, we ask what the firm's preferred second measure is. The firm can opt for a measure which is customized in that it perfectly tracks the second division's effort. Or, the firm may emphasize standardization, relying on a metric that is measured in a manner similar to that used for the first division.[1] While standardization comes with the downside of having a less targeted measure for the second division, it also means that errors in the two measures will naturally be linked. As a result, the standardized measures lend themselves to more effective comparative evaluation schemes.

With this issue as a backdrop, we provide conditions under which employing standardized measures is preferred to using customized measures. The decision rule is crisp: if the correlation between standardized measures is large enough, standardized measures are preferred due to their relative performance evaluation benefits. Even though standardized measures are less precise, they are more informative when both divisions' incentives are in view.[2] In other words, when the firm must rely on one imperfect measure, perfection is not necessarily the goal of the second measure since blemishes in the second measure may help clean up problems with the first.

The use of standardized metrics for relative performance evaluation comes with an added cost, the potential for tacit collusion. When a division's performance is evaluated relative to its peers (rather than comparison to a mechanistic index), peer divisions may implicitly coordinate in "lowering the bar." In particular, in our setting, the optimal performance evaluation contract based on standardized measures introduces a second equilibrium in the employees' subgame wherein both divisions shirk their responsibilities (as in Demski and Sappington 1984). The equilibrium is especially pressing since it provides each employee with a higher expected utility than the equilibrium the firm prefers.

To overcome this concern of gaming with standardized measures, the firm can write a contract that induces dominant strategy incentives for one division, thereby ensuring acquiescence by the other.[3] Under such an approach, standardized measures can still be preferred to customized measures, but the correlation requirement is more stringent. This tact also requires the firm to decide which division is provided with dominant strategy incentives. With ex ante identical divisions, the decision is a toss up. However, if the two divisions share different characteristics or beliefs, one may be a better candidate than another.

In particular, a firm may assign the dominant strategy contract to the division whose beliefs are harder to discern. This is because the dominant strategy contract has an added feature of being effective no matter what the division believes about the correlation in the measures. We are unaware of another paper on multi-agent moral hazard that makes the observation that uncertainty about beliefs may justify assigning a dominant strategy contract. We conjecture that a feature of our model in driving this result is that the agents' measures depend on each other only through the correlation in errors–the production technologies are individual (depend on each agent's effort) rather than joint (depend directly on both agents' efforts).

While the result here is derived in a simple two division, two performance measure setting, the idea that relative performance evaluation may justify standardization of performance measures presumably has broader implications. As an example, in university admissions, the debate over standardization is hotly contested. While a student's entire admissions package can be useful in evaluating ability, admissions tend to rely heavily on standardized test results due to their inherent comparability. Even if a standardized test is an imperfect measure of performance, its measurement errors are common among all applicants. The same cannot be said of GPAs, since student GPAs emanate from a variety of testing and grading environments. For this reason, the College Board has been steadfast in its insistence on following rigid guidelines in administering the SATs (*WSJ* 11/20/02).

As an example of this issue in a more conventional business context, consider the case of the Balanced Scorecard. In implementing the Balanced Scorecard, many firms place a heavy reliance on measures that are common among divisions (standardized) rather than picking measures unique to each division. And, since common measures tend to be backward-looking financial metrics while unique measures tend to be forward-looking nonfinancial indicators, the concern is that many scorecards in reality lack balance (e.g., Lipe and Salterio 2000). The results here suggest that the news may not be so grim. A somewhat unbalanced reliance on common measures may simply reflect the importance of removing common noise from measures so as to make them better proxies for unobservable employee actions.

On a related note, the use of a balanced scorecard is predicated on the use of a limited number of measures. This idea underlies our model also. While theory suggests that additional measures cannot hurt, practical considerations place a cap on the number of performance metrics that can be used by a firm. And, since vast amounts of information are costly to generate and difficult to process, effective performance measurement is often an exercise in efficiently condensing information. As Demski (1994, 6) puts it, "predigested, codified, and summarized presentations are the norm." This paper highlights the care needed in such choices, since the selection of one measure may affect the allure of another.

The role of performance measures in single-agent, multi-task settings has been extensively studied (e.g., Holmstrom and Milgrom 1991; Feltham and Xie 1994). In such settings, the optimal use of a performance measure takes into account two factors: the link between the measure and firm objectives (congruity) and the measure's noise. The ground swell of support for more targeted performance measures comes primarily on the first dimension. For example, using a mix of lag measures and leading indicators can provide timely feedback that more closely aligns agent incentives with a firm's long-term objectives (Dikolli 2001; Hemmer 1996). Moreover, such timeliness can improve evaluation when contractual concerns such as the potential for renegotiation preclude effective use of delayed measures (Sliwka 2002).

In our setting with multiple agents, the deciding factor in picking a set of performance measures is not congruity with firm objectives or even measure noise (a customized measurement system wins out on both counts). Rather, the crucial factor is how a measure's noise covaries with that of other measures when strategic interactions and parsimony in measurement are critical. Since the presence of one measure may introduce a complementarity with another, performance measurement design requires a holistic view of an organization, one that focuses on potential interactions among participants.

MODEL

A firm (principal) hires two employees (agents) to assist in operations. Each agent privately and opportunistically chooses the supply of his input (effort). Denote Agent i's effort by e^i, $e^i \in \{e_L,e_H\}$ and $i \in \{1,2\}$. While the principal cannot directly observe the agents' efforts, she can design a performance measurement system to motivate the appropriate efforts. Here, the principal has two options.

One possibility is for the principal to install a performance measurement system that relies on standardized measures. This corresponds to tracking x^1 and x^2, which are imperfect proxies for the agents' inputs. In particular, x^i, $x^i \in \{x_L,x_H\}$, depends not only on e^i but also on θ^i, $\theta^i \in \{\theta_L,\theta_H\}$, the productivity parameter of the environment in which Agent i operates. The precise relationship between the measures, agents' efforts, and productivity parameters is as follows: $x^i = x_H$ if and only if $e^i = e_H$ and $\theta^i = \theta_H$, i.e., Agent i's measure is high if and only if he exerts high effort and his productivity parameter turns out to be favorable.

Though θ^1 and θ^2 are unobservable, their probability distributions are common knowledge. The (marginal) probability $\theta^i = \theta_L$ is p, $0 < p < 1$, and the (conditional) probability $\theta^j = \theta_L$ given $\theta^i = \theta_L$ is r, $p \leq r \leq 1$. In effect, r represents the correlation in the agents' environments: $r = p$ reflects no correlation and $r = 1$ is perfect correlation.

The principal's second option is to customize the performance measure for Agent 2. In this case, the measurement system tracks x^1 and a second, distinct measure y^2. Consistent with the idea of customized measures being more congruent with specific agent actions and to seemingly stack the deck against relying on standardized measures, we assume y^2, $y^2 \in \{y_L,y_H\}$, is actually a perfect proxy for Agent 2's effort: $y^2 = y_k$ if $e^2 = e_k$.

Of course, the principal would prefer to track all three measures, x^1, x^2, and y^2. In practice, however, the costs of generating and processing information often place limits on the number of measures that are tracked and included in contractual relationships. The performance measurement choices described above seek to reflect such a tension by restricting the principal to picking only two of the three possible measures.

In short, the principal's choice in our setting entails selecting a performance measure for Agent 2, wherein the choice of x^2, an imperfect proxy for e^2, corresponds to picking standardized measures and the choice of y^2, a perfect proxy for e^2, corresponds to picking customized measures. Table 1 summarizes the probability distributions induced by the agent's actions and the principal's measurement system.[4]

Insert Table 1 here.

To highlight the incentive problem, we assume the principal wants to motivate each agent to choose e_H. The timeline of events is as follows. First, the principal chooses her performance measures and offers contracts to the agents. The contracts stipulate payments to each agent conditioned on observations from the chosen measurement system. With standardized measures, Agent i's payments are $s^i(x^1,x^2)$; with customized measures, his payments are $s^i(x^1,y^2)$. Second, each agent privately chooses his effort. Finally, performance measures are observed, and each agent is paid according to the contract.

The agents are risk and effort averse. Given payment, s^i, and effort, e^i, Agent i's utility is $U^i(s^i,e^i) = u(s^i) - v(e^i)$, where $u:R \rightarrow R$, $u'(s^i) > 0$, $u''(s^i) < 0$, and $v(e_H) > v(e_L) = 0$. The principal is risk-neutral. In designing the contract, the principal's goal is to minimize expected compensation subject to the following constraints. The individual rationality constraints, (IR^i), require that the contract provides each agent with at least his reservation utility, $\overline{U}$. The incentive compatibility constraints, (IC^i), require that each agent choosing $e = e_H$ is a (Nash) equilibrium in the agents' subgame. Programs (P^S) and (P^C) present the principal's contracting problems under standardized measures and customized measures, respectively.

Program (P^S)

$$\underset{\{s^i(x_j,x_k)\}_{i=1,2}}{\text{Min}} \sum_{j,k=L,H} \Pr(x_j,x_k|e_H,e_H)[s^1(x_j,x_k) + s^2(x_j,x_k)]$$

s.t.

$$\sum_{j,k=L,H} \Pr(x_j,x_k|e_H,e_H)u(s^i(x_j,x_k)) - v(e_H) \geq \overline{U} \qquad i = 1,2 \qquad (IR^i)$$

$$\sum_{j,k=L,H} \Pr(x_j,x_k|e_H,e_H)u(s^1(x_j,x_k)) - v(e_H) \geq \sum_{j,k=L,H} \Pr(x_j,x_k|e_L,e_H)u(s^1(x_j,x_k)) \qquad (IC^1)$$

$$\sum_{j,k=L,H} \Pr(x_j,x_k|e_H,e_H)u(s^2(x_j,x_k)) - v(e_H) \geq \sum_{j,k=L,H} \Pr(x_j,x_k|e_H,e_L)u(s^2(x_j,x_k)) \qquad (IC^2)$$

Program (P^C)

$$\underset{\{s^i(x_j,y_k)\}_{i=1,2}}{\text{Min}} \sum_{j,k=L,H} \Pr(x_j,y_k|e_H,e_H)[s^1(x_j,y_k) + s^2(x_j,y_k)]$$

s.t.

$$\sum_{j,k=L,H} \Pr(x_j,y_k|e_H,e_H)u(s^i(x_j,y_k)) - v(e_H) \geq \overline{U} \qquad i = 1,2 \qquad (IR^i)$$

$$\sum_{j,k=L,H} \Pr(x_j,y_k|e_H,e_H)u(s^1(x_j,y_k)) - v(e_H) \geq \sum_{j,k=L,H} \Pr(x_j,y_k|e_L,e_H)u(s^1(x_j,y_k)) \qquad (IC^1)$$

$$\sum_{j,k=L,H} \Pr(x_j,y_k|e_H,e_H)u(s^2(x_j,y_k)) - v(e_H) \geq \sum_{j,k=L,H} \Pr(x_j,y_k|e_H,e_L)u(s^2(x_j,y_k)) \qquad (IC^2)$$

Denote the solution to (P^S) and (P^C) by $s^{*i}(x^1,x^2)$ and $s^{*i}(x^1,y^2)$, respectively.

RESULTS

In the setup, standardized measures seemingly have little to offer. After all, the standardized measure x^2 is an imperfect proxy, while the customized measure y^2 is a perfect proxy, for Agent 2's act. Despite this advantage of y^2, there may still be justification for using x^2 based on its ability to help evaluate Agent 1.

If the two agents' environments are uncorrelated, there is clearly no advantage to using x^2 in Agent 1's evaluation. However, correlation between the agents' environments introduces a spillover from Agent 2's measure to Agent 1's evaluation. In fact, the greater the correlation, the more useful this relative performance evaluation spillover becomes. The value of correlation in relative performance evaluation is confirmed in the following Lemma.

Lemma. With standardized measures, expected compensation is decreasing in r.

Proof.

Given ex ante symmetric agents and symmetric performance measures, we solve (P^S) only for one agent's, say Agent 1's, payments. Also, for simplicity, we write the payment $s^1(x_j,x_k)$ as s_{jk} and $s^{*1}(x_j,x_k)$ as s^*_{jk}. The principal's program for Agent 1 can be rewritten as:

$$\min_{s_{jk}} \; f(s^1(x^1,x^2);r)$$

s.t.

$$g(s^1(x^1,x^2);r) \geq \overline{U} \qquad (IR^1)$$

$$h(s^1(x^1,x^2);r) \geq 0 \qquad (IC^1)$$

From Table 1:

$f(s^1(x^1,x^2);r) = prs_{LL} + p(1-r)s_{LH} + p(1-r)s_{HL} + (1-p-p(1-r))s_{HH}$,

$g(s^1(x^1,x^2);r) = pru(s_{LL}) + p(1-r)u(s_{LH}) + p(1-r)u(s_{HL}) + (1-p-p(1-r))u(s_{HH}) - v(e_H)$, and

$h(s^1(x^1,x^2);r) = g(s^1(x^1,x^2);r) - [pu(s_{LL}) + (1-p)u(s_{LH})]$.

As is common in moral hazard models, for any $r < 1$, both (IR^1) and (IC^1) bind at the optimal solution. Denoting the positive multipliers on (IR^1) and (IC^1) by λ and μ, respectively, the first-order conditions for the program are:

$$\frac{1}{u'(s_{LL})} = \lambda - \mu\frac{1-r}{r}$$

$$\frac{1}{u'(s_{LH})} = \lambda - \mu\frac{1-p-p(1-r)}{p(1-r)}, \text{ and}$$

$$\frac{1}{u'(s_{HL})} = \frac{1}{u'(s_{HH})} = \lambda + \mu.$$

The first-order conditions imply $s^*_{HL} = s^*_{HH} > s^*_{LL} \geq s^*_{LH}$ (with equality iff r = p).

Denote the optimal objective function value, the minimum expected compensation, by F. Then, by the envelope theorem:

$$\frac{dF}{dr} = \frac{\partial f(s^{*1}(x^1,x^2);r)}{\partial r} - \lambda\frac{\partial g(s^{*1}(x^1,x^2);r)}{\partial r} - \mu\frac{\partial h(s^{*1}(x^1,x^2);r)}{\partial r}.$$

Using the expressions for f, g, and h provided earlier, and setting $s^*_{HL} = s^*_{HH}$, yields:

$$\frac{dF}{dr} = p[s^*_{LL} - s^*_{LH} - (\lambda+\mu)(u(s^*_{LL}) - u(s^*_{LH}))].$$

Hence, $\frac{dF}{dr} < 0$ if $s^*_{LL} - s^*_{LH} - (\lambda+\mu)(u(s^*_{LL}) - u(s^*_{LH})) < 0$, or $\frac{u(s^*_{LL}) - u(s^*_{LH})}{s^*_{LL} - s^*_{LH}} > \frac{1}{1+\mu}$.

To confirm the last inequality, note that the first-order conditions stipulate $s^*_{LL} \geq s^*_{LH}$ and thus, by the concavity of u, $\frac{u(s^*_{LL}) - u(s^*_{LH})}{s^*_{LL} - s^*_{LH}} \geq u'(s^*_{LL})$. Further, by the first-order conditions, $u'(s^*_{LL}) > u'(s^*_{HH}) = \frac{1}{\lambda+\mu}$. ■

Another way of demonstrating the monotonicity result is to argue that higher values of r introduce more dispersion in the likelihood ratios for the relevant acts. In other words, check Kim's (1995) condition for information system ranking in single-agent settings. To do so, define the likelihood ratio $z(x_j,x_k;r) = 1 - \frac{\Pr(x_j, x_k \mid e_L, e_H)}{\Pr(x_j, x_k \mid e_H, e_H)}$, and let L(z;r) denote the distribution function over the ratio under the act combination (e_H,e_H).

Using Table 1, and listing z in ascending sequence, z takes on the values $1 - \frac{1-p}{p(1-r)}$, $1 - \frac{1}{r}$, and 1 with corresponding probability p(1-r), pr, and 1-p, respectively. Note two properties. First, the likelihood ratio is mean zero for all r. Second, for $r_2 > r_1$, $\int_{-\infty}^{t} L(z;r_2)dz \geq \int_{-\infty}^{t} L(z;r_1)dz$ for all $t \in R$, with the inequality strict for $t \in (1 - \frac{1-p}{p(1-r_2)}, 1 - \frac{1}{r_2})$. Hence, the likelihood ratio distribution for r_2 is a mean preserving spread of that for r_1, thus satisfying Kim's condition. Intuitively, the r_2-system provides more effective information for the principal by offering a more dispersed set of data.[5]

Under standardized measures, Agent i's moral hazard problem is softened by higher r values due to the spillover from x^j. In contrast, under customized measures, r has no effect on either agent's contract. These forces can make standardized measures strictly preferred, as stated in the first proposition.

Proposition 1.

There exists an interior cutoff r^* such that:

(i) for $r < r^*$, the principal (strictly) prefers customized measures.

(ii) for $r > r^*$, the principal prefers standardized measures.

Proof.

The proof follows from three observations.

1) In an uncorrelated environment (r = p), there is no reason for relative performance evaluation and, hence, customized measures are strictly preferred. The absence of relative performance evaluation can be confirmed by the first-order conditions in the Lemma that

stipulate the single-agent solution: $s_{HL} = s_{HH}$ and $s_{LH} = s_{LL}$ at r = p under standardized measures. Hence, expected compensation for Agent 1 is equivalent under standardized and customized measures. Further, since y^2 is clearly more informative of Agent 2's effort than x^2, expected compensation for Agent 2 is strictly less under customized measures.

2) In a perfectly correlated environment (r = 1), standardized measures are strictly preferred since, with such measures, relative performance evaluation benefits allow the principal to achieve the first-best solution for each agent. The first-best for Agent 1 is accomplished by setting $s_{LL} = s_{HH} = u^{-1}[v(e_H) + \overline{U}]$, and $s_{LH} = s_{HL} = M$, where the off-equilibrium payment M is chosen sufficiently small to satisfy the (IC^i) constraints. Under customized measures, however, first-best expected compensation is only possible for Agent 2 (the noisy proxy x^1 is all that is available to motivate Agent 1).

3) From the Lemma, expected compensation under standardized measures is decreasing in r. In contrast, expected compensation under customized measures is unaffected by r (see Table 1). This, coupled with the above two observations, yields the cutoff property of the proposition and ensures the cutoff is interior. ■

Note, the main advantage to using standardized measures is exploiting relative performance evaluation. Specifically, under the solution to (P^S), an agent is punished if the other agent's outcome is greater than his. This interlinkage in agent incentives serves to ease each agent's moral hazard problem. However, a new problem of multiple equilibria in the agents' subgame (tacit collusion) arises.

The presence of multiple equilibria is particularly pressing in our setting since the collusive equilibrium under (P^S) is strictly preferred by both agents. As Kreps (1990, 702) writes, “if a mechanism admits several Nash equilibria, some of which are worse for the designer (and, more importantly, better for the participants) than is the equilibrium that is desired, then one worries that the participants will find their way to the wrong equilibrium." This unintended consequence of standardized measures is formally presented in the following proposition.

Proposition 2.

For any interior r, the use of standardized measures introduces a tacit collusion problem. In particular, under the solution to (P^S): (i) each agent choosing e_L is also an equilibrium in the agents' subgame and (ii) from the agents' perspective, this collusion outcome Pareto dominates the equilibrium that has each agent choosing e_H.

Proof.

Again, without loss of generality, we focus on Agent 1's contract and strategy. Given $s^{*i}(x^1,x^2)$, if Agent 2 chooses e_L, Agent 1 (at least weakly) prefers to choose e_L if:

$u(s^*_{LL}) \geq pu(s^*_{LL}) + (1\text{-}p)u(s^*_{HL}) - v(e_H)$, or equivalently

$(1\text{-}p)[u(s^*_{LL}) - u(s^*_{HL})] + v(e_H) \geq 0.$

Plugging in $v(e_H)$ from the binding (IC^1) constraint, the condition becomes

$$(1 - p - p(1\text{-}r))[u(s^*_{LL}) - u(s^*_{HL}) + u(s^*_{HH}) - u(s^*_{LH})] \geq 0.$$

Recall, from the Lemma, $s^*_{HL} = s^*_{HH}$, and, hence, the condition further simplifies to

$$(1 - p - p(1\text{-}r))[u(s^*_{LL}) - u(s^*_{LH})] \geq 0.$$

Finally, since $s^*_{LL} > s^*_{LH}$, the above inequality holds. Hence, $e^i = e_L$ is an equilibrium in the agents' subgame under the solution to (P^S).

Not only does the $e^i = e_L$ equilibrium exist, it is also preferred by each agent to the $e^i = e_H$ equilibrium. Since (IC^i) binds, each agent's expected utility under the $e^i = e_H$ equilibrium is equal to $pu(s^*_{LL}) + (1\text{-}p)u(s^*_{LH})$. Under the $e^i = e_L$ equilibrium, each agent's expected utility is $u(s^*_{LL})$. Since $s^*_{LL} > s^*_{LH}$, the agents prefer the $e^i = e_L$ equilibrium. ■

Though the prospect of tacit collusion is disconcerting, it is not insurmountable. In particular, the principal can design a contract under which (e_H,e_H) is the unique equilibrium in the agents' subgame thereby quelling any fears of tacit collusion. It turns out that such a requirement does not ruin the benefits of standardized measures, but simply shifts the r-cutoff at which standardized measures become preferred.

Proposition 3.

There exists an interior cutoff r^{**}, $r^{**} > r^*$, such that for $r \in (r^{**},1]$, the principal prefers standardized measures, even with the added requirement that the desired equilibrium in the agents' subgame is unique.

Proof.

Under (P^C), the agents are treated independently and, hence, have dominant strategy incentives to choose e_H. Therefore, a unique equilibrium in the agent' subgame is assured by augmenting the solution to (P^C) with an arbitrarily small $\varepsilon > 0$ added to $s^1(x_H,y_k)$ and $s^2(x_j,y_H)$. Because of relative performance evaluation, the situation is more difficult with standardized measures. The only way to ensure a unique equilibrium in the agents' subgame is to add the requirement of dominant strategy incentives to choose e_H for at least one agent.

To confirm this claim, note that if neither agent has dominant strategy incentives at the solution, then each has a strict incentive to choose e_L if the other chooses e_L. Hence, (e_L,e_L) is a second equilibrium. Given this, the solution to the principal's problem entails providing dominant strategy incentives to one agent (of course, the best contract to provide to the other is as before, incurring an ε-cost to make the preference strict). Since the agents are ex ante identical, the principal is indifferent as to who gets dominant strategy incentives, so without loss of generality assume it is Agent 1. In this case, (P^S) is solved as before, with the added constraint that Agent 1 prefer e_H if $e^2 = e_L$. As with customized measures, the weakly dominant strategy contract that is obtained can be adjusted at an ε-cost to ensure a strict dominant strategy and a unique equilibrium in the agents' subgame. Denote the added constraint by ($IC^{1'}$), with the associated multiplier γ. Specifically, ($IC^{1'}$) is:

$$t(s^1(x^1,x^2)) \geq 0, \text{ where}$$

$$t(s^1(x^1,x^2)) = pu(s_{LL}) + (1\text{-}p)u(s_{HL}) - v(e_H) - u(s_{LL})$$

The first-order conditions for this more constrained program are:

$$\frac{1}{u'(s_{LL})} = \lambda - \mu\frac{1-r}{r} - \gamma\ \frac{1-p}{pr},$$

$$\frac{1}{u'(s_{LH})} = \lambda - \mu\ \frac{1-p-p(1-r)}{p(1-r)},$$

$$\frac{1}{u'(s_{HL})} = \lambda + \mu + \gamma\ \frac{1-p}{p(1-r)}\ \text{, and}$$

$$\frac{1}{u'(s_{HH})} = \lambda + \mu$$

The remainder of the proof proceeds in three steps.

Step 1. $\mu > 0$ and $\gamma > 0$ for all interior r.

If $\gamma = 0$, then the above first-order conditions are the same as for (P^S) in the Lemma.

In this case, from Proposition 2, we know that any solution that satisfies the first-order conditions violates ($IC^{1'}$).

Likewise, If $\mu = 0$ and $\gamma > 0$, the first-order conditions imply $s_{HH} = s_{LH}$. It is then easy to verify that any solution in which $s_{HH} = s_{LH}$, and which satisfies ($IC^{1'}$) as an equality, violates (IC^1).

Step 2. If Agent 1 is given dominant strategy incentives and Agent 2 is given Nash incentives using standardized measures, expected compensation is decreasing in r.

Denote the minimum expected compensation when the program is constrained by (IR^1), (IC^1), and ($IC^{1'}$) by $\widetilde{F}$. With a slight abuse of notation, again denote the solution by $s^{*1}(x^1,x^2)$. From the envelope theorem:

$$\frac{d\widetilde{F}}{dr} = \frac{\partial f(s^{*1}(x^1,x^2);r)}{\partial r} - \lambda\frac{\partial g(s^{*1}(x^1,x^2);r)}{\partial r} - \mu\frac{\partial h(s^{*1}(x^1,x^2);r)}{\partial r} - \gamma\frac{\partial t(s^{*1}(x^1,x^2))}{\partial r}.$$

Using the expressions for f, g, h, and t provided earlier yields:

$$\frac{d\widetilde{F}}{dr} = p[s^*_{LL} - s^*_{LH} - s^*_{HL} + s^*_{HH} - (\lambda+\mu)(u(s^*_{LL}) - u(s^*_{LH}) - u(s^*_{HL}) + u(s^*_{HH}))].$$

From Step 1, (IC^1) and ($IC^{1'}$) bind, which implies $u(s^*_{LL}) - u(s^*_{LH}) = u(s^*_{HL}) - u(s^*_{HH})$, so

$$\frac{d\widetilde{F}}{dr} = p[(s^*_{LL} - s^*_{LH}) - (s^*_{HL} - s^*_{HH})].$$

The first order conditions (and $\gamma > 0$) stipulate s^*_{HL} is the largest payment. That, coupled with $u(s^*_{LL}) - u(s^*_{LH}) = u(s^*_{HL}) - u(s^*_{HH})$ and the concavity of u yields $\frac{d\widetilde{F}}{dr} < 0$. And, from the Lemma, the same monotonicity holds for Agent 2.

Step 3. r^{**} exists and $r^{**} > r^*$.

The existence of an interior cutoff follows precisely the same arguments as in the proof of Proposition 1. The only difference is that the first-best contract for Agent 1 with r = 1 now sets the off-equilibrium payment s_{HL} sufficiently high to ensure e_H is a dominant strategy for Agent 1.

From Step 1, ($IC^{1'}$) binds for all interior r, so expected compensation when dominant strategy incentives are provided to Agent 1 with standardized measures is strictly higher than under (P^S). This implies the revised cutoff, r^{**}, is greater than r^*. ■

Note that resolving the collusion problem entails providing dominant strategy incentives to one agent and Nash incentives to the other. The optimal dominant strategy contract continues to take advantage of relative performance evaluation. With the dominant strategy requirement, however, not only is there a punishment for an agent having lower output than his peer ($s^1(x_L,x_H) < s^1(x_L,x_L)$), but there is also an added reward for having a higher output ($s^1(x_H,x_L) > s^1(x_H,x_H)$). Thus, a relative performance evaluation benefit of standardized measures persists. Figure 1 provides a visual depiction of the results in Propositions 1 and 3.

Insert Figure 1 here.

An ancillary observation associated with resolving collusion is the asymmetric treatment of otherwise identical agents. A related question is: to whom will the principal assign the dominant strategy contract? With identical agents, a coin flip works as the decision rule. But, some differences between agents can give rise to a clear preference.

In particular, say one agent has different beliefs about correlation in the measures than the other parties, and that the principal is unaware of that agent's beliefs. The principal may seek to assign dominant strategy incentives to such an agent. As discussed in Mookherjee and Reichelstein (1992), dominant strategy contracts in adverse selection models have the appealing feature of being prior independent. In our moral hazard model, the dominant strategy contract has a similar benefit in that it provides incentives regardless of the agent's beliefs about r.

To see this, note that from Step 1 in the proof of Proposition 3, (IR^1), (IC^1), and ($IC^{1'}$) bind when Agent 1 is provided dominant strategy incentives. Solving (IR^1), (IC^1), and ($IC^{1'}$) as equalities yields:

$$u[s^1(x_L,x_H)] = u[s^1(x_H,x_H)] - \frac{v(e_H)}{(1-p)},$$

$$u[s^1(x_L,x_L)] = \frac{v(e_H)-(1-p)u[s^1(x_H,x_H)]+\overline{U}}{p}, \text{ and}$$

$$u[s^1(x_H,x_L)] = \frac{v(e_H) - (1-p)^2 u[s^1(x_H, x_H) + (1-p)\overline{U}}{p(1-p)}.$$

Hence, the principal's problem under dominant strategy incentives amounts to picking $s^1(x_H,x_H)$ subject to the above relationships. Since the above relationships (constraints) stipulate payments that do not directly depend on r (r comes into the contract only through the principal's choice of $s^1(x_H,x_H)$), a dominant strategy contract provides appropriate incentives to Agent 1 no matter what he believes about r.

Of course, if the principal is unaware of either agent's beliefs, she may have to resort to offering dominant strategy incentives to both agents. Consistent with intuition, the provision of such incentives reduces but does not derail the attractiveness of standardized measures. That is, there exists an interior r-cutoff, greater than r^{**} of Proposition 3, such that for all higher r values, the principal prefers standardized measures to customized measures.

Another reason for preferring one agent over another in assigning the dominant strategy incentives arises if the agents have different traits which lead to different incentive problems. An example demonstrates this point. Say p = 0.3, r = 0.55, $\overline{U}$ = -7/4, and $u(s^i) = -e^{-0.01s^i}$. To reflect potential differences in the agents, say Agent 1 has an effort cost, $v(e_H)$, of 1, while Agent 2 has an effort cost of k. Introducing such agent asymmetry does not alter the derivation of the equilibrium, since the problem for each agent can be solved separately as in our main analysis with symmetric agents. However, it does have the potential to alter the desired assignment and measurement policy. The following table details expected compensation under each measurement system for two k-parameterizations of the example.

Insert Table 2 here.

In the example, when the agents are ex ante identical ($k = 1$), it of course does not matter which agent is assigned the dominant strategy contract. However, when the agents are different ($k = 0.8$), this decision proves critical. Since Agent 2 is easier to motivate in the first place, placing the added restrictions of a dominant strategy on him is less costly. Note the implications for performance measurement design. If dominant strategy incentives are provided to Agent 2, standardized measures outperform customized measures. An improper provision of dominant strategy incentives, however, results in customized measures being preferred. In short, selecting performance measures is a delicate exercise that requires a close eye on agent traits and the role of interconnectedness among measures.

Pushing the above theme further, consider the conventional view that suggests that standardized measures are most appropriate in situations where division problems are themselves standard. Contrary to this view, when $k = 1$ and incentive problems are identical across divisions, customization is called for in the example. And, standardization is preferable when the divisions' problems are different ($k = 0.8$). While this counterintuitive result does not generally hold, it does provide a word of caution to those seeking to extrapolate conformity (differentiation) among divisions simply by observing conformity (differentiation) among their performance measures.

CONCLUSION

This paper studies the choice among performance measures with a focus on relative performance evaluation. Customized measures provide a clearer picture of employee performance, but this benefit comes at a cost. Unlike customized measures, standardized measures share common features, among them errors in measurement. And, the commonality in errors has vital comparative evaluation benefits, since a measure for one employee can provide information about errors in the measure for another employee.

However, in exploiting relative performance evaluation, the firm has to be wary of inadvertently introducing opportunities for tacit collusion among employees.

With these forces at work, we provide conditions under which a set of customized measures and a set of standardized measures are each preferred. The tradeoff we consider may provide a more benign explanation for firms' reluctance to replace conventional accounting metrics with a smorgasbord of measures that are custom made to the goals and strategies of individual divisions.

As a parting comment, we note that these issues mirror the broader debate in accounting over standards. While firms may be able to more accurately reflect value by utilizing more specialized accounting choices, moving away from widely-accepted practices comes with the cost of diminishing comparability across firms. Presumably, well-designed standards reflect a balancing act between customization and the inherent comparability (relative evaluation) that comes with standardization.

FOOTNOTES

[1] Consider an example adapted from Lipe and Salterio (2000), where one division is in charge of selling to retail customers while the other takes catalog orders from commercial customers. Given sales growth is used as a measure for the commercial sales division, the firm must choose between using sales growth for the retail division or getting a more targeted measure of sales effort such as evaluations provided by mystery shoppers.

[2] Stressing the bond between informativeness (Holmstrom 1979) and relative performance evaluation, Laffont and Martimort (2002, 170) write: "The most spectacular applications of the Sufficient Statistic Theorem arise in multiagent environments. In such environments, it has been shown that the performance of an agent can be used to incentivize another agent if their performances are correlated, even if their efforts are technologically unrelated."

[3] This approach arises when the desired equilibrium in the subgame is required to be unique. An alternative approach is to expand the contract to include a more complex indirect mechanism in which whistle-blowing is encouraged (e.g., Ma et al. 1988).

[4] The specification of the two measurement systems is not mere labeling. Customized measures naturally have distinct errors, while standardized measures have commonality in errors. In Table 1, this translates to r being present under standardized measures but missing under customized measures.

[5] This is not to say expected compensation is decreasing in r if both tasks are performed by a single agent. In this case, multiple incentive constraints come into play, thereby precluding Kim's condition. The reader can confirm expected compensation is actually increasing in r in the single-agent case. As a result, with one agent taking both acts, customized measures are always preferred.

Standardized Measures: $Pr(x^1,x^2 \mid e^1,e^2)$

	(e_L,e_L)	(e_L,e_H)	(e_H,e_L)	(e_H,e_H)
(x_L,x_L)	1	p	p	pr
(x_L,x_H)	0	1-p	0	p(1-r)
(x_H,x_L)	0	0	1-p	p(1-r)
(x_H,x_H)	0	0	0	1-p-p(1-r)

Customized Measures: $Pr(x^1,y^2 \mid e^1,e^2)$

	(e_L,e_L)	(e_L,e_H)	(e_H,e_L)	(e_H,e_H)
(x_L,y_L)	1	0	p	0
(x_L,y_H)	0	1	0	p
(x_H,y_L)	0	0	1-p	0
(x_H,y_H)	0	0	0	1-p

TABLE 1. Probability Distributions Under Each Measurement System.

	k = 1	k = 0.8
Customized Measures	91.4	67.8
Standardized Measures (Nash Implementation)	92.6	59.2
Standardized Measures (Dominant Strategy for Agent 1)	107.5	74.2
Standardized Measures (Dominant Strategy for Agent 2)	107.5	63.6

TABLE 2. Expected Compensation in the Example.

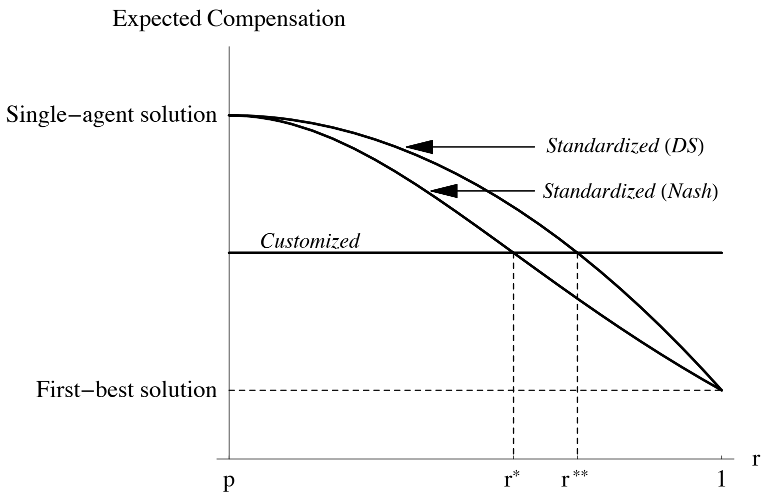

FIGURE 1. Expected Compensation as a Function of r.

REFERENCES

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

Demski, J. 1994. *Managerial Uses of Accounting Information*. Boston, MA: Kluwer.

Demski, J., and D. Sappington. 1984. Optimal incentive contracts with multiple agents. *Journal of Economic Theory* 33: 152-171.

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

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

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

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

Ittner, C., and D. Larcker. 1998. Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research* 10: 205-238.

Kim, S. 1995. Efficiency of an information system in an agency model. *Econometrica* 63: 89-102.

Kreps, D. 1990. *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.

Kronholz, J. 2002. Pros debate whether standardized tests should be timed–comparing performances is easier with clocking, but some advocate letting kids demonstrate mastery. *The Wall Street Journal* November 20: B1.

Laffont, J., and D. Martimort. 2002. *The Theory of Incentives*. Princeton, NJ: Princeton University Press.

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

Ma, C., J. Moore, and S. Turnbull. 1988. Stopping agents from 'cheating', *Journal of Economic Theory* 46: 355-372.

Mookherjee, D., and S. Reichelstein. 1992. Dominant strategy implementation of bayesian incentive compatible allocation rules. *Journal of Economic Theory* 56: 378-399.

Sliwka, D. 2002. On the use of nonfinancial performance measures in management compensation. *Journal of Economics and Management Strategy* 11: 487-511.