Performance Evaluation Metrics for Information Systems Development: A Principal-Agent Model

Rajiv D. Banker

*Carlson School of Management
University of Minnesota
Minneapolis, Minnesota 55455*

Chris F. Kemerer

*Sloan School of Management
Massachusetts Institute of Technology
50 Memorial Drive, E53-315
Cambridge, Massachusetts 02142-1347*

The information systems (IS) development activity in large organizations is a source of increasing cost and concern to management. IS development projects are often over-budget, late, costly to maintain, and not done to the satisfaction of the requesting user. These problems exist, in part, due to the organization of the IS development process, where information systems development is typically assigned by the user (principal) to a systems developer (agent). These two parties do not have perfectly congruent goals, and therefore a contract is developed to specify their relationship. An inability to directly monitor the agent requires the use of performance measures, or metrics, to represent the agent's actions to the principal. The use of multiple measures is necessary given the multi-dimensional nature of successful systems development. In practice such contracts are difficult to develop satisfactorily, due in part to an inability to specify appropriate metrics. This paper develops a principal-agent model that provides a set of decision criteria for the principal to use to develop an incentive compatible contract for the agent. These criteria include the *precision* and the *sensitivity* of the performance metric. After presenting the formal model, some current software development metrics are discussed to illustrate how the model can be used to provide a theoretical foundation and a formal vocabulary for performance metric analysis. The model is also used in a positive (descriptive) manner to explain why current practice emphasizes metrics that possess relatively high levels of sensitivity and precision. Finally, some suggestions are made for the improvement of current metrics based upon these criteria.

Software engineering—Management of computing and information systems

1. Introduction

Information systems (IS) development in large organizations is a source of increasing cost and concern to management.[1] IS development projects are often over budget, late, costly to maintain, and not done to the satisfaction of the requesting user.[2] It has been suggested that these problems exist, in part, due to the organization of the IS development process, where information systems development is typically assigned by the user (principal) to a developer (agent) (Gurbaxani and Kemerer 1989, 1990; Beath and Straub 1989; Klepper 1990; Whang 1992; Richmond et al. 1992). These two parties do not have perfectly congruent goals, and therefore a contract is developed to specify their relationship. An inability to monitor the agent directly requires the use of performance measures, or metrics, to represent the agent's actions to the principal. The use of multiple measures is necessary given the multidimensional nature of successful systems development. In practice such contracts are difficult to develop satisfactorily, due in part to an inability to specify appropriate metrics.

There is much current interest in industry in general related to performance contracting, and specific issues related to software development contracting are growing in currency with the increased awareness and interest in outsourcing of the systems development and delivery functions.[3] In order for organizations to enter into such arrangements with vendors formal contracts are required, and such contracts require valid performance evaluation metrics in order for both parties to reach agreement.

The difficulties that principals have in specifying performance metrics can be illustrated easily with a few examples from current practice. It is well documented that over an information system's useful life the maintenance costs typically exceed the development cost (Swanson and Beath 1990). Yet, in practice, software developers are typically evaluated by criteria such as on-time and on-budget delivery of the initial system, and rarely, if ever, on the likely maintainability of the system that they have just delivered (Gode et al. 1990). Izzo notes that, "Maintenance, long considered one of the most important product support services a business provides, is considered a secondary responsibility in information systems" (1987, p. 25). Therefore, the question remains, since developers understand this relationship, why don't their contractual arrangements reflect it?

Another example comes from a recent study of 11 large federal government systems integration projects.[4] The most frequent definition of success was "user satisfaction," yet the report notes that "Agencies such as the US GAO . . . ignore long-term user satisfaction and focus instead on cost and budget issues because they are easy to measure." Even interpreting this statement in a relative manner, i.e., ". . . are easier to measure," it is not obvious why this should be the case. Tracking cost and schedule data typically requires the implementation and use of a project management system

---

[1] The term "development" is used here to mean all the activities that constitute the systems life cycle, including systems maintenance. Activities solely related to new systems exclusive of any maintenance activity will be referred to as "new development."

[2] See, for example, Kemerer and Sosa (1991), Mehler (1991), Rothfeder (1988) and Ware (1990).

[3] See, for example, Bennett (1991) and Kirkpatrick (1991).

[4] The projects ranged in size from $42 M to $443 M (Anthes 1991).

devoted to the task. Developers need to record their time spent, and such actual data must be matched against previously budgeted milestones in order to generate the appropriate management information. Therefore, "easier to measure" must refer to conceptual rather than practical concerns. What is it that makes "user satisfaction" a desirable but underused performance metric?

In order to understand these apparent paradoxes of user and developer behavior this paper develops a principal-agent model that is analyzed to identify a set of decision criteria for the principal to use to specify the contract. This model results in two criteria, the *precision* and the *sensitivity* of the performance metric which influence the emphasis on various metrics. In particular, the model suggests that metrics that are relatively more precise and more sensitive will be preferred in the long term by both the principal and the agent in establishing the contract. These general results are then applied to two mini-case studies, one an internal IS group and one an external provider, to illustrate the application of these concepts in an IS development context.

The model provides a theoretical foundation and a formal vocabulary for performance metric evaluation in the general context of a multidimensional performance contract. The results of the model are applied to two organizations to illustrate the model's use in a positive (descriptive) manner to suggest explanations for the current relative emphasis in practice on cost and schedule. Additional discussion of the results shows how these results could be used in a normative manner to improve current metrics and develop new metrics that are more likely to be adopted.

This paper is organized as follows. The formal model is developed and shown in §2. Section 3 first develops a simple framework of IS development project performance metrics, and then applies the model results to two mini-case studies. Section 4 presents a broader discussion of both the ramifications and limitations of the model outside the context of the two organizations studied. Finally, some concluding remarks are presented in §5.

2. General Model

Information systems (IS) development is modeled as a principal-agent problem, with the client (the principal) desiring information systems to be developed to meet her goals.[5] She contracts with an IS project manager (the agent) to perform this work, due to specialized expertise on the part of the agent. The normal principal-agent model assumptions are made: (i) the goals of the agent are only imperfectly aligned with those of the principal (goal incongruence) and (ii) the agent's actions can only be imperfectly observed by the principal (information asymmetries). The principal is assumed to be risk-neutral and the agent is assumed to be risk and effort averse. Considerable prior work exists in this area, including Ross (1973), Jensen and Meckling (1976), Holmstrom (1979) and Harris and Raviv (1979). The current work builds directly on prior work by Banker and Datar (1989).

The principal is assumed to be interested in the outcome along n dimensions, which are represented by the vector $\mathbf{x} \equiv (x_1, \ldots, x_i, \ldots, x_n)$. The agent can increase

---

[5] Following Beath and Straub (1989) the use of "she/her" will refer to the principal, and "he/him" will refer to the agent in order to make pronoun references easier to follow. The model will focus on only these two parties, and excludes from consideration any possible agency relationship between the principal requesting the work and her superior, for instance, as suggested by Gurbaxani and Kemerer (1990). Therefore, it is applicable to situations involving either external or internal developers.

the likelihood of obtaining a better outcome x_i by devoting more effort a_i towards that outcome. More formally, let

$$\partial m_i/\partial a_i > 0, \qquad \partial m_i/\partial a_j = 0, \qquad i, j = 1, 2, \ldots, n, j \neq i$$

where $m_i = E(x_i \mid a_i)$ is the expected value of outcome x_i.

The outcomes cannot be observed jointly by the principal and the agent with perfect accuracy. The agent's efforts $\mathbf{a} \equiv (a_1, \ldots, a_i, \ldots, a_n)$ cannot be perfectly observed by the principal without incurring prohibitive monitoring costs. For performance evaluation purposes, therefore, appropriate metrics $\mathbf{y} \equiv (y_1, \ldots, y_i, \ldots, y_n)$ are developed to provide (imperfect) signals about the true outcomes.

More formally, let

$$y_i = x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n$$

where ϵ_i represents random variations (noise) for each of the n outcomes of interest.

In order to provide incentives for the agent to exert greater effort to produce higher levels of the outcomes of interest to the principal, the principal bases the agent's compensation on the jointly observable metrics $s = s(\mathbf{y})$ where s represents the agent's compensation. The monetary value of the outcomes to the principal is represented by w, where w is a function of $\mathbf{x}$, and therefore the risk-neutral principal seeks to maximize the expected value of $w(\mathbf{x}) - s(\mathbf{y})$. The agent, due to his risk and effort aversion, must be compensated at the end of the contractual time period (Lambert 1983). The principal understands the agent to be economically rational, and knows that a compensation contract based on y will influence the agent's actions $\mathbf{a}$. The agent seeks to maximize the expected value of $u(s) - v(\mathbf{a})$ where $u(\cdot)$ represents his utility for compensation, $s(\cdot)$, and $v(\cdot)$ represents his disutility for effort, with $u'(\cdot) > 0$, $u''(\cdot) < 0$, and $v'(\cdot) > 0$. The principal's problem can now be formulated as follows:

$$\max_{s(\cdot), \mathbf{a}} E[w(\mathbf{x}) - s(\mathbf{y})] \tag{1}$$

subject to

$$E[u(s(\mathbf{y})) - v(\mathbf{a})] \geq u_0, \tag{2a}$$

$$\partial E[u(s(\mathbf{y})) - v(\mathbf{a})]/\partial a_i = 0 \qquad \text{for} \qquad i = 1, \ldots, n, \tag{2b}$$

$$s \in [s_L, s_H], \qquad \mathbf{a} \in [\mathbf{a}_L, \mathbf{a}_H]. \tag{2c}$$

The objective function simply maximizes the expected benefit $w(\mathbf{x})$ to the principal of the information systems outcomes $\mathbf{x}$ net of compensation s paid to the agent. The first constraint ("individual rationality") ensures that the contract guarantees the agent a minimum expected utility level, u_0, equaling at least his best alternative employment possibility. The next set of n constraints ("self-selection") ensures that the agent's effort level choices a_i, $i = 1, 2, \ldots, n$, maximize his own expected utility level, and thus provide incentive compatibility with the second best actions. This set of first order optimization conditions is assumed to characterize the optimal action choices for the agent (Rogerson 1985). The final constraints specify a bounded feasible space to ensure the existence of an optimal solution to the principal's constrained maximization problem (Holmstrom 1979).

This program, solved repeatedly for different values of u_0, will generate the Pareto efficient frontier of possible contracts whereby neither the principal nor the agent can be made better off without the other being made worse off. The principal seeks to design the compensation contract that will maximize her own utility. The model can

be solved by setting the agent's expected utility at the level u_0. This amount is assumed to be determined by the market for the agent's skills. Therefore, in terms of the model, improved metrics (metrics which more closely approximate the actual outcomes) in the short term only benefit the principal, since solving the model involves selecting a fixed expected utility for the agent. However, in the long term improved metrics will lead to more effective monitoring, which will lead to actions by the agent that will improve his marginal product, which will move the entire Pareto efficient frontier outward, which will result in both parties being better off, under the assumption that the market will prevent the principal from capturing all of the marginal rents resulting from such a shift. Therefore, better metrics ultimately will be preferred by both the principal and the agent.

The Euler-Lagrange optimization conditions for the mathematical program above are given by the following:

$$\frac{1}{u'(s)} - \lambda + \sum_{i=1}^{n} \mu_i \frac{\int [\partial f(\mathbf{x}, \mathbf{y}; \mathbf{a})/\partial a_i] d\mathbf{x}}{\int f(\mathbf{x}, \mathbf{y}; \mathbf{a}) d\mathbf{x}}, \tag{3}$$

$$\frac{\partial^2}{\partial a_i^2} E[w(\mathbf{x}) - s(\mathbf{y})] + \sum_{j=1}^{n} \mu_i \frac{\partial^2}{\partial a_i^2} E[u(s) - v(\mathbf{a})] = 0$$

$$\text{for each} \quad i = 1, \ldots, n. \tag{4}$$

Here, λ and μ_i, $i = 1, \ldots, n$, are Lagrange multipliers for the $(n + 1)$ constraints. The joint probability density function of the outcomes $\mathbf{x}$ and the metrics $\mathbf{y}$ is embodied in $f(\cdot)$, and $\partial f(\cdot)/\partial a_i$ denotes its partial derivative with respect to effort dimension a_i. The condition in (3) reflects pointwise optimization for each observable value of the metric vector $\mathbf{y}$. Since the actual outcomes $\mathbf{x}$ are not jointly observable, the incentive contract cannot be based on it, and therefore integration is performed over all possible values of $\mathbf{x}$ in condition (3). Let

$$f(\mathbf{x}, \mathbf{y}; \mathbf{a}) = g(\mathbf{x}|\mathbf{y}; \mathbf{a})h(\mathbf{y}; \mathbf{a}) \tag{5}$$

where $g(\cdot)$ is the probability density function of $\mathbf{x}$ conditional on the observed value of $\mathbf{y}$, and $h(\cdot)$ is the marginal probability density function of $\mathbf{y}$. Now

$$\int [\partial f(\cdot)/\partial a_i] d\mathbf{x} = \int [\partial g(\cdot)/\partial a_i]h(\cdot) d\mathbf{x} + \int g(\cdot)[\partial h(\cdot)/\partial a_i] d\mathbf{x}. \tag{6}$$

But, $\int g(\cdot) d\mathbf{x} = 1$ because $g(\cdot)$ is a probability density function, and therefore

$$\int [\partial g(\cdot)/\partial a_i]h(\cdot) d\mathbf{x} = h(\cdot) \frac{\partial}{\partial a_i} \int g(\cdot) d\mathbf{x} = 0. \tag{7}$$

It follows from (5), (6) and (7) that

$$\frac{\int [\partial f(\mathbf{x}, \mathbf{y}; \mathbf{a})/\partial a_i] d\mathbf{x}}{\int f(\mathbf{x}, \mathbf{y}; \mathbf{a}) d\mathbf{x}} = \frac{[\partial h(\mathbf{y}; \mathbf{a})/\partial a_i]}{h(\mathbf{y}; \mathbf{a})}. \tag{8}$$

Returning to the condition in (3)

$$\frac{1}{u'(s)} = \lambda + \sum_{i=1}^{n} \mu_i \frac{[\partial h(\mathbf{y}; \mathbf{a})/\partial a_i]}{h(\mathbf{y}; \mathbf{a})}. \tag{9}$$

Differentiating (9) with respect to a particular y_j, $j = 1, \ldots, n$, yields

$$\left[-\frac{u''(s)}{(u'(s))^2} \right]\left[\frac{\partial s^*(\mathbf{y})}{\partial y_j} \right] = \sum_{i=1}^{n} \mu_i \frac{\partial}{\partial y_j} \frac{[\partial h(\cdot)/\partial a_i]}{h(\cdot)}. \tag{10}$$

In order to derive the distribution of the performance metrics $\mathbf{y}$, some additional structure is imposed. In particular, it is assumed that the stochastic variables x_i, given

the agent's choice of efforts a_i, are statistically independent[6] and are normally distributed with means m_i and variances η_i^2. The measurement error ϵ_i in the metric y_i is also assumed to be distributed normally with mean zero and variance σ_i^2. The errors ϵ_i are assumed to be distributed independent of x_i, x_j and ϵ_j, $j \neq i$. It follows, therefore, that the metrics $y_i = x_i + \epsilon_i$ are distributed independent of the other stochastic variables described above.

The conditional distribution of each y_i given a_i, being a convolution of two random variables following a bivariate normal distribution, is itself normal with mean

$$E(y_i|a_i) = E(x_i|a_i) + E(\epsilon_i) = m_i + 0 = m_i$$

and variance

$$V(y_i|a_i) = V(x_i|a_i) + V(\epsilon_i) = (\eta_i^2 + \sigma_i^2).$$

(In the analysis presented here, it is assumed that only the mean $m_i(a_i)$ is affected by the agent's actions. However, this approach could be extended to address the case where the variance of x_i can be influenced by the agent's actions.) The probability density function $h_i(y_i|a_i)$ is then given by

$$h_i(y_i|a_i) = \exp\{\tfrac{-1}{2} \ln 2\pi V(y_i|a_i) - [y_i - m_i(a_i)]^2/2V(y_i|a_i)\}.$$

Further, since the y_i are independently distributed,

$$h(\mathbf{y}|\mathbf{a}) = \prod_{i=1}^{n} h_i(y_i|a_i) \qquad \text{and}$$

$$\frac{\partial h(\mathbf{y}|\mathbf{a})/\partial a_i}{h(\mathbf{y}|\mathbf{a})} = \frac{\partial \ln h(\mathbf{y}|\mathbf{a})}{\partial a_i} = \frac{\partial \ln h_i(y_i|a_i)}{\partial a_i}$$

$$= [y_i - m_i(a_i)][\partial m_i(a_i)/\partial a_i]/V(y_i|a_i)$$

$$= [y_i - m_i(a_i)][\partial m_i(a_i)/\partial a_i]/[\eta_i^2 + \sigma_i^2].$$

Therefore,

$$\frac{\partial}{\partial y_j}\frac{[\partial h(\mathbf{y}|\mathbf{a})/\partial a_i]}{h(\mathbf{y}|\mathbf{a})} = 0 \qquad \text{for} \qquad j \neq i \qquad \text{and}$$

$$= [\partial m_i(a_i)/\partial a_i]/[\eta_i^2 + \sigma_i^2] \qquad \text{for} \qquad j = i.$$

It follows from equation (10) that

$$\frac{\partial s^*(\mathbf{y})}{\partial y_i} = \frac{-(u'(\cdot))^2}{u''(\cdot)} \cdot \frac{\mu_i[\partial m_i(a_i)/\partial a_i]}{[\eta_i^2 + \sigma_i^2]}. \qquad (11)$$

Recall that the goal is to characterize the optimal compensation contract, determined as a function of the available metrics. The principal's problem can be decomposed into two steps, one being the aggregation of the multiple performance metrics (the primary interest of the current analysis), and the other being the transformation

---

[6] The agent will trade off allocations of efforts a_i to different activities i, and to that extent the model captures the interdependent nature of the outcomes. The statistical independence assumption is maintained for expositional convenience; the principal results extend to the case of correlated stochastic variables.

of this aggregated signal into the ultimate compensation paid to the agent, the unidimensional $s^*(\mathbf{y})$. Since the right-hand side of equation (11) is independent of $\mathbf{y}$, it follows that the optimal compensation contract $s^*(\mathbf{y})$ can be written as $s^*(\mathbf{y}) = s^*(s_2^*(\mathbf{y}))$ where $s_2^*(\mathbf{y})$ is linear in $\mathbf{y}$ and can be interpreted as the aggregated performance evaluation metric, and s_1 is the mapping of the aggregate into compensation. It follows from equation (11) that

$$s_2^*(\mathbf{y}) = \sum_{i=1}^{n} \rho_i \xi_i y_i \qquad (12)$$

where $\rho_i = [\eta_i^2 + \sigma_i^2]^{-1}$ is the *precision* of the metric y_i which is inversely related to $V(x_i | a_i)$ and $V(\epsilon_i)$, and $\xi_i = \mu_i \partial m_i(a_i)/\partial a_i$ is the *sensitivity* of the outcome x_i (and the metric y_i) to the agent's action a_i.

Precision is a measure of the degree to which the value of the metric can be predicted, given a set of actions. The lack of precision, or increase in the variance, can be seen as being due to two sources. The first is that the relationship between an outcome x_i and corresponding action a_i may contain a great deal of *uncertainty* due to the effect of factors outside the purview of the agent. A second source may be a *lack of accuracy*, or "noise" in measuring x_i, i.e., large variations in the values of ϵ_i. More formally, the inverse of the precision measure can be decomposed into its two constituent components, as follows:

$$\mathrm{var}(y_i | \mathbf{a}) = \mathrm{var}(x_i | \mathbf{a}) + \mathrm{var}(\epsilon_i)$$

where the first term on the RHS corresponds to the uncertainty component (the amount of variance in the outcome given a set of agent's actions) and the second term corresponds to the inaccuracy component (the variance of the noise in measuring the outcome).[7] *Less formally, precision is a measure of the degree to which random factors may augment or countervail the agent's efforts to bring about the outcomes valued by the principal.* All else being equal, an agent will be better monitored by metrics with higher precision since the same incentives can be provided to the agent while imposing a reduced level of risk. Therefore, a metric with higher precision will be preferred by the principal since it will be more informative about the agent's action choice. This is true whether the greater precision results from greater certainty, greater accuracy, or some combination.

In equation (12), $\xi_i = \mu_i \partial m_i(a_i)/\partial a_i$ is the *sensitivity* of the outcome x_i (and the metric y_i) to the agent's action a_i. Using standard sensitivity analysis in optimization theory (Ioffe and Tihomirov 1979, pp. 292–298) the ξ_i is seen to correspond to the change in the principal's expected utility relative to the change in the agent's expected utility when, at the optimal solution, the agent's incentive compatibility constraint for the choice of a_i is perturbed marginally. In other words, ξ_i is the marginal value to the principal of providing the incentive to the agent to increase his effort a_i by a marginal unit. *Less formally, the degree of sensitivity of a metric (or outcome) can be seen as a measure of the impact that a unit of the agent's effort has on outcomes of importance to the principal.* The principal will want to encourage the agent's actions that most increase the final payoff to her, and therefore metrics that correspond to

---

[7] Precision is the reciprocal of the sum of the variances, which is generally not equal to the sum of the reciprocals of the variances.

these "high payoff" activities that are most sensitive to the agent's actions will be preferred by her relative to those with less impact.[8] For a metric to exhibit high sensitivity it must exhibit significant changes during the evaluation period in response to the agent's actions. A very sensitive metric would show a large change in the value to the principal, on average, for even a small additional amount of disutility to the agent resulting from an increase in effort. In terms of the optimal contract, more weight will be placed on metrics with high sensitivity relative to those with low sensitivity. Specifically, in the optimal performance evaluation measure the relative weight on each metric is directly proportional to its sensitivity times its precision.[9]

It is relatively easy to see that precision and sensitivity are independent concepts. Recall that $y_i = x_i + \epsilon_i$, $i = 1, 2, \ldots, n$ where the ϵ_i represent random variations (noise) for each of the n outcomes of interest. Suppose the vector of outcomes $\mathbf{x}$ is $\mathbf{x} = k\mathbf{a} + E$, where k reflects the agent's ability to influence the outcome, and E represents the effect of external factors, assumed to be a normally-distributed stochastic variable. Combining these two equations results in $\mathbf{y} = k\mathbf{a} + E + \epsilon$. The $k\mathbf{a}$ term represents the degree of sensitivity of the metric, and the second and third terms represent the two components of precision, certainty and accuracy.

Sensitivity and precision need not move together, i.e., metrics may score relatively high on one dimension and relatively low on the other. A simple two-metric example may help to illustrate this point. Imagine a compensation contract between the owner of a high technology corporation (the principal) and the firm's CEO (the agent). The owner may be ultimately interested in the total cash flow stemming from her investment in the firm's stock but, given the difficulty in observing this during the time period of a typical performance contract, may elect to use the level of short-term and/or long-term profits as the metrics (y_i), for purposes of the contract. The question then is how much weight to place on either metric. Short-term profits are a relatively more precise metric since the variance surrounding the effects of the agent's actions are relatively smaller than in the case of longer term profits, when many external factors (e.g., general economic conditions; actions of successor managers) may have unaccounted for effects. However, short-term profits may exhibit less sensitivity than long-term profits in that decisions that the agent may take today, e.g., technology selection, may have influence only in the longer run. That is, however hard the agent works he cannot do much to increase short-term profits. In other words, even in a world where no other forces countervailed (making the metric very precise), the sensitivity of the short-term profits metric to the agent's action may still be low. Therefore, the choice of weights for the two metrics would involve balancing these competing effects.

Of course, an actual contract may be based on several (>2) metrics. In particular, in the case of software development it will be argued below that this is the appropriate form for contracts to take. Since the true levels of the agent's efforts $\mathbf{a}$ are unobserved,

---

[8] With multidimensional tasks the agent has tradeoff possibilities and it is therefore possible that a particular μ_i could be very small. Therefore, it would be optimal at the margin to not devote additional effort to that task. In the event that the precision and sensitivity of the associated metric are low, effort devoted to that dimension is likely to become extremely small. This result is complementary to that of Holmstrom and Milgrom (1990)

[9] This extends the concepts of precision and sensitivity (Banker and Datar 1989) to the case of multiple actions ($\mathbf{a}$), multiple outcomes of interest ($\mathbf{x}$) and imperfect performance metrics ($\mathbf{y}$).

for incentive contracting purposes the principal and the agent agree on a set of performance evaluation metrics **y** that can be observed. This multidimensionality poses a dilemma for the principal: how to establish a contract that maximizes the agent's efforts appropriately across dimensions; in particular, which metrics to emphasize or weight in the agent's performance evaluation.

In order to effect the appropriate behaviors, the principal will base the agent's compensation in part upon the value of the performance metrics **y**.[10] Since the **y** are likely to be imperfect surrogates for the **x** and underlying effort choices **a**, some uncertainty is present. Therefore, an extreme form of compensation contract involving total reliance on performance evaluation metrics and assurance of certain utility for the principal is unlikely, since this places extreme risk on the agent, who is assumed to be risk averse. Conversely, however, the opposite extreme of zero reliance on the performance evaluation metrics is also unlikely, as this does not allow the principal to offer any incentives for appropriate behavior. These notions are, of course, predicated on the idea that the information costs related to gathering and reporting the **y** do not swamp the benefits to be gained from superior contracts.

It should further be noted that these results for use of the metrics for performance evaluation purposes are not dependent upon the customary assumption of risk neutrality of the principal.[11] In a case where both the principal and the agent are risk averse the central results for performance evaluation are unchanged. To evaluate the performance of the agent, the metrics **y** will be aggregated with weights reflecting sensitivity and precision as described above. In addition, the metrics will also be used for optimal risk sharing when both the principal and the agent are risk averse, where the exact weights for this purpose will depend on their relative risk tolerances.

However, within the range of likely contract forms, there is still room for considerable variation in terms of the choice of individual metrics (the y_is) and the weight that is to be assigned each metric in the compensation scheme. A metric that is more precise will receive more weight when all of the metrics are aggregated to determine the final performance evaluation than an otherwise identical metric. Similarly for a metric that is more sensitive. A potential issue is the mapping of the weighted performance evaluation metrics to the actual rewards. However, as shown above in equation (12), this third step is straightforward in this analysis, as the rewards will depend directly upon the weighted aggregate of the individual y_is. Therefore, the critical decision problem for the principal is the selection and use of appropriate metrics.

3. Application of the Model to IS Development

In this three-part section the model developed in §2 is applied to the domain of Information Systems (IS) development. Section A describes the broad overall dimensions of performance evaluation in IS development and gives illustrative examples of the typical metric used in each category. Section B presents specific metric operationalizations of these dimensions gleaned from two mini-case studies. Section C interprets the case study data in light of the model results.

---

[10] The emphasis here is on the use of a set of metrics to evaluate performance. The form of the actual reward, be it cash, stock options, promotion, time off, etc., will clearly vary due to individual preference, prevailing industry norms, etc., and will not be considered here.

[11] The standard assumption of risk aversion on the part of the agent is essential in that with a risk neutral agent no monitoring is required, and therefore no interesting managerial problem exists (Harris and Raviv 1979).

A. *Performance Evaluation in IS Development*

The principal seeks to motivate the agent to take actions that increase gross benefits and decrease costs.[12] It is assumed that higher effort on the part of the agent increases the expected value of the gross benefits to the principal. In an IS development context the costs and benefits have both long-term and short-term components. In the short term the emphasis is on initial *systems development costs,* most prominently labor costs. However, there are also longer term *maintenance* costs associated with each system. Numerous studies have shown that over half of all systems moneys are spent on maintenance (Lientz and Swanson 1981, Boehm 1987) and, most recently, that for every dollar spent on development, nine will be spent on maintenance (Corbi 1989). While many factors (including exogenous factors such as future changes in the business environment) may affect maintenance costs, for information systems development contracting purposes the principal can only attempt to ensure that the system developed by the agent can be maintained at the least possible foreseeable cost.

Benefits have traditionally been much more difficult to quantify, but can also be seen as having both a short- and long-term component. The principal requesting the system can begin to benefit only when the system is completed. Further, the business use of the new system may have to be coordinated with several other business activities, and considerable other resources may have to be committed at the anticipated implementation time for the system, particularly for larger systems. Therefore, if the system is delivered on time, the principal is likely to be better off, ceteris paribus, than if it were delivered late. This corresponds to the notion of *timeliness,* the ability to deliver the system on or before the deadline. However, in the long term, the ultimate value of the system may be due to the provision of user-desirable functionality which improves organizational performance. This is the notion of *effectiveness,* and it can only be interpreted in a longer term context.

Therefore, for model illustration purposes, the focus is on four outcomes for the principal to apply the efforts of the agent, represented as x_1 (initial development cost), x_2 (maintainability), x_3 (timeliness), and x_4 (effectiveness).[13] These are perhaps best presented by means of a 2×2 matrix.

Table 1 presents four outcomes (x) of interest to the principal requesting the information system.[14] (It will be useful to bear in mind that "initial development cost" will be an outcome to be reduced, in contrast to the other outcomes which are to be increased.) The principal and the agent must jointly agree on a set of perfor-

---

[12] In a recent review and analysis of potential IS effectiveness evaluation approaches, Cooper and Mukhopadhyay note that only three approaches, cost/benefit analysis, information economics, and microeconomic production functions, are suitable for use in performance evaluation, and that of these, only the first is of current practical applicability (1990, p 5 and Figure 1) Therefore, for illustrating the model in terms of current practice, the focus is on the cost/benefit approach to performance evaluation (See also Mukhopadhyay and Cooper 1991)

[13] Note that the research problem of interest here is the measurement of project results, which are the principal's typical concern, especially in the case of an external agent There may be extra-project organizational-level effects, (e.g., the degree to which a project furthered the professional development of its staff, which in turn may increase their value on some future, as yet unspecified project) but these are only secondary effects in terms of an individual project and therefore are not considered here.

[14] Note that while this framework is meant to be illustrative of performance dimensions used in practice, it maps well to other published frameworks For example, see Berger (1988) whose list consists of "cost, timeliness, accuracy, and quality" (p 78), which map relatively directly to Table 1

TABLE 1

Classification Matrix of IS Development Project Outcomes

	Short Term	Long Term
Cost	Initial Development Cost	Maintainability
Benefit	Timeliness	Effectiveness

mance evaluation metrics y for the compensation contract. If the x are observable by both the principal and the agent in the contractual period, then these may serve as the y. However, if that is not the case, then the principal and the agent must determine surrogate metrics that are jointly observable.

B. *Performance Evaluation Metric Operationalizations*

In order to determine the type and extent of project measurement used, two sites were selected as mini-case studies, one an internal development organization and the other an external firm. They are believed to be representative of typical current practice in information systems development.[15]

The internal organization is located within a large commercial bank. The information systems development group consists of approximately 450 professional staff members who work at developing and maintaining financial application software for the bank's internal use. The applications are largely on-line transaction processing systems, operating almost exclusively in an IBM mainframe COBOL environment. The bank's systems contain over 10,000 programs, totaling over 20 million lines of code. The programs are organized into application systems (e.g., Demand Deposits) of typically 100–300 programs each. Some of the bank's major application systems were written in the mid-1970s and are generally acknowledged to be more poorly designed and harder to maintain than more recently written software. The bank has made some attempts to upgrade its systems development capability. These steps include the introduction of a commercial structured analysis and design methodology, the institution of a formal software reuse library, and the use of some CASE tools on a few pilot projects.

The external organization is a major systems consulting and integration firm that operates nationally. Their staff consists of over 2,000 systems development professionals who are recruited from leading colleges and universities. They develop custom applications and sell customizable packages to a variety of public and private clients. Their various divisions are organized around a small number of specific industries, such as financial services. These divisions tend to focus on software and hardware platforms that are widespread in their respective market segments, although there is some firmwide commonality across divisions via a standardized development methodology and toolset. An emphasis is placed on very large systems integration projects that are often multiyear engagements. A state of the art development environment is maintained, with the firm being an early adopter of most software engineering innovations.

B.1 *Initial Development Cost—Empirical Observations.* At the bank, development cost is tracked through a project accounting system that is used to chargeback systems developer hours to the requesting user department. Hours are charged on a

---

[15] See §4.A for some external validation of this assumption

departmental average basis, with no allowance for the skill or experience level of the developer being incorporated into the accounting system. Mainframe computer usage is also charged back to the user, at a "price" designed to fully allocate the annual cost of operating the data center to the users. However, labor costs are generally believed to constitute 80% of the cost at this organization (Kemerer 1987).

At the consulting firm, development costs are tracked through a sophisticated project accounting and billing system, with the main entry being the biweekly timesheets of the professional staff, who may be simultaneously working on multiple projects for different clients. Time and materials contracts typically have multiple hourly rates whereby more project team members are billed at higher rates. Other direct project charges are also administered through this system, especially travel. Development is typically done at the client's site, and therefore hardware chargeback is typically unnecessary.

B.2 *Maintainability—Empirical Observations.* Long term maintenance costs are, in part, a function of the maintainability of a system (Banker et al. 1991a, 1992). While there are many factors outside the control of both the principal and the agent that can affect maintenance costs (e.g., changes in external business conditions such as regulatory changes), the principal desires that the agent deliver a system that can be maintained at the least possible cost. Therefore, the outcome that is desired is a high level of maintainability. Unfortunately, even the growing recognition of the significant magnitude of maintenance efforts has not yet produced a well-accepted metric for maintainability. The closest approximation to such a notion are the class of software metrics known as complexity metrics (McCabe 1976, Halstead 1977, Banker et al. 1991b). The general notion is that, as systems become more complex they become more difficult to maintain. The various complexity metrics provide a means of measuring this complexity, and therefore can be used both to predict maintenance costs, and as an input to the repair/rewrite decision (Gill and Kemerer 1991, Banker et al. 1991a, 1992).

At the bank, while maintenance projects are recognized as the primary information systems development activity, no attempt was made to measure and manage the maintainability of the applications, although most recently interest has been expressed in using the McCabe cyclomatic complexity metrics to aid management in this area. Similarly, at the consulting firm no maintainability measures are tracked, even though the ongoing maintenance of the developed system by the firm is a requirement of many projects.

B.3 *Timeliness—Empirical Observations* On the benefit row of Table 1, the short run benefit is provided by delivering the system on schedule, what is referred to as system timeliness. Of course, the appropriate duration of a systems development project is very much dependent upon such factors as the size of the system and the productivity of the development staff. Therefore, the timeliness metric is generally stated in relative terms, rather than absolute terms, most typically in relation to a deadline. Thus, a system is delivered "on time" or "two months late." Of course, this metric is really a difference result, and therefore an agent seeking to minimize the difference can direct effort both towards maximizing the time period (deadline) allowed during the project planning stage, as well as towards actually developing the system in such a way as to minimize the delay from the delivery date. However, a tendency on the part of developers to estimate or propose excessively long develop-

ment times will be mitigated by other controls, i.e., an external developer is unlikely to be awarded such a contract, and an in-house developer may find that the principal chooses not to do the system at all. Therefore, a timeliness metric can be assumed to provide at least partial motivation to develop the system promptly.

At the bank, project schedules are published and the larger projects are tracked via a regular status meeting chaired by the most senior vice president in charge of the information systems function. Project adherence to intermediate milestones is checked, and late projects are flagged for discussion. At the consulting firm, adherence to schedule is monitored through use of a development methodology with standardized milestones. Deliverable deadlines are an important part of many contracts, with clients' desire to implement systems by certain fixed dates a key contributor to their decision to use an external developer. Some contracts contain penalty clauses for late delivery.

B.4 *Effectiveness—Empirical Observations.* The fourth and final cell in Table 1 is a long-term benefit, or effectiveness. Effectiveness metrics are much sought, but little or no general agreement has been reached on such metrics. Crowston and Treacy note that:

> Implicit in most of what we do in MIS is the belief that information technology (IT), has an impact on the bottom line of the business. Surprisingly, we rarely know if this is true (1986, p. 299)

They go on to review the existing literature in this area for the previous ten years and conclude that until more progress is made in identifying performance variables, the best current metrics can only test whether systems engender user satisfaction. This finding was recently reaffirmed by a study of large federal government systems integration projects, where a survey of the program managers revealed that user satisfaction was the most frequently cited measure of success [ADAPSO 1991]. Therefore, commonly accepted effectiveness metrics tend to take the form of surveys of user satisfaction that could be administered at the end of the project.[16]

At the bank, no formal mechanisms are in place to measure user satisfaction, although occasional efforts are made to interview key users about their needs. At the consulting firm, while user satisfaction is deemed to be highly relevant in terms of its linkage to follow-on contracts, until very recently, no standardized mechanism existed to capture this information. Of course, contractual provisions typically guarantee some minimum level of performance. Beyond this, a small number of newer projects are experimenting with a user satisfaction survey.

C. Application of Model Results

The results of the previous sections are now combined by applying the measurement criteria from the model to the commonly used operationalizations of IS performance evaluation metrics. From this application some observations are made with regard to the model criteria about the relative emphasis on the current operationalizations in practice.

---

[16] Criticisms of this work point out that it is not theoretically based and that results of these surveys will be subject to users' prior expectations about the system (Chismar et al. 1986, Melone 1990). More recent work proposes "system/task fit," or "user satisfactoriness" as a theoretically-based alternative measure of system effectiveness (Goodhue 1986, 1988; Miller 1989).

TABLE 2

Metric Operationalizations of IS Development Project Outcomes

	Short Term	Long Term
Cost	Budget	Complexity metrics
Benefit	Schedule	User satisfaction

Table 2 summarizes the empirically observed operationalizations of the project outcome dimensions from Table 1.

C.1 *Precision.* The first criterion is the precision of the performance evaluation metrics. The two principal components of precision, lack of certainty, as defined by $Var(x_i | a)$, and lack of accuracy, as defined by $Var(\epsilon_i)$, are considered in turn. Certainty, as defined above, is directly related to the outcome, while accuracy is related to outcome through a specific metric.

There are relatively few factors external to the agent's actions that influence development cost and timeliness, as compared to the long-term outcomes, in that the agent can propose a budget and schedule and then staff the project in such a way as to attempt to meet those goals. Of course, the influence of external factors is not absent. In particular, the interface with other projects can be a source of disruption for the agent. A parallel project may not complete its portion in time for the agent's project to keep to its critical path and therefore its schedule. Changes in project scope are also an important influence, unless the agent carefully manages the changes by ensuring that the schedule and budgets are revised accordingly.

Interruptions to the project are likely to have greater effects on schedule than on the budget. This is because if work on the project is delayed it is often possible to temporarily reassign staff to work that does not have them charging time to the project, thus avoiding a budget overrun. On the other hand, "time marches on" as far as the deadline goes, with any delay in the critical path making the project late. Therefore, the certainty component of the precision of the budget metric will be higher than that for schedule.

Maintainability as operationalized by complexity metrics would rate a relatively middle score on a certainty scale. The agent's actions can clearly improve complexity metric scores, but he may be constrained by outside limitations, such as the need to reuse portions of existing systems that are relatively complex. Also, there are many dimensions to software complexity, and a metric like cyclomatic complexity measures only one aspect. In fact, it may be argued that overly strict reliance on one complexity metric can merely transfer the complexity to other, unmeasured dimensions, e.g., data complexity. Therefore, complexity metrics are relatively less certain than budget metrics.

Finally, least certain of all is the system's effectiveness. The system may have been poorly conceived initially by the requester, and therefore, the delivered system, while perhaps meeting the agreed upon technical specifications, may not prove to be valuable. Or, the principal may have done an inadequate job of making the organizational changes necessary for the success of the new system, e.g., reassignment of tasks, retraining, and adjustment of compensation systems. In support of these notions there is a growing body of descriptive work that suggests that many completed systems are never used (Rothfeder 1988, Kemerer and Sosa 1991).

In the accuracy component, the short-term measures clearly allow for more accuracy than the long-term measures. Project management systems routinely track project expenditures and deadlines, and these provide metrics that are relatively objective and accurate versus either maintainability (subject to limitations in measurement and the impact of the unknown nature of future change requests) or effectiveness (subject to the lack of reliability of the measurement instrument and the unknown impact of future changes in the business). For example, if user satisfaction metrics are used, it may be in the interests of the user to not report satisfaction as high, in order to extract additional effort or attention from the developer. These problems with maintainability and effectiveness reduce the precision of metrics for those performance evaluation variables.

Summing these two components of precision, certainty and accuracy, it can be seen that, at these two sites, development cost scores relatively the best on both components, while effectiveness scores relatively the worst. Timeliness and maintainability rate in the middle of these two extremes in terms of their precision.

C.2 *Sensitivity.* If sensitivity is high, then for a small amount of disutility the agent can significantly increase the utility of the principal. Development cost, operationalized at both sites primarily as labor work months, is a sensitive metric, that is, it possesses a relatively high value for $\mu_1 \partial m_1(a_1)/\partial a_1$. A project manager can change the expected development cost by deciding which staff members are to be assigned and how they are to be deployed, and by providing leadership and supervision during the development process. In addition, a manager may also influence project cost by under-reporting his own hours, as a means of adding value to a project without exceeding the budget. However, project managers at the consulting firm can typically exert more leverage than can their counterparts at the bank since at least some differential labor rate structures exist. The consulting firm agent can exploit different mixes of high and low cost staff in an attempt to keep within the budget. At the bank all staff are charged to projects at the bank's average labor cost, and therefore a bank project manager has somewhat less flexibility.

One concern with this analysis might be the notion that a project manager at the bank could essentially "game" meeting a particular budget by assigning a staff of say, for example, more productive than average people to a project with a tight budget since everyone is charged at the same $40/hour rate. However, this scenario does not ultimately change the sensitivity rating, due to the following logic. If the project manager is assumed to be at a low level where he has only one project in whose outcome he is interested, then he might attempt such an optimization. However, at the bank there are multiple projects and hence multiple project managers, all of whom would like to game the situation this way, and therefore, through competition for resources, this strategy is not likely to obtain. Alternatively, one might posit a "super project manager," responsible for all the current projects. In this case this individual is presumably interested in the outcomes of all the projects and cannot staff them all with "above average" personnel.

The other short-term measure, timeliness, as operationalized by the degree to which the deadline is met, is also a sensitive metric. However, timeliness is not very highly sensitive, since while assigning less or more expensive personnel can directly affect the project cost, the influence on timeliness is less direct. An example of this is Brooks's research which has been summarized into the aphorism that "adding staff to

a late project makes it later," denying the ability of the agent to move the timeliness metric in the desired direction in a substantial way (Brooks 1975, p. 25). The less sensitive nature of schedule performance depends in part upon the project specification being sufficiently concrete as to disallow the possibility of significant "gaming," i.e., undocumented reductions in scope that allow the appearance of on time delivery of what in reality is significantly reduced functionality. This is the situation at both of the case study sites, particularly the external consulting firm where formal contracts are the norm. However, where this is not the case it might be expected that timeliness would be the most sensitive metric.

In terms of the longer-term metrics, the cost side is reflected by maintainability, possibly operationalized by complexity metrics (although not done at either site), and the benefit side is referred to as effectiveness, possibly operationalized by user satisfaction (although not done regularly at either site). Maintenance, despite its growing economic importance, is a relatively unstudied and therefore poorly understood phenomenon. Since the relationships among agents' efforts and their impact on maintainability are not well understood, and since metrics for measuring maintainability are immature, it follows that the relationship among agent's efforts and complexity metrics are even less well understood. The project manager's ability to influence maintainability is limited, and thus the sensitivity of maintenance metrics can only be described as relatively low.

Conversely, the user satisfaction metrics used to indicate effectiveness should show relatively high sensitivity. Often, the inclusion of a seemingly small feature can greatly improve the user's perceived or even actual value for the application. If the IS development agent is aware of user needs and preferences, particularly regarding user interface issues, he is often able to greatly influence user satisfaction. Since the literature notes the strong influence played by expectations in user satisfaction, a talented agent may be able to greatly control expectations, and therefore the value of the metric at the end of the project.

In summary, for these two sites, the relative sensitivity of the commonly used metrics are as follows. If used, user satisfaction exhibits relatively high sensitivity. Timeliness and development cost may also be relatively sensitive, with the consulting firm agents often having a greater ability to influence this than bank project managers. Finally, maintainability, with the current poor understanding of the relationship between complexity and maintenance, is relatively the least sensitive of the four.

C.3 *Summary.* In examining all of the performance evaluation metrics relative to the criteria defined by the model, it is proposed that at these two sites development cost and timeliness rate well in terms of both sensitivity and precision. User satisfaction seems sensitive, but fares poorly in terms of its precision, while maintainability is only moderately sensitive and moderately precise.

Recall equation (12):

$$s_2^*(\mathbf{y}) = \sum_{i=1}^{n} \rho_i \xi_i y_i. \tag{12}$$

This result shows that a linear aggregation of the scores will produce the correct ranking of performance evaluation metrics. In other words, metrics with relatively higher levels of precision and sensitivity will receive more weight in the final aggregated evaluation. As shown in Table 3, an ordinal ranking of the metrics discussed in

TABLE 3

Relative Metric Values

| Metric | Precision | | | Ordinal Score |
	Certainty	Accuracy	Sensitivity	
Budget Performance	High	High	Medium to High (bank) (consultants)	1
Schedule Performance	Medium	High	High	1
User Satisfaction	Low	Medium	High	2
Maintenance complexity	Medium	Medium	Low	3

the mini-case studies would find budget and schedule performance at the top, followed by user satisfaction and then followed by maintainability.

To summarize, this ranking is based on observations at the two sites. In practice, both sites emphasize measurement on two dimensions, cost and timeliness, that are seen to possess relatively the most precision and sensitivity, as predicted by the model. How these dimensions might fare at other sites, or at these same sites in the future, are discussed in the following section.

4. Discussion

In this section the generalizability and implications of the results shown in §3 are discussed. Limitations of the model and possible extensions to it are also presented.

A. *Generalizability and Implications of the Results*

In examining the results presented above, one possible concern might be with the representativeness of the two mini-case studies. While their measurement practices are as predicted by the model, to what degree are they believed to be representative of current practice?

Three other sources of data on the current state of measurement suggest that the two mini-cases may be quite typical of current practice. The first source is a survey of over 140 medium to large IS departments conducted in 1988, in which managers were asked what measures they currently used (Howard 1988). By far the leading measures were work-hours per project, a measure of development cost (78% of managers surveyed), and adherence to delivery dates, a measure of timeliness (72%). The third most used measure was computer resource usage, which was only mentioned by 27% of the respondents. All other measures were less frequently reported, and, in particular, "module size," a potential measure of maintainability, was reported by only 8% of the respondents.

A second independent source is some descriptive data from the text by Jones (1991). His reports about the status of software measurement in various industries are worth quoting at length:

> Companies such as Exxon and Amoco were early students of software productivity measurement, and have been moving into . . . user satisfaction as well . . . The leading insurance companies such as Hartford Insurance, UNUM, USF&G, John Hancock, and Sun Life Insurance tend to measure productivity, and are now stepping up to . user satisfaction measures as well. . In the manufacturing, energy, and wholesale/retail segments the use

of software productivity measurement appears to be proportional to the size of the enter-
prise: the larger companies with more than a thousand software professionals such as Sears
Roebuck and J C Penney measure productivity, but the smaller ones do not. . . user
satisfaction measurement are just beginning to heat up within these industry segments. . .
Companies such as Consolidated Edison, Florida Power and Light, and Cincinnati Gas and
Electric are becoming fairly advanced in software productivity measure. Here too,   . user
satisfaction measures have tended to lag behind. (pp. 16–18)

Jones's use of "productivity" here is in a broad sense that receiving more output for
the work hours input to the project will result in better performance on both budget
and schedule relative to less productive projects. Note that maintainability metrics
are conspicuous by their absence from this list, and that user satisfaction metrics tend
to lag schedule and budget metrics.

A third source is the work of Humphrey on software process maturity (Humphrey
1988, p. 74). He notes that the first measures adopted by organizations are cost and
schedule metrics, and it is not until stage four of the five-stage model that more
comprehensive measures are expected to be implemented. It should be noted that the
vast majority of software development organizations in the United States are
currently at stages one or two.

These independent observations corroborate what was observed at the two mini-
case studies. Budget and schedule metrics are in wide use, while effectiveness mea-
sures in the form of user satisfaction metrics are less widely adopted. Measures of
maintainability are completely absent from these discussions, which is consistent
with the results in Table 3 which suggest that they are the least likely of the four to be
adopted.

The implications for this choice of adoption are worthy of managerial concern.
The emphasis on short-term results may produce decisions on project planning,
staffing, and technology adoption that are suboptimal for the organization in the long
term. For example, the almost total lack of measurement of the maintainability
impacts of project decisions implies that only minimal effort will be devoted towards,
for example, useful design and code documentation or adherence to structured cod-
ing precepts, to the extent that these activities are viewed as costly or otherwise
compete for resources with different activities that are measured. Similarly, an em-
phasis on schedule and budget measurements in preference to effectiveness measures
implies an emphasis on delivering any product on-time, rather than a better product
later, where this latter option might be the preferred alternative for the organization.

Another application of these results could be on the part of external IS develop-
ment firms. As agents typically bidding on competitive contracts, one method of
increasing the desirability of their services to the principal is by incurring so-called
bonding costs (Jensen and Meckling 1976). These bonding costs are actions by the
agent to provide assurances to the principal that possible goal incongruencies on the
part of the agent will be offset by such costs. One way for IS development agents to do
this would be to develop performance "guarantee" metrics that have relatively high
levels of precision and sensitivity upon which a contract can be based. For example,
the external consulting firm portrayed in this mini-case study could provide sug-
gested maintainability and effectiveness measures that it was willing to adhere to as
part of its proposal. Such a proposal would be viewed more favorably by the principal
than one without, all other things being equal.

Most importantly, while some of these conclusions may have been made by other

observers, the current research provides a theoretically grounded formal model which provides concepts that predict the choice of performance metrics in information systems organizations. These concepts conceivably can be used then to diagnose and improve current metrics and support the development of new metrics. With an informed understanding of why it is that budget and schedule metrics are preferred in practice, managers who wish to provide more balanced project outcomes by, for example, seeking to incorporate measures of maintainability into the development contract, should seek to discover and/or develop maintainability metrics that possess high levels of precision and sensitivity. For example, if code complexity metrics such as McCabe's cyclomatic complexity are shown to be good predictors of future maintenance costs, and if the agent can be given sufficient control over the code, perhaps through automated restructurers, such that he can influence these metrics in the appropriate direction, then inclusion of such measures in performance evaluation contracts can be expected to increase (Gill and Kemerer 1991).

B. *Limitations and Possible Extensions to the Results*

The discussion so far has been limited to application of the results of the model to examples in traditional information systems development that are currently observed. Two obvious extensions to this analysis would be to (a) apply the model to different development environments, and (b) speculate as to future trends that may have some impact on these results.

Different environments may have available metric operationalizations that exhibit higher precision or sensitivity or both versus their counterparts in traditional information systems. For example, the effectiveness dimension is traditionally perceived as difficult to quantify. However, in another environment this may not be the case. For example, in a safety critical application, such as real-time control of a nuclear power plant, software reliability may be the overwhelming criterion, and therefore the degree to which the software has been tested and can be "proven" correct may swamp all other possible effectiveness considerations. To the degree that metrics for reliability exhibit higher precision relative to the equivalent user satisfaction metric of traditional information systems, and to the degree that reliability is a highly valued outcome dimension, it will be weighted more heavily. Another example might be the effectiveness of a real-time military fire control system which may depend almost solely on its operational performance (speed). This may lend itself to easily definable metrics that possess desirable properties.

One change that may occur over time within the commercial information systems environment is greater recognition of the ability to measure and improve software maintainability (Swanson and Beath 1989, Chapter 8). While the importance of the maintenance activity has been recognized for over a decade (Lientz and Swanson 1981) it is only recently that research has linked measures of complexity to maintainability (Gibson and Senn 1989, Gill and Kemerer 1991, Banker et al. 1991a, 1992). This realization has been accompanied by the commercial availability of automated tools that deliver the metric values. To the degree to which these static analysis tools are delivered within CASE environments, rather than having to be justified and purchased as stand-alone tools, their use can be expected to increase. Therefore, over time a greater understanding and refinements of software complexity metrics as operationalizations of the maintainability dimension may improve the precision of this metric. The sensitivity of maintenance metrics may also improve as research in

this field provides clearer direction to managers to best apply their efforts in reducing maintenance requirements.

A further interpretation of the results from the model would be to move beyond the positive or descriptive aspects and use the results to argue for greater emphasis on development and improvement of metrics for both effectiveness and maintainability, as these are the two dimensions least well represented by current metrics. For example, the effectiveness dimension would be emphasized more if there were a more precise metric than the current user satisfaction metric. It should be noted that this result for effectiveness, derived from the agency theory perspective, matches well with some current calls from practitioners for better measures of the "business value" of IS development (Banker and Kauffman 1988) and with movements toward user-centered design within the human-computer interaction research community (Grudin 1991)..

5. Concluding Remarks

This paper has developed a principal-agent model that provides a common conceptual framework to illuminate current and future practice with regard to performance evaluation metrics for information system development. Given the principal-agent nature of most significant scale IS development, insights that will allow for greater alignment of the agent's goals with those of the principal through incentive contracts will serve to make IS development both more efficient and more effective. An important first step in this process is gaining a better understanding of the behav-

The current research provides a theoretically grounded formal model which defines criteria that predict the choice of performance metrics in information systems organizations. The insights available from the model both suggest explanations as to the current weighting of the dimensions of IS development performance, and provide insights into where better metrics are needed if the current largely unsatisfactory situation is to be remedied. These concepts can conceivably be used to diagnose and improve current metrics and support the development of new metrics.

In terms of future research, a natural follow-on would be to perform a formal empirical validation of the proposed relative weightings given a set of performance evaluation metrics. This will require the development of an instrument to measure the model's sensitivity and precision constructs. The ultimate value of such research will be in an increased understanding of how best to evaluate current systems development performance, so as to provide guidance to managers on how best to improve that performance. Given the key role played by systems development in enabling strategic uses of information technology, such improvement is of critical importance to the management of organizations.*

Bibliography

ADAPSO. "Observations on Successful Federal Systems Integration Programs. An ADAPSO Survey of Industry and Federal Agency Program Managers," ADAPSO Federal Information Systems Integration Committee Technical Report, June 1991.

Anthes. G. H., "Oversight Agencies Neglect Role of User Satisfaction," *Computerworld*, 81 (July 22 1991).

Banker, R. D. and S. M. Datar, "Sensitivity, Precision and Linear Aggregation of Signals for Performance Evaluation," *Journal of Accounting Research*, 27. 1 (Spring 1989), 21–39.

————— , ————— , C. F. Kemerer and D. Zweig, "Software Complexity and Software Maintenance Costs," MIT Sloan School Working Paper #3155-90. Jaunary 1992.

Banker, R. D , S. M. Dater and D. Zweig, "Complexity, Structuredness, and the Rate of Errors in Application Programs," University of Minnesota Working Paper, 1991a.

————— , ————— and ————— . "Software Complexity Metrics: An Empirical Study," University of Minnesota Working Paper, 1991b.

————— and R. Kauffman, "Strategic Contributions of Information Technology: An Empirical Study of ATM Networks," *Proceedings of the Ninth International Conference on Information Systems*. Minneapolis, Minnesota, 1988. 141–150.

Beath, C. M. and D. Straub. "Managing Information Resources at the Department Level: An Agency Perspective," *Proceedings of the 22nd Hawaii International Conference on Systems Sciences,* Kailua-Kona, Hawaii, January 1989, 151–159.

Bennett, A., "Paying Workers to Meet Goals Spreads, but Gauging Performance Proves Tough," *Wall Street Journal*, B1 (September 10, 1991).

Berger, P., "Selecting Enterprise-Level Measures of IT Value," in *Measuring Business Value of Information Technologies*, P. Berger, J. G. Koielus and D. E. Sutherland (eds.), ICIT Press, Washington DC, 1988. 59–92.

Boehm, B., "Improving Software Productivity," *Computer*, 20. 9 (September 1987), 43–57.

Brooks, F., *The Mythical Man-Month*, Addison-Wesley, 1975.

Chismar, W., C. H. Kriebel and N. P. Melone, "A Criticism of Information Systems Research Employing 'User Satisfaction.' " Graduate School of Industrial Administration. Carnegie Mellon University Working Paper 24-85-86, 1986.

Cooper, R. and T. Mukhopadhyay, "Toward the Formulation and Validation of a Microeconomic Production Approach for Management Information System Effectiveness." Carnegie Mellon University Working Paper, November 1990

Corbi, T. A., "Program Understanding: Challenge for the 1990s," *IBM Systems Journal*, 28, 2 (1989). 294–306.

Gibson, V. R. and J. A. Senn, "System Structure and Software Maintenance Performance " *Communications of the ACM*, 32. 3 (March 1989), 347–358.

Gill, G. K. and C. F. Kemerer, "Cyclomatic Complexity Density and Software Maintenance Productivity," *IEEE Transactions on Software Engineering*, 17, 12 (December 1991). 1284–1288.

Gode, D. K., A. Barua and T. Mukhopadhyay, "On the Economics of the Software Replacement Problem," *Proceedings of the 11th International Conference on Information Systems* Copenhagen, Denmark, pp. 159–170, December 1990

Goodhue, D , "IS Attitudes: Toward Theoretical and Definition Clarity," *Proceedings of the Seventh International Conference on Information Systems*. San Diego, California, December 1986, 181–194.

————— , "I/S Satisfactoriness: An Outcomes Measure for MIS Research," University of Minnesota MISRC Working Paper MISRC-WP-89-05, October 1988.

Grudin, J., "Interactive Systems: Bridging the Gaps between Developers and Users," *IEEE Computer*, 24, 4 (April 1991), 59–69.

Gurbaxani. V and C F Kemerer, "An Agent-theoretic Perspective on the Management of Information Systems," *Proceedings of the 22nd Hawaii International Conference on System Sciences,* Kailua-Kona, Hawaii, 1989, 141–150

————— and ————— , "An Agency Theory View of the Management of End User Computing," *Proceedings of the 11th International Conference on Information Systems*. Copenhagen, Denmark, 279–290. December 1990.

Halstead, M., *Elements of Software Science*, Elsevier North-Holland, New York, NY, 1977.

Harris, M and A. Raviv, "Optimal Incentive Contracts with Imperfect Information," *Journal of Economic Theory*, 20, (1979), 231–259.

Holmstrom, B., "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 1 (Spring 1979), 74–91.

———— and P. Milgrom, "Multi-task Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," Yale University Working Paper 45, May 1990

Howard, P. C., "Trends in Application Development Productivity," *System Development*, 8, 11 (November 1988), 1–5

Humphrey, W. S., "Characterizing the Software Process: A Maturity Framework," *IEEE Software*, 5, 3 (March 1988), 73–79.

Ioffe, A. D. and V. M. Tihomirov, *Theory of Extremal Problems*, North-Holland, New York, NY, 1979.

Izzo, J., *The Embattled Fortress Strategies for Restoring Information Systems Productivity*, Jossey-Bass, San Francisco, CA, 1987.

Jensen, M. and W. Meckling, "Theory of the Firm: Managerial Behaviour, Agency Costs, and Ownership Structure," *Journal of Financial Economics*, 3, 305–360, 1976.

Jones, C., *Applied Software Measurement*, McGraw-Hill, New York, NY, 1991.

Kemerer, C. F., *Measurement of Software Development Productivity*, Unpublished Carnegie Mellon University Ph.D. Thesis, 1987.

———— and G. L. Sosa, "Systems Development Risks in Strategic Information Systems," *Information and Software Technology*, 33, 3 (April 1991), 212–223

Kirkpatrick D., "Why Not Farm Out Your Computing?," *Fortune* (September 23, 1991), 103–112.

Klepper, R., "An Agency Theory Perspective on Information Centers," *Proceedings of the Twenty-Third Annual Hawaii International Conference on System Sciences*, Kailua-Kona, Hawaii, January 1990, 251–259

Lambert, R. A., "Long-term Contracts and Moral Hazard," *Bell Journal of Economics*, (1983), 441–452

Lientz, B. P. and E. B. Swanson, "Problems in Application Software Maintenance," *Communications of the ACM*, 24, 11 (November 1981), 31–37.

McCabe, T. J., "A Complexity Measure," *IEEE Transactions on Software Engineering*, SE-2, 4 (1976), 308–320

Mehler, M., "Reining in Runaway Systems," *Information Week*, 351 (December 16, 1991), 20–24.

Melone, N. P., "Theoretical Assessment of User-Satisfaction Construct in Information Systems Research," *Management Science*, 36, 1 (January 1990), 76–91.

Miller, J., "Information Systems Effectiveness: The Fit Between Business Needs and System Capabilities," *Proceedings of Tenth International Conference on Information Systems*, Boston, MA, December 1989, 273–288

Mukhopadhyay, T. and R. B. Cooper, "Impact of Management Information Systems on Decisions," *Omega* 20, 1 (1992), 37–49

Richmond, W. B., A. Seidmann and A. B. Whinston, "Incomplete Contracting Issues in Information Systems Development Outsourcing," *Decision Support Systems*, in press, 1992.

Rogerson, W., "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 1985.

Ross, S., "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, (1973), 134–139.

Rothfeder, J., "It's Late, Costly, and Incompetent—But Try Firing a Computer System," *Business Week*, (November 7, 1988), 164–165.

Swanson, E. B. and C. M. Beath, *Maintaining Information Systems in Organizations*, John Wiley & Sons, New York, 1989.

———— and ————, "Departmentalization in Software Development and Maintenance," *Communications of the ACM* 33, 6 (June 1990), 658–667.

Ware, R., "Gong-Ho Projects," *Journal of Systems Management*, 41, 12 (December 1990), 18.

Whang, S., "Contracting for Software Development," *Management Science* 38, 3 (March 1992), 307–324.