12-31-1994

The Measurement of Information Systems Effectiveness: Evaluating a Measuring Instrument

Judy Scott
University of California, Irvine

Benn Konsynski
Emory University

Robert Blanning
Vanderbilt University

David King
COMSHARE

THE MEASUREMENT OF INFORMATION SYSTEMS EFFECTIVENESS: EVALUATING A MEASURING INSTRUMENT

Judy E. Scott
Graduate School of Management
University of California, Irvine

ABSTRACT

Information system effectiveness is an important phenomenon for both researchers and practitioners. Despite widespread interest, and the importance of the uses, there have been no efforts to validate *Computerworld's* Premier 100 rankings of information system effectiveness. This paper uses structural equation modeling in an attempt to validate the measuring instrument used to derive the *Computerworld* rankings.

Alternative models for the measuring instrument are proposed. Using a reflective model, the findings raise doubts as to the reliability of the rankings, and both content validity and construct validity are also suspect. The reliability and validity are problematic because multiple indicators of the same construct must be homogeneous for it to make sense to combine them into a composite index. A solution to this problem is to represent information system effectiveness as a multidimensional construct that is part of a causal model. Based on previous research in the area, suggestions are offered to improve the measuring instrument.

1. INTRODUCTION

For the past six years, *Computerworld* (CW) has been rating Fortune 500 corporations on their information systems effectiveness. Organizations that score in the top one hundred are designated as the Premier 100 users of information systems. As one of the few readily available secondary data sources in the information technology domain, this data set has generated much interest among researchers as well as IS practitioners. Researchers have attempted to find a relationship between IT and business success (Strassman 1990; Mahmood and Mann 1991; Ahituv and Giladi 1993; Sethi, Hwang and Pegels 1993), while IS practitioners have justified budgets and benchmarked themselves on the basis of the Premier 100 rankings.

Despite this widespread interest in the CW ratings, there have been no efforts to validate these rankings as an instrument measuring the "quality" or "effectiveness" of the IS function. This paper uses structural equation modeling in an attempt to validate the measuring instrument used to derive the CW rankings. A validated measuring instrument provides a standardized evaluation mechanism that enables comparisons across departments, systems, users, organizations, and industries (Baroudi and Orlikowski 1988; Ives, Olson and Baroudi 1983); it promotes research utilizing a tested instrument and thus supports the triangulation of results; allows replication; and saves time for development of new instruments (Baroudi and Orlikowski 1988; Straub 1989). Such formal instrument validation helps build a cumulative research tradition; it provides improved measurement of research variables; it helps improve the clarity of research questions; and it results in more meaningful variable relationships (Baroudi and Orlikowski 1988; Straub 1989). The use of unvalidated instruments, on the other hand, causes uncertainty in interpreting research findings and offers no protection against the effects of confounding variables (Straub 1989). As pointed out by Straub, inadequate instrument validation has been a major shortcoming of MIS research.

Confidence in a measuring instrument depends on its reliability and its validity. An instrument with high reliability measures consistently without error; reliability can be assessed by techniques such as Cronbach's alpha or squared multiple correlations (SMCs) in structural equation modeling (SEM), which represent the proportion of variance in a measure that is explained by the variables that directly affect the indicator. This latter approach is able to cope with correlated errors of measurement and is more general in allowing the specification of models with more than one latent or observed variable (Bollen 1989).

Validity in measuring instruments is concerned with whether a variable measures what it is supposed to measure (Bollen 1989). Content validity assures the theoretical meaningfulness of a concept (Bagozzi 1979, 1980) and the logic of the underlying analysis (Pedhazar and Pedhazar Schmelkin 1991), while construct validity assesses the quality of correspondence between a theoretically-based construct and its operationalized measures.

The next section of this paper reviews the information systems effectiveness literature. Section 3 discusses CW's measuring instrument and evaluates its content validity by comparing CW's approach with the research literature. Alternative models for IS effectiveness are proposed, highlighting the difference between formative and reflective models. Section 4 attempts to develop certain reflective and formative measurement models corresponding to the CW measures and assesses their reliability and construct validity using confirmatory factor analysis with LISREL. In addition, descriptive statistics and Cronbach's alpha tests are used to search for further evidence of reliability and validity in the reflective model. The results of the analysis are summarized in section 5 and section 6 offers some concluding remarks.

2. THE INFORMATION SYSTEMS EFFECTIVENESS CONSTRUCT

Measuring IS effectiveness has been an important research issue in the literature. DeLone and McLean (1992) offer the most comprehensive examination of previous research in this area, citing over 180 articles. Their taxonomy of information systems success yields six underlying categories: system quality, information quality, use, user satisfaction, individual impact and organizational impact. They show relationships between the six categories with an interdependent IS success model that implies a causal relation from system and information quality to individual and organizational impacts.

Since IS effectiveness is presumed to increase user satisfaction, measures of user information satisfaction (Bailey and Pearson 1983; Ives, Olson and Baroudi 1983; Baroudi and Orlikowski 1988) and end-user information satisfaction (Cheney, Mann and Amoroso 1986; Doll and Torkzadeh 1988, 1989, 1991; Etezadi-Amoli and Farhoomand 1991; Henderson and Treacy 1986; Igbaria 1990; Rivard and Huff 1988) have been frequently suggested to measure IS effectiveness. Also, user involvement has been linked to MIS success, system usage and information satisfaction (Ives and Olson, 1984; Baroudi, Olson and Ives 1986).

Srinivasan (1985) examined alternative measures of system effectiveness. He compared behavioral measures of system effectiveness and system usage with Jenkins and Ricketts' (1979) measures of perceived effectiveness and concluded that system usage is not an appropriate surrogate for perceived system effectiveness in the case of model-based decision support systems.

Hamilton and Chervany (1981) define system effectiveness in terms of a goal-centered view, which compares performance to objectives through specific criterion measures, and a normative systems-resource view with expected standards for the quality of the system, service levels and human resource issues. They present a framework of effectiveness-oriented objectives which suggests a multidimensional IS effectiveness construct and measures performance at three levels: (1) information and support provided; (2) use process and user performance; and (3) organizational performance.

In a similar vein, Daft (1989) and Sethi, Hwang and Pegels (1993) consider resource and goal paradigms. Organizational theory looks at organizational effectiveness along the dimensions of internal versus external focus (Daft 1989). Externally, their open systems model deals with resource acquisition and their rational goal model centers on productivity and profit, while internally they posit models of internal processes and human relations. Raghunathan and Raghunathan (1991) view information systems planning effectiveness as a multidimensional construct as well. Effectiveness criteria relate to the level of achievement of objectives such as predicting future trends, improving short-term and long-term IS performance, and resource allocation.

In Figure 1, we integrate the models of Daft (1989), Hamilton and Chervany (1981), and DeLone and McLean (1992) into an input, process, output framework that is commonly used in economics and IT business value research (Crowston and Treacy 1986). In the next section we assess the CW IS effectiveness measure, map the CW measures to the underlying theoretical constructs, discuss the content validity of the measuring instrument and present alternative measurement models.

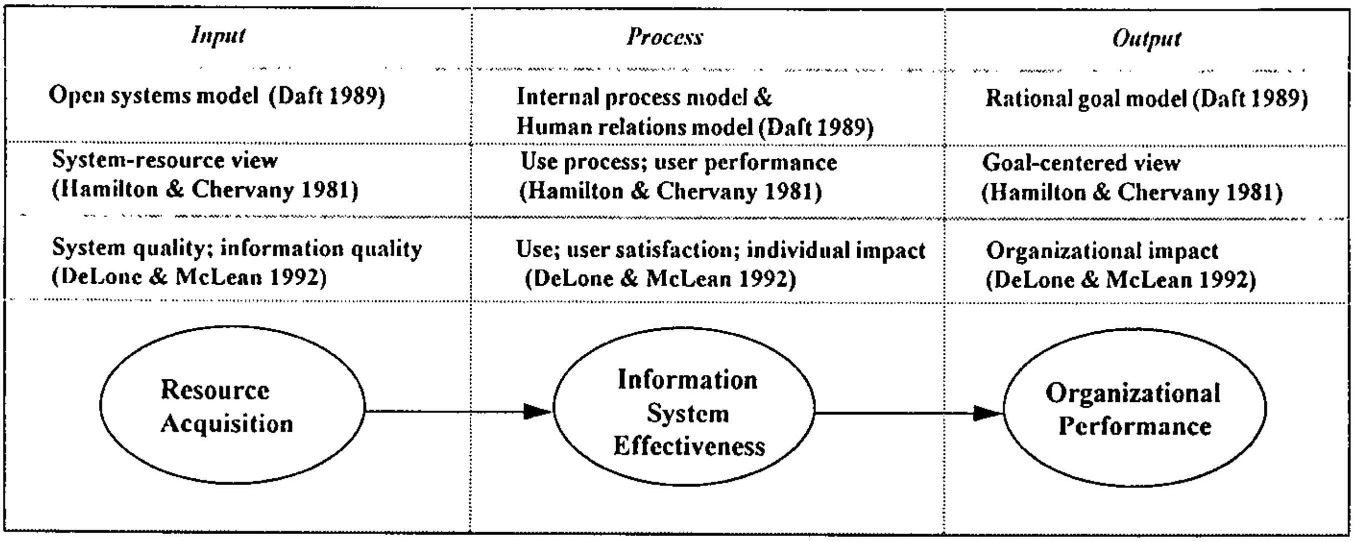


Figure 1. An Input Process Output Framework for IS Effectiveness

3. ASSESSING AN INFORMATION SYSTEMS EFFECTIVENESS MEASURE

3.1 The Evolution of the *Computerworld* Information System Effectiveness Measure

CW has rated large companies, most of which are in the Fortune 500, on the basis of their IS effectiveness, for the past six years, using weighted additive models. In 1988, the first year of the study, the measures used were *variance of the IS budget as a percentage of revenue from the industry average, mainframe/mini processor market value as a percentage of revenue, staff spending as a percentage of IS budget, training spending as a percentage of IS budget, PCs and terminals per employee* and *average profit increase over the previous five years*. In 1989 and 1990 the measuring instrument was essentially identical, while in 1991 and 1992 *peer assessment* was added, and then in 1993, ratings based on various criteria from *CEOs* and *IS managers* were included (see Figure 2).

Although CW described the criteria for peer, CEO and Management ratings, they did not report data for these criteria, only reporting composite scores for each rating.

The weightings of the components constituting CW's total scores, shown in Table 1, demonstrate a decreasing emphasis of the IS budget. In particular, the weighting of the IS budget drops from 30% in the early years (1988-1990) to 10% for 1993. Another trend is the increasing emphasis on subjective ratings, with peer rating weighting increases from 15% in 1991, to 30% in 1992, and 20% in 1993. Also, when two more subjective measures are added in 1993, management rating and executive rating, 40% of the total score becomes dependent on subjective measures.

3.2 Advantages and Disadvantages of *Computerworld's* Measuring Instrument for IS Effectiveness

CW's measuring instrument for information system effectiveness has benefitted from the inputs of several consultants and the experience of the CW staff. Data have been collected for a large sample: approximately 600 companies were surveyed in most years. Until recently most of the measures were objective and thus are likely to be relatively unbiased in contrast to perceptual measures such as user information satisfaction. The quality of such objective data is probably high; the IS budget, processor market value, percentage of budget spent on staff and training and PCs/terminals per employee are likely to be well documented and available to respondents. Perceptual measures, such as ratings by peers, CEOs and IS managers are more susceptible to bias, but have the advantage of assessing IS effectiveness more directly and attempt to measure process rather than inputs or outputs (see Figure 3).

Peer	CEO	Management
Competitive IS Integrated strategy Bottom-line advantage Cost-effectiveness	User satisfaction Support strategic plans Manage risk World-class tools Understand corporate goals Involvement Develop systems quickly IS education of users Integration	Org. & IS structure compatible Strategic alignment CIO reporting level Top mgmt technology perception User feedback from UIS surveys IT investment criteria Focus of activities (reengineering) Application quality User spending IS budget % for R&D Staff deployment % maintenance

Figure 2. The Criteria for Peer, CEO and Management Ratings

Input	*Process*	*Output*
Open systems model (Daft 1989)	**Internal process model & Human relations model (Daft 1989)**	**Rational goal model (Daft 1989)**
System-resource view (Hamilton & Chervany 1981)	**Use process; user performance (Hamilton & Chervany 1981)**	**Goal-centered view (Hamilton & Chervany 1981)**
System quality; information quality (DeLone & McLean 1992)	**Use; user satisfaction; individual impact (DeLone & McLean 1992)**	**Organizational impact (DeLone & McLean 1992)**
Resource Acquisition →	**Information System Effectiveness** →	**Organizational Performance**
Budget Value Staff Training PCs & Terminals World-class tools (CEO) IT investment criteria (Mgmt) Application quality (Mgmt) User spending (Mgmt) IS budget % for R&D (Mgmt) Staff deployment % maintenance (Mgmt)	Integrated strategy (Peer) User satisfaction (CEO) Support strategic plans (CEO) Manage risk (CEO) Understand corporate goals (CEO) Involvement (CEO) Develop systems quickly (CEO) IS education of users (CEO) Integration (CEO) Org. & IS structure compatibility (Mgmt) Strategic alignment (Mgmt) CIO reporting level (Mgmt) Top mgmt technology perception (Mgmt) User feedback from UIS surveys (Mgmt) Focus of activities (reengineering) (Mgmt)	Profit Bottom-line advantage (Peer) Cost-effectiveness (Peer) Competitive IS (Peer)

Figure 3. Mapping *Computerworld* Measures to a Theoretical Framework

Table 1. Weightings of Components

	1988	1989	1990	1991	1992	1993
Sample					600	600
Data Source	IS executive	IS executive	IS executive	IS executive	IS executive	IS executive CEO Peers
IS budget/revenue variance	30%	30%	30%	15%	15%	10%
Processor value/revenue	15%	15%	15%	15%	15%	10%
IS budget for staff	10%	10%	10%	10%	10%	10%
IS budget for training	15%	15%	15%	15%	15%	10%
PCs/terminals/employee	15%	15%	15%	15%	15%	10%
Avg. 5 year profit increase	15%	15%	15%	15%	15%	10%
Peer rating	n/a	n/a	n/a	15%	30%*	20%
Management rating	n/a	n/a	n/a	n/a	n/a	10%
Executive rating	n/a	n/a	n/a	n/a	n/a	10%

*These weights as reported in *Computerworld* 1992 sum to 115%.

Longitudinal data from six consecutive years of conducting this study provide data over time that could be useful for establishing causality in hypothesized relationships. Since 1991, firms have been ranked only within their industry rather than in the entire Premier 100 group, implying that comparison across industries is relatively meaningless. In addition, the data collected for the model are fairly parsimonious, although the trend over the years has been to increase this number.

CW's sample is biased toward large companies, mainly Fortune 500 organizations. The effect of this bias is not known, but higher coordination requirements are likely and consequently more difficulties with systems integration and possibly lower IS effectiveness. On the other hand, with larger IS budgets, these companies have more opportunities to invest in emerging technologies that might facilitate IS effectiveness.

3.3 The Relationship of *Computerworld's* Measures to the Underlying Theoretical Constructs

CW's weighted additive model does not address the multiple levels of measures that previous research has posited. Many items used to measure IS effectiveness are actually measures of antecedents (Treacy 1985; Doll and Torkzadeh 1991; Sethi, Hwang and Pegels 1993) better represented with a causal model. A causal model shows antecedents determining other constructs that follow in time. Budget, processor market value, and PCs and terminals are inputs, resources or information and support provided, while financial measures such as profit increase are measures of outputs, organizational performance, or a goal-centered view. So, if IS effectiveness is defined in terms of the relevant processes, then profit is not a measure of IS effectiveness per se. Although profit is possibly a result of IS effectiveness, confounding factors often make this difficult to ascertain (Strassman 1990; Mahmood and Mann 1991; Ahituv and Giladi 1993; Sethi, Hwang and Pegels 1993; Brynjolfsson 1993).

The perceptive ratings by peers, CEOs and IS management are intended to measure the process aspects of IS effectiveness, but on closer inspection of the actual criteria used for these ratings in 1993 (see Figures 2 and 3), it becomes clear that they reflect different perspectives, spanning attributes of inputs, processes and outputs. In particular, the peer rating is externally focused while the IS management rating is concerned mostly with the internal IS department and the CEO rating is internal but with an enterprise outlook. Nevertheless, each realizes the importance of strategic alignment between IS and the corporation. In addition, both the CEO and IS management ratings consider user issues, such as user satisfaction, education of users, frequency of user feedback, and user spending.

3.4 Content Validity

Bollen (1989) sees content validity as

> a qualitative type of validity where the concept is made clear and the analyst judges whether the measures fully represent the domain....To know the domain of a concept, we need a theoretical definition...to reflect the meanings associated with the term in prior research.

Content validity indicates representativeness and sampling adequacy (Bollen 1989). Relevant theories, research findings and literature need to be examined for construct definition (Bollen 1989; Bagozzi 1979, 1980). The definition of the construct should avoid being vague, ambiguous and tautological, and consistency is required between the content of items to be used as measures of the construct and its definition. Bollen stresses the importance of specifying the scope of the construct in the definition, since this scope helps in identifying the underlying dimensions of the construct.

It is difficult to provide a detailed definition of information systems effectiveness. Because of the complexity and lack of a standardized precise definition (Bakos 1985), the dimensions of the IS effectiveness construct are difficult to clearly identify and the scope of the construct is not made explicit (Bollen 1989). Nevertheless, using our framework (see Figure 3), we find support to show that CW, with its extensive experience in IT reporting and the advice of consultants, is on the right track and making a worthwhile contribution in this regard. As seen on Figure 3, the CW measures map to the input-process-output theoretical framework, showing consistency between theory and operationalizations. However, there are problems with components of composite ratings (peer, CEO and management) crossing over between theoretical constructs.

3.5 Alternative Measurement Models

Measurement models typically use effect indicators to represent the construct (Byrne 1989; Pedhazar and Pedhazar Schmelkin 1991) and these models are called reflective or factor-analytic models. On the other hand, formative or causal measurement models that use causal indicators are often appropriate,1 although relatively neglected in the literature. Information system (IS) effectiveness cannot be measured directly and so is a latent variable that CW operationalizes with their observed variable measures. We need to consider whether the CW measures are cause or effect indicators of IS effectiveness. Effect indicators *depend on* one or more latent variable, while cause indicators *influence* latent variables (Bollen 1989).

CW uses a weighted additive index to measure IS effectiveness, which could be represented by either a reflective or a formative measurement model, depending on the direction of influence between the unobservable and observable variables. Reflective indicators attempt to account for observed variables while formative indicators are used to minimize residuals in the structural relationship (Fornell and Bookstein 1982).

If the measurement model underlying the CW data is reflective, there is the assumption of an underlying unidimensional IS effectiveness construct. This would contradict the substantial literature suggesting that IS effectiveness is of a multidimensional nature (Hamilton and Chervany 1981; Daft 1989; Raghunathan and Raghunathan 1991; DeLone and McLean 1992; Sethi, Hwang and Pegels 1993). Thus, the CW instrument is likely to have measurement and validity problems if it implements a reflective measurement model. We use SEM to investigate this in sections 4 and 5.

Alternatively, the CW data may be intended to represent a formative measurement model (such as an index for GNP). In this case, the indicators are not necessarily highly correlated. In some years, certain components are clearly reflective (e.g., ratings by peers, CEO and IS management), which implies a formative/reflective hybrid model (Fornell and Bookstein 1982, Jöreskog and Sörbom 1989, Bollen 1989) as the underlying measurement model. In sections 4 and 5, we attempt to use SEM to investigate the validity of this type of measurement model as well.

4. METHODOLOGY

4.1 Measurement Models

Structural equation modeling is a methodology that can be used to validate measurement models. Using structural equation modeling conventions, models are represented in a path diagram. The latent (unobservable) variables are enclosed in ovals, observed variables are enclosed in rectangles, error variables are not enclosed and the direction of the arrows between indicators and their latent variable depends on the direction of influence. Figure 4 represents the CW data as a unidimensional reflective measurement model, with arrows indicating the influence of the latent variable on its effect indicators, which are assumed to have some measurement error and not be perfectly measured. (The description of the variables can be found in Section 5, Table 2.)

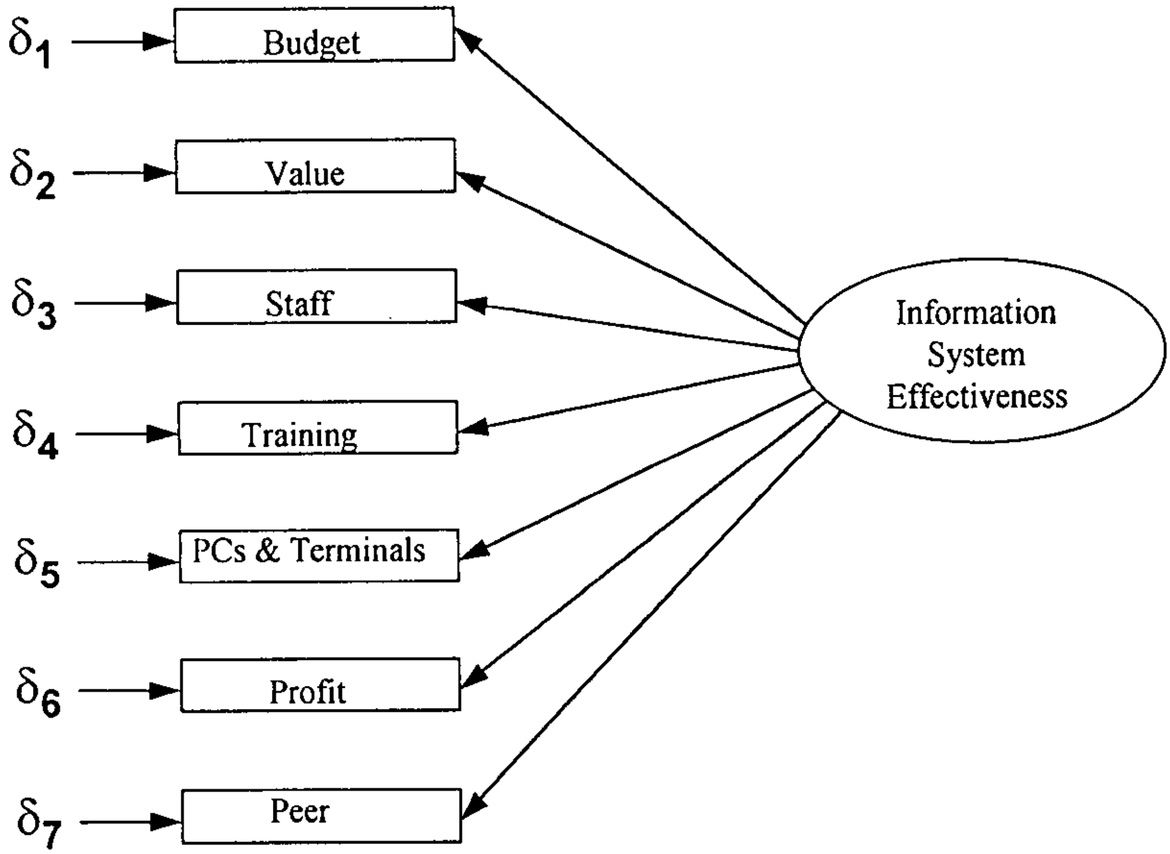


Figure 4. *Computerworld's* 1991 Reflective Measurement Model for IS Effectiveness

CW has published data on most of the previously discussed measures for the Premier 100 companies over the period 1988 to 1993. We analyzed this data using confirmatory factor analysis with LISREL, a SEM software package, to determine the reliability and construct validity of the reflective measuring instrument shown in Figure 4. Section 5 presents the results for reliability (Table 2) and the results for construct validity (Table 4). We examined the reliability of this measuring instrument using Cronbach's alpha (the results for this test are in Table 3). We also analyzed the data with scatterplots (see the appendix) and correlations of measures to total scores (shown in Table 6).

Figure 5 shows a formative model that separates peer and profit from the other measures, based on the 1991[2] data. The rationale, consistent with previous research, is that budget, value, staff, training and PCs do not really measure IS effectiveness but rather are possible predictors (Treacy 1985), while peer ratings measure the construct more directly and profit results from increased organizational performance, to which IS effectiveness may be a contributor. Measures of organizational performance other than profit could be obtained from secondary data sources, but this is beyond the scope of the present study. In any case, we realize that the link between IS effectiveness and organizational performance is tenuous. For example, recent studies question the sustainability of competitive advantage through IT investments. Frequently, successful IT investments are copied by competitors and so may not increase profits to any firm in the industry but instead result in competitive parity (Clemons and Row 1991).

4.2 Determining the Appropriate Measurement Model

Fornell and Bookstein state there are three major considerations in making the choice between reflective and formative modes: (1) the study objective; (2) theory; and (3) empirical contingencies. For example, objectives that explain unobserved variance, theory that conceives constructs as explanatory combinations of indicators and low

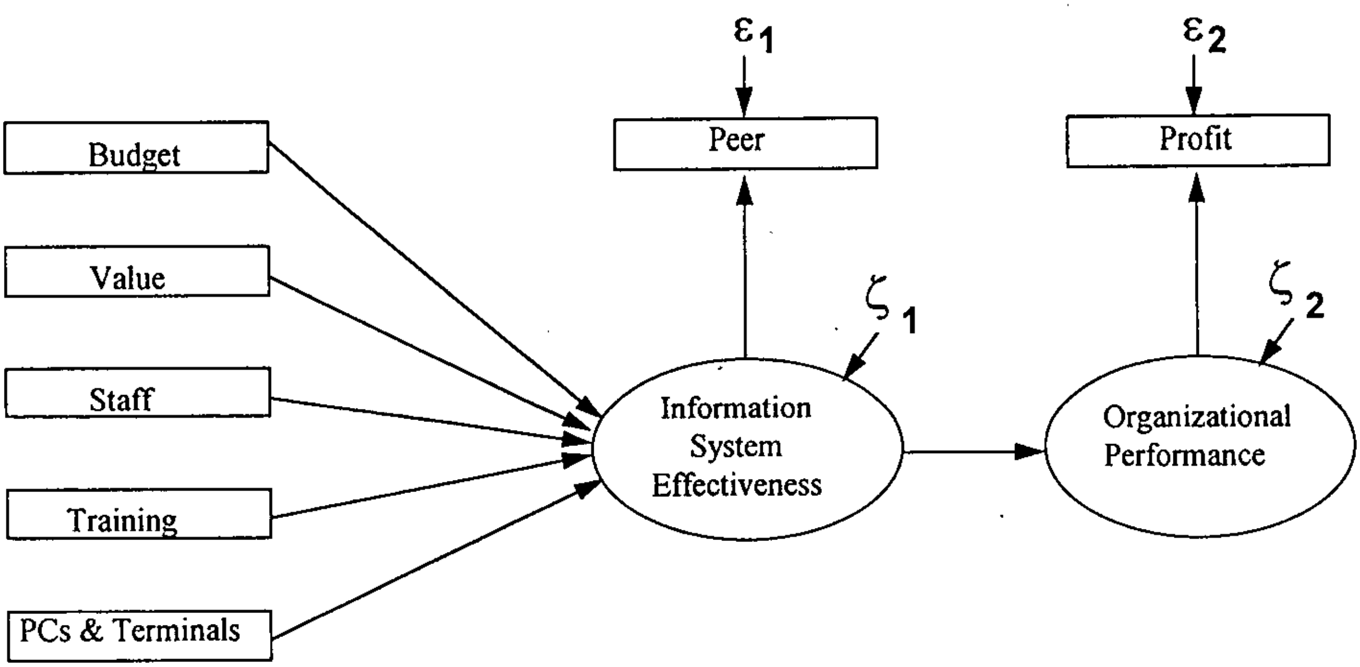


Figure 5. Formative Model (1991 Data)

collinearity between observed variables, lend weight to using formative indicators or a combination of formative and reflective modes. Similarly, Bollen uses a "mental experiment" test to choose the appropriate model. A change in the latent variable is imagined and then the researcher decides whether it is reasonable to expect a subsequent change in the observed variables. If so, the reflective mode is likely. Alternatively, the reverse scenario represents the formative model.

Bollen's "mental experiments" and Fornell and Bookstein's considerations lend credence to identifying Budget, Value, Staff, Training and PCs as causal indicators, but Profit and the subjective Peer, CEO and Management ratings are not likely to influence IS effectiveness and should thus remain as reflective indicators. The resulting model is shown in Figure 6.

Using multiple indicators for IS effectiveness would result in a MIMIC model; these models contain observed variables that are Multiple Indicators and MultIple Causes of a single latent variable (Fornell and Bookstein 1982, Jöreskog and Sörbom 1989; Bollen 1989). According to Fornell and Bookstein, the latent variable in a MIMIC model is an exact linear combination of its indicators, with the error term of the unobservable variable attributed to measurement errors for the correlated effect indicators. Bollen shows that the constraint of zero variance for the latent variable makes it a "weighted index" of its cause indicators. Since formative indicators and the latent variable are assumed to have no errors (in LISREL terminology the error covariance matrix theta delta and disturbance term zeta are zero), issues of reliability become irrelevant for the formative indicators. However, error terms still apply to the reflective indicators in the MIMIC model.

Correlation matrices[3] for all years show low correlations (ranging from .0004 to .464 with most below .3) between the measures, making it unlikely that homogeneity or internal consistency will exist in an additive model. Although this would be a problem for a reflective model, homogeneity and internal consistency are not expected for formative indicators. For the MIMIC model in Figure 6, for example we would not necessarily expect high correlations between Budget, Value, Staff, Training or PCs (as defined in Table 2); high correlations would still be desirable, however, among the subjective ratings of Peer, CEO and Management.

5. RESULTS

All methods of analysis reveal problems with the reflective measuring instrument. Scatterplots show little if any relationship between the measures and the total score (see the appendix). Correlation matrices show low and even

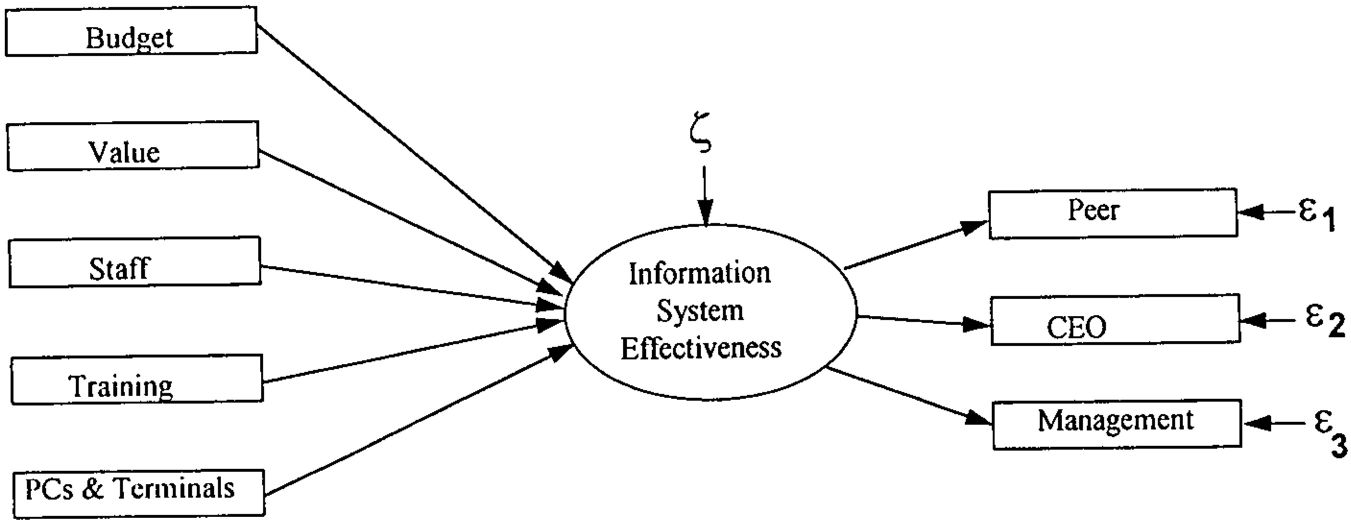


Figure 6. MIMIC Model (1993 Data)

some negative correlations. The only measure that CW posits as negatively correlated with IS effectiveness is the percentage of the IS budget allocated to staff, yet many of the other measures are in fact negatively correlated to the total score. In the following subsection, we show the results of the confirmatory factor analysis with LISREL, including reliability and construct validity assessments. We also use Cronbach's alpha to test the reliability of the reflective measuring instrument.

5.1 Reliability of the Reflective Model

An instrument with high reliability measures consistently with little or no error. The systematic variance is high, while the remaining variance due to random error is low. Squared multiple correlations (SMCs) represent the proportion of variance in a measure that is explained by the variables that directly affect the indicator. The SMC is a measure of the strength of relationship between the observed variable and the latent variable it attempts to measure. Thus the SMCs show how well the observed variables serve *separately* as measurement instruments for the latent variable (Jöreskog and Sörbom 1982). Similarly, the coefficient of determination measures how well the observed variables serve *jointly* as measurement instruments for the latent variable (Jöreskog and Sörbom 1982).

Table 2 contains reliability estimates obtained from LISREL; the SMCs indicate the reliability of the observed variables and the coefficient of determination indicates the reliability of the entire model. These reliability estimates are expected to be positive and less than one. The smaller the error variance, the better the reliability. SMC estimates above one (budget in 1988, value in 1990) and negative estimates (1993) are not meaningful and are associated with non positive definite matrices, probably due to misspecification of the model (Bollen 1989). In 1988 and 1990, the error covariance matrices (theta delta) are not positive definite, and in 1993, the covariance matrix (phi) is not positive definite. Indefinite matrix estimates are typically caused from model misspecification, too little information provided by the data, outliers and nonnormalities, too many parameters in the model or empirical underidentification (Wothke 1993). Jöreskog and Sörbom (1982) consider unreasonable results, such as non positive definite matrices, to indicate that the model is fundamentally wrong and is not suitable for the data. Consequently, despite the high overall goodness of fit for the model with the 1988, 1990 and 1993 data (Table 5), results for these years should be viewed as suspect and in disagreement with the measurement model of Figure 4.

Squared multiple correlation values for the variables in 1989, 1991, and 1992 range from a low of .000 to a high of .447, indicating that over half the variance of the measure is due to random error and that the latent variable, IS effectiveness, which CW's total score is presumed to measure, accounts for at most .447 of the variance of any

Table 2. Reliability (Squared Multiple Correlations LISREL)

Variable	Description	1988	1989	1990	1991	1992	1993
Budget	IS budget/revenue variance from industry average	1.303	.418	.211	.226	.447	-.018
Value	Mainframe and mini processors market value/revenue	.085	.313	1.028	.279	.436	-.001
Staff	Percentage of IS budget spent on IS staff	.042	.103	.001	.051	.000	-.030
Training	Percentage of IS budget spent on IS staff training	.002	.145	.064	.049	.000	-.109
PCs and Terminals	Percentage of PCs and terminals per employee	.000	.033	.015	.319	.246	-.001
Profit	Profit increase over the previous five years		.003		.225	.024	
Peer	Perceived IS effectiveness by competitors	n/a	n/a	n/a	.147		-.001
CEO	Perceived IS effectiveness by CEO	n/a	n/a	n/a	n/a	n/a	-.403
Mgmt	Perceived IS effectiveness by IS management	n/a	n/a	n/a	n/a	n/a	-.072
	Total coefficient of determination		.599		.632	.659	-1.010

Table 3. Reliability Cronbach's Alpha

	1988	1989	1990	1991	1992	1993
Cronbach's Alpha	.059	.032	.080	.102	-.148	.094
Standardized Item Alpha	.275	.398	.367	.363	.180	.179

measure. Staff and training measures are particularly unreliable, with not more than .145 variance explained by IS effectiveness. Although Byrne evaluates SMCs of around .3 and .4 as reasonable, explained variance values greater than .8 have been considered acceptable in other IS studies (Baroudi and Orlikowski 1988), so we consider the SMC values from this analysis as too low for satisfactory reliability.

The coefficient of determination shows how well the indicators jointly measure the latent variable and is a generalized measure of reliability for the whole measuring instrument. This coefficient has values of .599, .632, and .659, respectively, for the 1989, 1991, and 1992 data, indicating a relatively poor relationship between the observed variables and the latent variable, implying that almost half the variance of the reflective measuring instrument is due to random error.

We also used SPSS to derive Cronbach's Alpha — an alternative statistical method for measuring reliability. Cronbach's alpha is dependent on the number of items and the variances and covariances of the items. According to Pedhazar and Pedhazar Schmelkin,

> To be substantively meaningful, a composite score has to be based on items "measuring the same phenomenon." In other words, responses to items comprising a measure of a construct are expected to be internally consistent. It is on this notion that internal-consistency reliability estimates are based.

Table 3 shows the results from this analysis and further supports the low reliability estimates from the LISREL analysis. Reliability as measured by Cronbach's alpha (Table 3) ranges from -.148 to .102. Since acceptable reliability is at least .5 or .6 and usually over .7 (Pedhazar and Pedhazar Schmelkin 1991), this reflective measuring instrument displays unsatisfactory reliability and its items are not homogeneous and do not seem to be measuring the same phenomenon. For example, the negative value of Cronbach's alpha in 1992 results from the sum of item covariances being negative. This means that increases in some items correspond with decreases in others, which is very unlikely if all items are measuring the same construct. It should be kept in mind, however, that the 1992 data is incomplete as CW gave peer assessment a weight of 30% but did not supply the individual peer scores, except a list of the top performer in each industry.

Table 4. Factor Loadings**

Variable	Description	1988	1989	1990	1991	1992	1993
Budget	IS budget/revenue variance from industry average	1.141	.647*	.460*	.476*	.669*	.211
Value	Mainframe and mini processors market value/revenue	.291	.560*	1.014*	.529*	.660*	-.040
Staff	Percentage of IS budget spent on IS staff	.205	.321*	-.024	.226	.011	-.272
Training	Percentage of IS budget spent on IS staff training	.049	.380*	.254	.221	-.010	.521
PCs and Terminals	Percentage of PCs and terminals per employee	.008	.181	.123	.565*	.496*	.038
Profit	Profit increase over the previous five years		-.052		-.475*	-.156	
Peer†	Perceived IS effectiveness by competitors	n/a	n/a	n/a	.383*		-.059
CEO†	Perceived IS effectiveness by CEO	n/a	n/a	n/a	n/a	n/a	1.000
Mgmt†	Perceived IS effectiveness by IS management	n/a	n/a	n/a	n/a	n/a	-.418

* T values in the output > 2 indicate significance of the measures at .05 probability.

** Standardized solution values reported except for 1993, which has only an intermediate solution. Methods used were Maximum Likelihood for 1989, 1991, and 1992; Unweighted Least Squares for 1988 and 1990. Loadings set to 1 were Budget in 1989; PCs in 1991; Budget in 1992; and CEO in 1993. The pattern (PA statement was used in 1988 and 1990.

† Only composite data values are available.

Since not all items are measured along the same scale, standardized scores are more appropriate, but even standardized item alpha ranges from .179 to a maximum of .398, still too low for satisfactory reliability. Without satisfactory reliability, validity is also suspect (Bollen 1989). In conclusion, assuming a reflective measuring model, the CW measuring instrument cannot be used with confidence.

5.2 Construct Validity of the Reflective Model

The purpose of construct validity is to validate the theory behind the construct. Internal-structure analysis (Pedhazar and Pedhazar Schmelkin 1991) and the internal consistency of operationalizations (Bagozzi 1979, 1980) require that multiple indicators of the same construct must be homogeneous for it to make sense to combine them into a composite index. Factor analysis is typically used to analyze the relationships between a set of indicators and the internal structure of a construct. The higher the factor loading between a construct and an indicator, the greater the effect of the factor on the indicator.

Confirmatory factor analysis using LISREL determines the loadings of each observed variable in the measurement model. Since the units of measurement are not consistent, for example, Budget, Processor Market Value, Staff, Training and PCs are percentages while the ratings are points, the standardized solutions are examined for appropriate weightings. The weights calculated by LISREL do not correspond to the ones used by CW (see Table 4).

High loadings of the underlying factors on instrument indicators are desirable because they confirm the validity of the proposed construct. For example, Ives, Olson and Baroudi found most loadings were greater than 0.5 in their validation of the user information satisfaction measuring instrument. Several of the loadings in Table 4, however, raise doubts about the probability that the corresponding indicators contribute to the measurement of the proposed IS effectiveness construct. Budget and IS value had significant loadings in years 1989 through 1992, although the value for IS value of 1.014 in 1990 is questionable, since loadings should be less than one. Staff and training are significant only in 1989, while PCs/terminals are significant in 1991 and 1992. Profit is significant in 1991, but negatively, which is not the expected sign and therefore suspect. Peer rating was also significant in 1991, but the low incidence of significant loadings overall does not provide support for construct validity. Even testing the construct as a two factor and three factor model did not result in substantial improvements.

Analyzing the results by year, the measurement models for 1989, 1991, and 1992 have the most significant loadings. In 1988, 1990, and 1993, there were problems with matrices being not positive definite. LISREL failed to recreate

Table 5. Goodness of Fit for the Model

	1988	1989	1990	1991	1992	1993
Chi-square/degrees of freedom	.810	1.400	.730	1.670	1.330	.550
Goodness of fit index	.985	.959	.989	.939	.962	.975
Adjusted goodness of fit index	.956	.903	.968	.877	.911	.955
Root mean square residual	.051	.067	.045	.082	.069	

Table 6. Correlations with Total Score

Peer Rating	-.066
CEO Rating	.018
IS Management Rating	.082

the correlation matrix using the predicted model and actual data and, as discussed above in the context of reliability, led to inadmissible results. According to Wothke, this can be due to model misspecification, too little information provided by the data, outliers and nonnormalities, too many parameters, or empirical underidentification. In view of the lack of internal consistency suggested by correlation matrices and Cronbach's Alpha, it is likely that the reflective measurement model is inappropriate (i.e., misspecified).

Despite problems with reliability and factor loadings, each year of analysis reveals high goodness of fit values in LISREL (see Table 5). This indicates that if the measures were more consistent, the models would possibly be adequate.

Another method of assessing construct validity is the correlation between an item and the adjusted total score, obtained by subtracting the item value from the total (Ives, Olson and Baroudi 1983; Baroudi and Orlikowski 1988). This technique makes most sense for the composite[4] items peer, CEO and IS management ratings that are scored directly and thus are not in different units to the total score.

The correlations found range from -0.066 for peer, to 0.018 for CEO, and 0.082 for IS management ratings (Table 6). Compared to other studies (Ives, Olson and Baroudi 1983, Baroudi and Orlikowski 1988, Doll and Torkzadeh 1988), which found most items correlated at greater than 0.5 with total scores, the reflective measuring instrument again performs poorly.

5.3 Results from the Formative Models

Using single indicators for latent variables makes it difficult to identify and separate the different sources of variability of these single indicators, resulting in an underidentified model (Pedhazar and Pedhazar Schmelkin 1991). Consequently, the proposed formative model in Figure 5 is not ideal. We cannot validate this model because the error term of profit was reported as "may not be identified" and the error matrix (psi) was not positive definite.

The second formative model tested was a MIMIC model using 1993 data as shown in Figure 6. The proposed MIMIC model assigns the three subjective ratings as reflective indicators of IS effectiveness and the measures of budget, value, staff, training and PCs as formative predictors of IS effectiveness. A MIMIC model is identified if it has at least two reflective indicators and at least one formative indicator (Bollen 1989). Interpreting IS effectiveness as a "weighted index" of its cause indicators (Bollen 1989), and an exact linear combination of its indicators (Fornell and Bookstein 1982), lends support to this model as the measurement model underlying CW's instrument. This cannot reconcile however, the addition of the effect indicators to the cause indicators in the determination of the total IS effectiveness score. Furthermore, the reflective indicators are not highly correlated as one would expect. In fact, the LISREL solution was non-admissible and the error matrix for the reflective indicators (Theta Epsilon) was not positive definite.

The appeal of this model is that the three "informants" are reporting on IS effectiveness directly, so content validity is enhanced. However, as seen in Figure 3, apart from the similar measure of strategic alignment and integration into company strategy, management, CEO, and peer are rating different concepts. In fact, the correlation between management rating and the other two ratings is quite low, at 0.035 with peer and 0.198 with CEO and the correlation between peer and CEO is -.010.

The model would be improved by standardizing the questions across all informants. Also, the low correlation between IS management and CEO ratings probably indicates some bias, as IS management attempts to promote its achievements, which CEOs do not always appreciate or may not be fully aware. Related to the issue of bias is the concept of different perspectives between the two groups and possibly lack of strategic alignment. With closer working relationships and goal congruence the ratings might converge. Despite the weaknesses of this model, it makes sense logically and has closer ties with previous research findings and theory than the reflective model.

6. CONCLUSION

As shown by the analysis in this paper, we do not yet have an adequate measuring instrument for information system effectiveness. Whether formative or reflective, the weighted model approach has reliability problems. In the formative case, the ratings by peers, CEO and IS management do not show consistency. Also, logical analyses do not support a completely additive model. Many, but not all, of the measures used are antecedents or predictors of information system effectiveness rather than measures of that actual construct. These cause indicators could contribute to a weighted additive index, but it is difficult to reconcile the further addition of the perceptive effect indicators or possible results, such as profit.

For the reflective model, problems exist with the reliability of most of the measures and so some doubt must exist as to the validity of the rankings and total scores. Using SEM, confirmatory factor analysis did not find support for construct validity nor for the specification of this model.

In addition, although availability of PCs and terminals, for example, is probably a necessary condition for IS effectiveness, research findings and the IS literature emphatically point out that the condition is not sufficient. Management practices deeply influence the attainment of systems success. Culture, policies and procedures affect implementation and use of information systems. The most effective users of IT have competent, well trained personnel to facilitate exploitation of the technology's potential.

For the future, we recommend measuring intermediate outputs (Bakos 1987, Bakos and Kemerer 1992) such as operational performance that is time-based, quality-based, and flexibility-based. Such improvement measures can be more easily related to information systems than financial performance. Another suggestion is to measure interorganizational impact from IS. EDI and other electronic links to entities outside the organization are major current contributors to IS effectiveness for many Premier 100 and other organizations.

Our findings point out the immaturity of this field of study and contribute to the refinement of the measuring instrument both for researchers and practitioners. Researchers will benefit from the synthesis of previous work and presentation of new ideas. Hopefully this will lead to further work in this line of research. Practitioners will become aware of the complexity of measuring models and become cautious in placing too much confidence in results of non-rigorous measuring instruments.

Some implications of this study are that further thought needs to be given to designing an IS effectiveness measuring instrument, and caution needs to be exercised in interpreting current resulting measures. Based on theoretical considerations, IS effectiveness is a multidimensional construct and so it seems very likely that a simple additive weighting model will not suffice. Instead, a more complex causal model with antecedents and resultants would probably be representative. We have proposed such models and suggest that the MIMIC model using 1993 data looks promising and probably would be valid if the ratings criteria were standardized.

Despite the fact that CW's measuring instrument does not pass rigorous statistical validity and reliability tests, one must keep in mind that it was not developed as a research tool, but rather as a practitioners' tool to enable benchmarking and to set standards for industries. As such, the data are useful for practitioners to compare their organizations along different dimensions, although it would be advisable to exercise caution in interpreting total scores. With the scarcity of publicly available IS related data, CW has provided a valuable service in publishing their data sets. Much interest has been generated by both practitioners and researchers showing there is a need and demand for such a service. Researchers have successfully used some items of the extended data set with additional data sets to challenge the productivity paradox (Brynjolfsson and Hitt 1993).

Finally, this study did not have available the full sample of approximately 400 firms that responded to the CW survey. The results could be somewhat different with a larger data set, so it would be instructive to repeat our analysis on the full sample.

8. ACKNOWLEDGMENTS

The author would like to thank Yannis Bakos, Vijay Gurbaxani, George Marcoulides and the anonymous ICIS reviewers for their helpful comments.

9. REFERENCES

Ahituv, N., and Giladi, R. "Business Success and Information Technology: Are They Really Related?" *Workshop on Information Systems and Economics*, Orlando, Florida, December 1993.

Bailey, J., and Pearson, S. "Development of a Tool for Measuring and Analyzing Computer User Satisfaction." *Management Science*, Volume 29, Number 5, May 1983, pp. 530-545.

Bakos, J. Y. "Dependent Variables for the Study of Firm and Industry-Level Impacts of Information Technology." In J. I. DeGross and C. H. Kriebel (Editors), *Proceedings of the Eighth International Conference on Information Systems*, Pittsburgh, Pennsylvania, December 1987, pp. 10-23.

Bakos, J. Y. "Toward a More Precise Concept of Information Technology." In L. Gallegos, R. Welke, and J. Wetherbe (Editors), *Proceedings of the Sixth International Conference on Information Systems*, Indianapolis, Indiana, December 1985, pp. 17-24.

Bakos, J. Y., and Kemerer, C. F. "Recent Applications of Economic Theory in Information Technology Research." *Decision Support Systems*, Volume 8, 1992, pp. 365-386.

Bagozzi, R. P. *Causal Models in Marketing*. New York: John Wiley and Sons, 1980.

Bagozzi, R. P. "The Role of Measurement in Theory Construction and Hypothesis Testing: Toward a Holistic Model." In O. C Ferrell, S. W. Brown, and C. W. Lamb, Jr. (Editors), *Conceptual and Theoretical Developments in Marketing*. Chicago: American Marketing Association, 1979, pp. 15-32.

Baroudi, J. J., Olson, M. H., and Ives, B. "An Empirical Study of the Impact of User Involvement on System Usage and Information Satisfaction." *Communications of the ACM*, Volume 29, Number 3, 1986, pp. 232-238.

Baroudi, J. J., and Orlikowski, W. J. "A Short Form Measure of User Information Satisfaction: A Psychometric Evaluation and Notes on Use." *Journal of Management Information Systems*, Volume 4, Number 4, 1988, pp. 45-59.

Bollen, K. A. *Structural Equations with Latent Variables*. New York: Wiley-Interscience, 1989.

Bollen, K. A., and Long, J. S. *Testing Structural Equation Models*. Newbury Park, California: Sage Publication, 1993.

Brynjolfsson, E. "The Productivity Paradox of Information Technology." *Communications of the ACM*, Volume 35, December 1993, pp. 66-77.

Brynjolfsson, E., and Hitt, L. "Is Information Systems Spending Productive? New Evidence and New Results." In J. I. DeGross, R. P. Bostrom, and D. Robey (Editors), *Proceedings of the Fourteenth International Conference on Information Systems*, Orlando, Florida, December 1993, pp. 47-64.

Byrne, B. M. *A Primer of LISREL*. New York: Springer-Verlag, 1989.

Cheney, P. H.; Mann, R. I.; and Amoroso, D. L. "Organizational Factors Affecting the Success of End-User Computing." *Journal of Management Information Systems*, Volume 3, Number 1, 1986, pp. 65-80.

Clemons, E. K., and Row, M. C. "Sustaining IT Advantage: The Role of Structural Differences." *MIS Quarterly*, Volume 13, Number 5, September 1991, pp. 274-292.

Crowston, K., and Treacy, M. E. "Assessing the Impact of Information Technology on Enterprise Level Performance." In L. Maggi, R. Zmud, J. Wetherbe (Editors), *Proceedings of the Seventh International Conference on Information Systems*, San Diego, California, December 1986, pp. 377-388.

Daft, R. L. *Organization Theory and Design*. St. Paul, Minnesota: West Publishing Company, 1989.

DeLone, W. H., and McLean, E. R. "Information Systems Success: The Quest for the Dependent Variable." *Information Systems Research*, Volume 3, Number 1, 1992, pp. 60-95.

Doll, W. J., and Torkzadeh, G. "A Discrepancy Model of End-User Computing Involvement." *Management Science*, Volume 35, Number 10, 1989, pp. 1151-1171.

Doll, W. J., and Torkzadeh, G. "The Measurement of End-User Computing Satisfaction." *MIS Quarterly*, Volume 12, Number 2, 1988, pp. 259-274.

Doll, W. J., and Torkzadeh, G. "The Measurement of End-User Computing Satisfaction: Theoretical and Methodological Issues." *MIS Quarterly*, March 1991, pp. 5-10.

Etezadi-Amoli, J., and Farhoomand, A.F. "On End-User Computing Satisfaction." *MIS Quarterly*, March 1991, pp. 1-4.

Fornell, C., and Bookstein, F. "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory." *Journal of Marketing Research*, Volume 19, 1982, pp. 440-452.

Hamilton, S., and Chervany, N. "Evaluating Information System Effectiveness — Part 1: Comparing Evaluation Approaches." *MIS Quarterly*, Volume 5, Number 3, September 1981, pp. 55-69.

Henderson, J. C., and Treacy, M. E. "Managing End-User Computing for Competitive Advantage." *Sloan Management Review*, Winter 1986, pp. 3-14.

Igbaria, M. "End-User Computing Effectiveness: A Structural Equation Model." *Omega-International Journal of Management Science*, Volume 18, Number 6, 1990, pp. 637-652.

Ives, B., and Olson, M. H. "User Involvement in Information Systems Development: A Review of the Research." *Management Science*, Volume 30, Number 5, 1984, pp. 586-603.

Ives, B.; Olson, M. H.; and Baroudi, J. J. "The Measurement of User Information Satisfaction." *Communications of the ACM*, Volume 26, Number 10, October 1983, pp. 785-793.

Jenkins, A. M., and Ricketts, J. A. "Development of an Instrument to Measure User Information Satisfaction with Management Information Systems." Unpublished working paper. Indiana University, Bloomington, November 1979.

Jöreskog, K., and Sörbom, D. *LISREL 7 User's Reference Guide*. Scientific Software, Inc., 1989.

Jöreskog, K., and Sörbom, D. "Recent Developments in Structural Equation Modeling." *Journal of Marketing Research*, Volume 19, 1982, pp. 404-416.

Mahmood, M. A., and Mann, G. J. "Measuring the Impact of Information Technology on Organizational Performance: A Key Ratios Approach." *Proceedings of the Twenty-fourth Hawaii International Conference on Systems Sciences*. Los Alamitos, California: IEEE Press, 1991, pp. 251-258.

Pedhazar, E. J., and Pedhazar Schmelkin, L. *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1991.

Raghunathan, B., and Raghunathan T. S. "Information Systems Planning and Effectiveness: An Empirical Analysis." *Omega-International Journal of Management Science*, Volume 19, Number 2-3, 1991, pp. 125-135.

Rivard, S., and Huff, S. L. "Factors of Success for End-User Computing." *Communications of the ACM*, Volume 31, Number 5, 1988, pp. 552-561.

Sethi, V.; Hwang, K. T.; and Pegels, C. "Information Technology and Organizational Performance: A Critical Evaluation of *Computerworld's* Index of Information Systems Effectiveness." *Information and Management*, Volume 25, Number 4, 1993, pp. 193-205.

Srinivasan, A. "Alternative Measures of System Effectiveness: Associations and Implications." *MIS Quarterly*, Volume 9, Number 3, 1985, pp. 243-253.

Strassman, P. A. *The Business Value of Computers*. New Canaan, Connecticut: Information Economics Press, 1990.

Straub, D. W. "Validating Instruments in MIS Research." *MIS Quarterly*, Volume 13, Number 2, 1989, pp. 147-169.

Sullivan-Trainor, M. (Editor). "The Premier 100." *Computerworld*, September 12, 1988, Section 2, CW Publishing, pp. 1-80.

Sullivan-Trainor, M. (Editor). "The Premier 100." *Computerworld*, September 11, 1989, Section 2, CW Publishing, pp. 1-84.

Sullivan-Trainor, M. (Editor). "The Premier 100." *Computerworld*, October 8, 1990, Section 2, CW Publishing, pp. 1-64.

Sullivan-Trainor, M. (Editor). "The Premier 100." *Computerworld*, September 30, 1991, Section 2, CW Publishing, pp. 1-80.

Sullivan-Trainor, M. (Editor). "The Premier 100." *Computerworld*, September 14, 1992, Section 2, CW Publishing, pp. 1-63.

Sullivan-Trainor, M. (Editor). "The Premier 100." *Computerworld*, September 13, 1993, Section 2, CW Publishing, pp. 1-59.

Treacy, M. E. "An Empirical Examination of a Causal Model of User Information Satisfaction." Working Paper, Center for Information Systems Research, Sloan School of Management, Massachusetts Institute of Technology, 1985.

Wothke, W. "Nonpositive Definite Matrices in Structural Modeling." In K. A. Bollen and J. S. Long (Editors), *Testing Structural Equation Models*. Newbury Park, California: Sage Publishing, 1993, pp. 256-293.

10. ENDNOTES

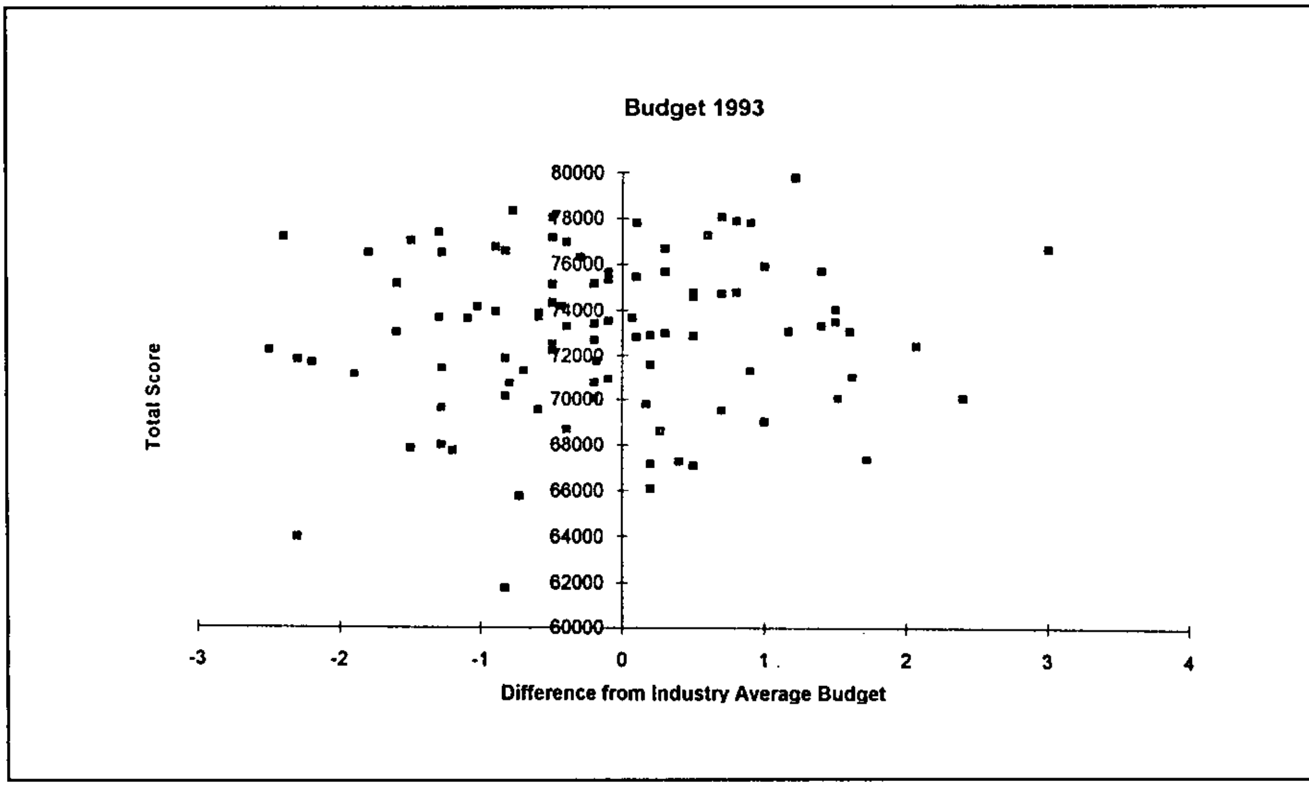
1. Causality is based on the criteria of precedence in time of cause to effect, association between cause and effect, and isolation from other effects (Bagozzi 1980; Bollen 1989).

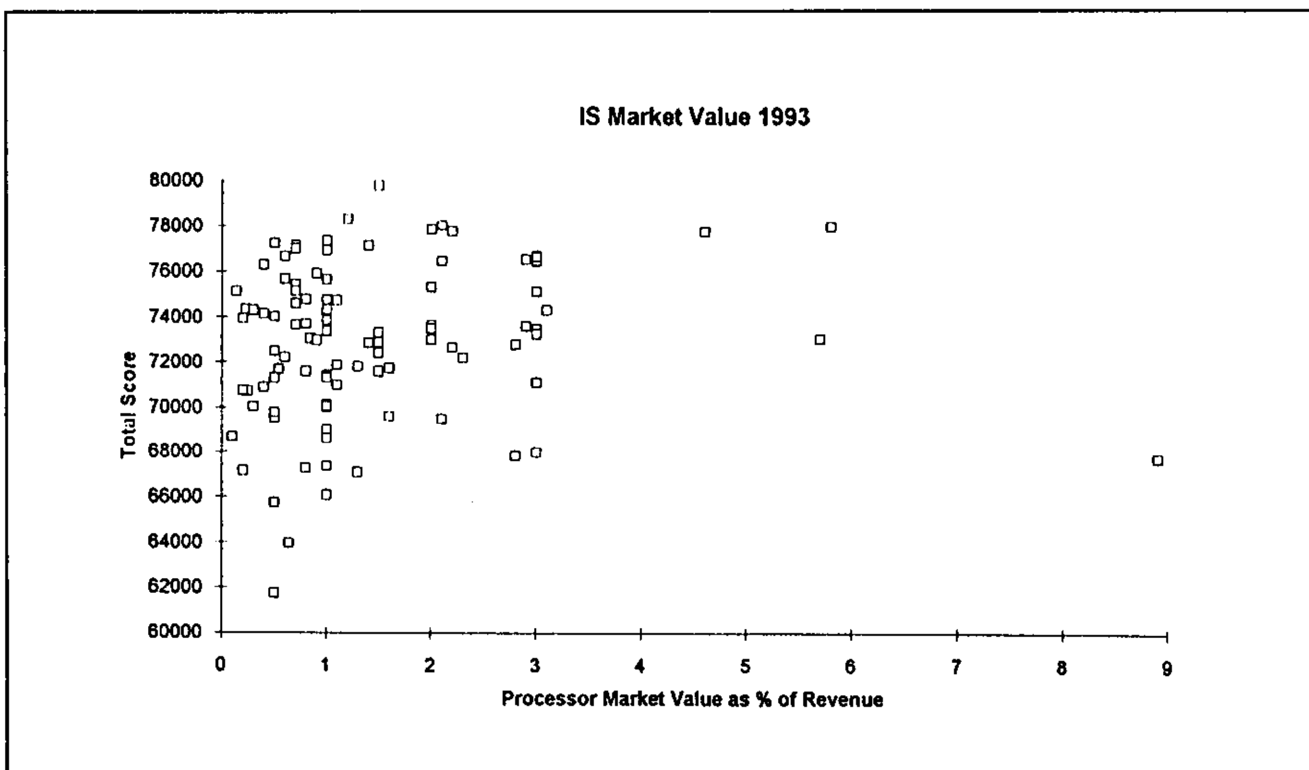
2. In 1992, the only peer data published was for leaders in each industry.

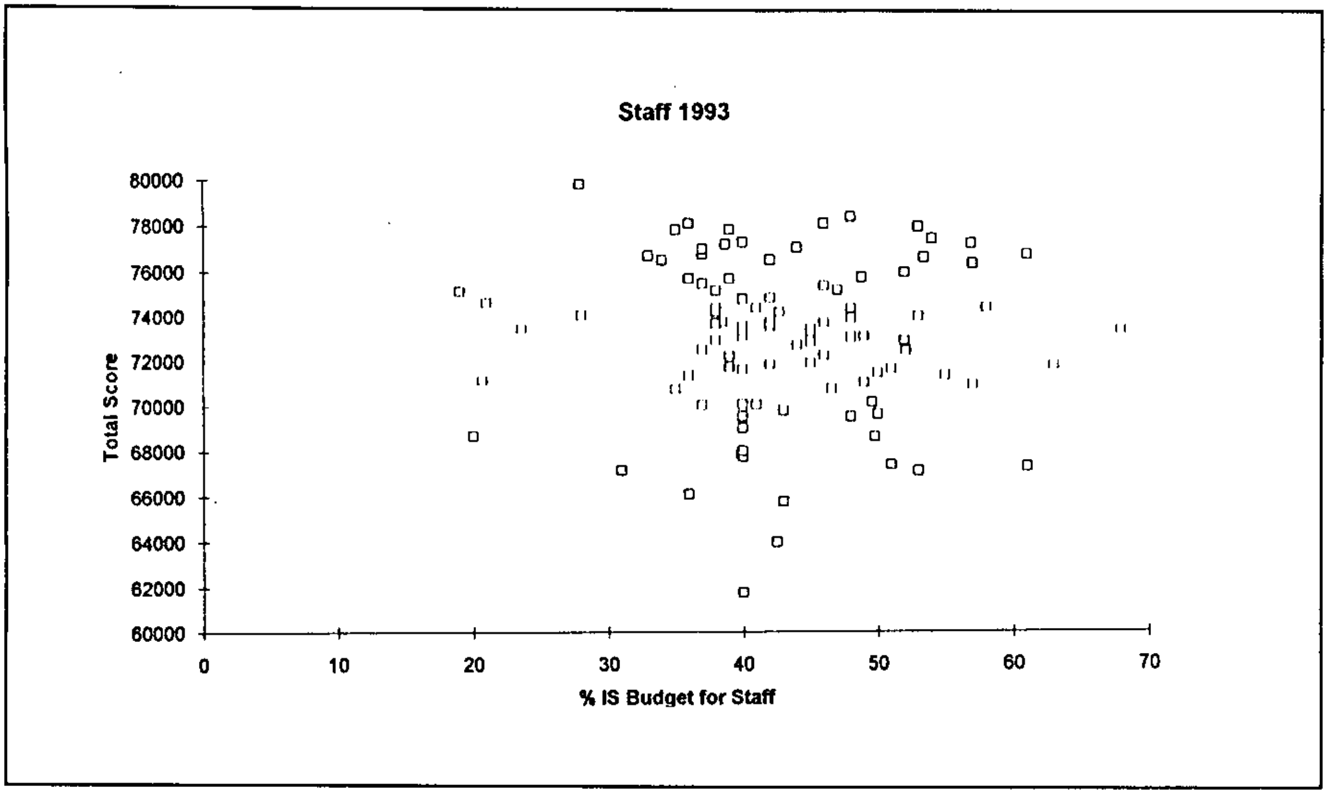
3. Available from author.

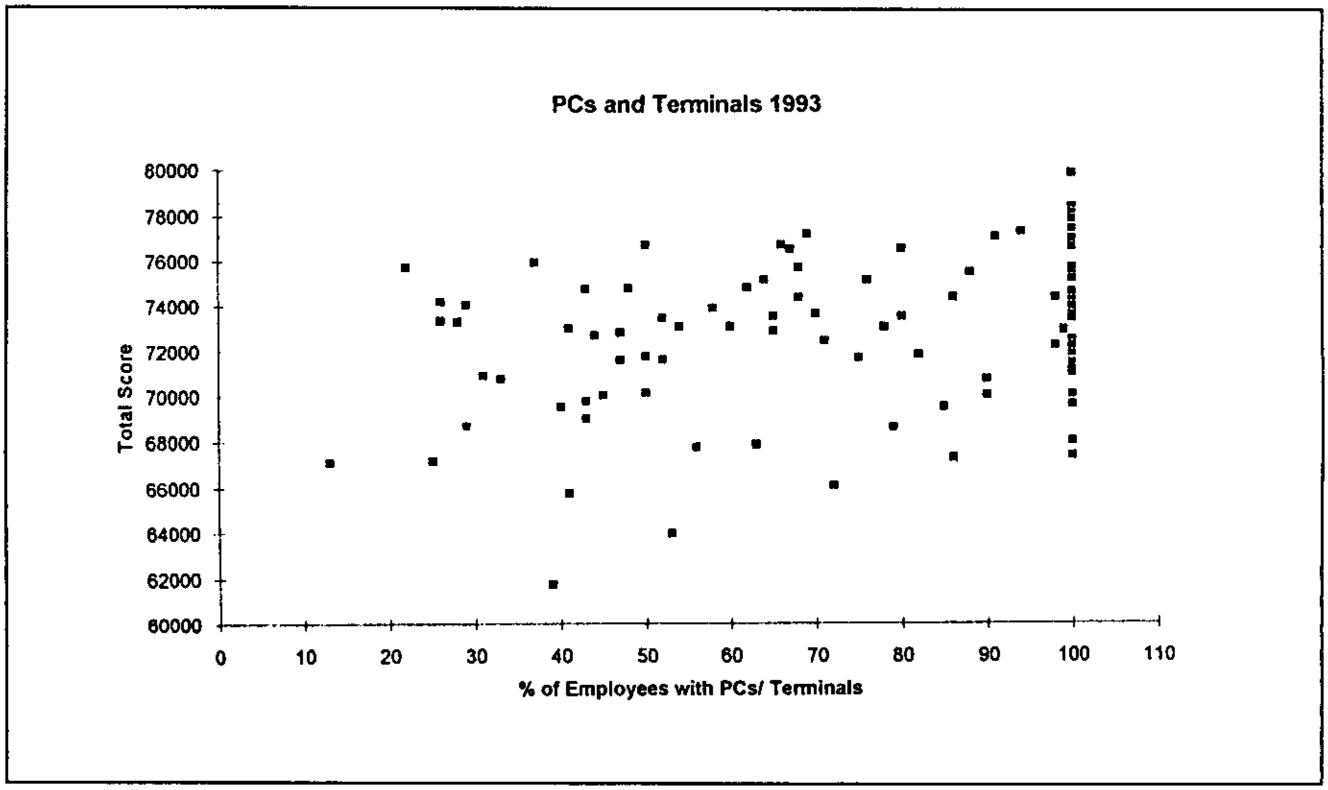
4. Only composite data values were published.

APPENDIX
SCATTERPLOTS OF 1993 TOTAL SCORES VERSUS MEASURES






Staff 1993
Total Score
80000
78000
76000
74000
72000
70000
68000
66000
64000
62000
60000
0
10
20
30
40
50
60
70
% IS Budget for Staff


PCs and Terminals 1993
Total Score
80000
78000
76000
74000
72000
70000
68000
66000
64000
62000
60000
0
10
20
30
40
50
60
70
80
90
100
110
% of Employees with PCs/ Terminals