Issues and Opinions

The Measurement of End-User Computing Satisfaction: Theoretical and Methodological Issues

Measurement issues are receiving increased attention among the MIS research community. This increased attention is quite appropriate. The productivity of substantive research activities depends upon efforts to improve theory and measurement development. In a companion article, Etezadi-Amoli and Farhoomand express several methodological concerns about the measurement of end-user computing satisfaction. Some of these concerns appear to be based upon underlying theoretical assumptions; others are related to confusion concerning the purpose for measuring end-user computing satisfaction or the procedures for developing Likert-type scales. First, we will indentify theoretical issues that guide instrument development. The purpose of the end-user computing satisfaction instrument (Doll and Torkzadeh, 1988) is explained in terms of the research domain in which it was designed to be useful and its role in that domain. Then, we respond to specific methodological concerns.

Theoretical Issues

Etezadi-Amoli and Farhoomand imply that: (1) the primary purpose for measuring end-user computing satisfaction is to predict certain behaviors; and thus (2) the measurement of end-user computing satisfaction should be somehow more closely tied to attitude-behavior theory. Addressing these concerns requires some discussion of the following theoretical questions. What is the domain of MIS research? Is end-user computing satisfaction (EUCS) a dependent or independent variable? What is the purpose of an EUCS instrument? Is it to evaluate an application or to predict behavior? What behavior? To what extent is the attitude-behavior research tradition of social and cognitive psychology applicable to MIS research? These theoretical questions should guide instrument development.

End-user computing satisfaction is an important theoretical construct because of its potential for helping us discover both forward and backward links in a causal chain (i.e., a network of cause and effect relationships that describe a large portion of the domain of MIS research) that are important to the MIS research community (see Figure 1). Thus, end-user computing satisfaction is potentially both a dependent variable (when the domain of one's research interest is upstream activities or factors that cause end-user satisfaction) or an independent variable (when the domain of one's research interest is downstream behaviors affected by end-user satisfaction). To date, the majority of what is considered MIS research has focused on upstream phenomena (e.g., research on design and implementation activities). In this upstream research domain, measures of system success (e.g., end-user computing satisfaction) have been used to evaluate the effectiveness of design and implementation activities.

The downstream research domain is not, as yet, well developed. In this downstream domain, end-user satisfaction assumes the role of an independent variable, and the emphasis is on the function attitudes serve for the user other than the evaluation of an information system or IS staff (Melone, 1990).

Theory should guide instrument development, but concepts should be borrowed from other disciplines with caution. The attitude-behavior literature evolved out of research domains in social and cogitive psychology that are quite different than those experienced by the MIS research community. The research in social and cognitive psychology contains a rich variety of behavioral phonomena and often focuses on emotionally charged issues (e.g., capital punishment, birth control, prejudice, etc.). Thus, attitude research in this domain emphasizes the affective rather than the cognitive (e.g., belief) dimension of attitudes (Chaiken and Stangor, 1987; Zajonc, 1984). Ostrom (1968) contends that the bulk of

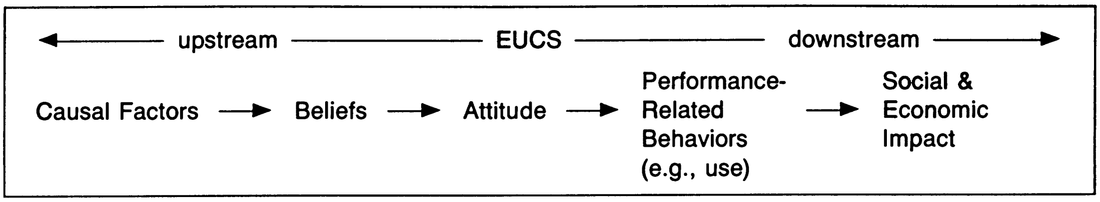

Figure 1. System to Value Chain

attitude research focuses primarily on affect to the detriment of understanding the cognitive and behavioral dimensions of attitudes.

One of the continuing dilemmas in IS research is the difficulty of specifying the performance-related behaviors that link user satisfaction with social and economic impacts. System usage has obvious limitations when use is mandatory. Melone (1990) recognizes this dilemma and calls for performance-related operationalizations (measures that consider the integrated context in which work is actually accomplished and the extent to which information is actually used) that would enhance the value of the system-use construct. However, these performance-related behaviors may be application-specific, making it difficult to develop measures that are generally applicable to a variety of applications. Until a richer variety of behavioral phenomena is identified and measured, predicting behavior (usage) by attitude (user satisfaction) is of limited theoretical interest.

The emphasis in social and cognitive psychology on emotional issues and measures that tap primarily the affective rather than the cognitive dimensions of attitudes also suggests the need for caution when borrowing concepts and techniques for use in the MIS arena. How much do we want evaluations of information systems to be based on affective (emotional) vs. cognitive responses from users? How much do MIS researchers want to venture in the world of feelings and emotions?

The attitude-behavior literature is complex and characterized by controversy (McGuire, 1985). There is little agreement about the definition of attitude or what aspects of attitude are worth measuring (Dawes and Smith, 1985). Moreover, research results have been disappointing. Simple models of the attitude-behavior relationship are yielding to more complex models of the conditions under which attitudes can predict behavior. Anyone going to this literature for an easy-to-read road map on how attitude-behavior theory can enhance research in MIS is likely to be disappointed.

Our purpose for developing a measure of end-user computing was to evaluate applications (e.g., to support upstream research). We didn't measure satisfaction to predict behavior (e.g., usage). We measured it to help us learn how to develop better applications and, thereby, realize social and economic benefits of investments in information technology.

This does not mean the end-user computing satisfaction instrument is not useful for predicting behavior. It may be quite useful. The problem in assessing its usefulness is that the performance-related behaviors that are critical to realizing the social and economic benefits of information technology have not been specified. Given the rather low or non-significant relations between attitude predictors and behavioral criteria, it is unlikely that research attempts to link satisfaction to behavior will be successful unless there is correspondence in target, action, context, and time between attitude and behavioral entities (Ajzen and Fishbein, 1977). It is difficult to develop an attitude measure that has correspondence with unspecified performance-related behaviors. The first step to tying end-user computing satisfaction more closely to attitude behavior theory is specifying a rich variety of performance-related behaviors.

Methodological Issues

Etezadi-Amoli and Farhoomand's questions concerning scaling and instrument development appear to be, in part, related to confusion concerning appropriate procedures for developing Likert-type scales to measure attitudes. An explicit objective of our research was to develop a Likert-type scale to measure end-user computing satisfaction (Doll and Torkzadeh, 1988). Irrespective of scaling technique, the resultant attitude score can represent an individual's location on an evaluative dimension vis-a-vis a given object. However, there are important differences between scaling procedures. In developing a Likert scale, attitude is a function of a respondent's set of beliefs about the application. A person's attitude score is obtained by summing across all his or her belief items (Fishbein and Ajzen, 1975 pp. 71-73).

In discussing item generation, Etezadi-Amoli and Farhoomand contend that the items in the original instrument are measures of beliefs rather than attitudes. This is true. The 38 items in the original questionnaire are belief items. Beliefs are defined as the probability dimension of a concept where the concept is a relational statement (Fishbein, 1963). In the end-user computing satisfaction instrument, the probability dimension of the beliefs is operationalized by the portion of the time (i.e., almost always to almost never) that the respondent feels the relational statement (e.g., Does the information content meet your needs?) is true. We suggest that this format is easy to respond to because it is experienced-based.

The expectancy-value model is useful for explaining how attitudes are measured using Likert-type scales. Fishbein (1963) maintains that an individual's attitude toward any object is a function of his or her beliefs about the object (i.e., the probability that the object is associated with other objects, concepts, values, or goals) and the evaluative aspect of those beliefs. This relationship can be expressed algebraically as follows:

$$A = \Sigma_i^N B_i \times E_i$$

Where:

- A is the resultant attitude score;
- B_i = belief "i" about the object;
- E_i = the evaluative aspect of B_i; and
- N = the number of beliefs.

In Likert scaling, each item is assumed to indicate either a favorable (+1) or unfavorable (−1) attitude, i.e., the E term has a value of +1 or −1. Attribute evaluations (Es) are not measured but are instead assumed to be the same for all subjects (Fishbein and Ajzen, 1975, p. 61). Items on a Likert scale do not reflect different degrees of favorableness, i.e., the extent of satisfaction (Fishbein and Ajzen, 1975, p. 80). Thus, Likert scales place greater weight on B than on E in computing attitude scores.

Several studies have suggested that attitudes might be estimated more accurately by considering both belief strength and evaluation of associated attributes (i.e., from sum of $B_i \times E_i$) rather than by using only the sum of beliefs (sum of B_i) or the sum of the evaluations (sum of E_i). However, this assumes that Es are a mixture of positive and negative items. When Es are either all positive or all negative, the sum of B_i alone tends to be highly correlated with the attitude, i.e., the Es are not necessary (Fishbein and Ajzen, 1975, p. 227). In our recommended instrument, each item measures a positive evaluative response.

Etezadi-Amoli and Farhoomand suggest including another scale evaluating the "degree of importance" of each item to the respondent. A "degree of importance" item was originally used by Bailey and Pearson (1983) in their user satisfaction instrument. In a later validation of Bailey and Pearson's instrument, Ives, et al. (1983) observed that "the weighted and unweighted scores are highly correlated, making the additional information provided by the importance rating unnecessary" (p. 787).

Fishbein and Ajzen (1975) do not recommend the use of "degree of importance" items. They contend that (1) the addition of an independent measure of importance merely provides redundant information, and (2) a measure of importance is not equivalent to evaluation. Although obtaining ratings of each attribute in terms of its importance has some intuitive plausibility, Fishbein and Ajzen argue that important items are typcially evaluated more positively or negatively than attributes that are unimportant. Also, people usually tend to have more information about things that are important to them, and thus they tend to be more certain and to have stronger beliefs about important than about unimportant attributes. Thus, studies have consistently found that including importance tends to attenuate the prediction of attitude (Fishbein and Ajzen, 1975).

Etezadi-Amoli and Farhoomand question the elimination of 15 items after the pilot study. They express concern that these elimination methods could have resulted in the deletion of some useful items. Pilot tests and the elimination of items is a standard procedure in the development of

Likert-type scales. In this case, there were several reasons for eliminating items. First, we were concerned that the items included be measures of satisfaction rather than measures of factors that cause satisfaction. In the early stages of research on new phenomena, cause and effect items are often grouped together to describe phenomena. Bailey and Pearson's (1983) work on the development of a user satisfaction instrument is a good example of this tendency. The items they used to measure user satisfaction included several factors, such as user involvement, top management involvement, documentation, relationship with EDP staff, and vendor support, that are often treated by others as variables that cause satisfaction. This concern was implicit in Treacy's (1985) call for a causal model of user satisfaction.

Second, the domain sampling model provides a rationale for excluding items. The key assumption in the domain sampling model is that all items, if they belong to the domain of the concept, have an equal amount of common core. If all the items in a measure are drawn from the domain of a single construct, responses to those items should be highly intercorrelated. The corrected item total correlation provides a measure of this (Churchill, 1979).

Third, Fishbein and Ajzen (1975)) describe the need to eliminate items in the construction of Likert scales. Items should indicate a favorable or unfavorable attitude toward the object in question. If the item is ambiguous or appears to indicate a neutral attitude, it should be immediately eliminated. Often the investigator makes this decision. In our research design, a measure of criterion-related validity (the two global items measuring perceived overall satisfaction and success of the application) was examined to identify items that did not indicate favorable or unfavorable attitudes.

Finally, Churchill (1979) describes the need to purify the measures before going beyond the pilot stage in research. Some researchers like to gather data and perform a factor analysis on the data before doing anything to purify the measures in the hope of determining the number of dimensions underlying the construct. Churchill contends that when factor analysis is done before purification, there seems to be a tendency to produce many more dimensions than can be conceptually identified. This effect is partly due to the "garbage items," which do not have the common core but do produce additional dimensions in the factor analysis. These "garbage items" confound the interpretation of the factor analysis.

Etezadi-Amoli and Farhoomand also question the elimination of five more items "because they represent the same aspect with only slightly different wordings." They ask why these five items were included in the instrument in the first place. The emphasis at the early stages of item generation is to develop a set of items that tap the construct. Churchill (1979) suggests that the researcher probably would want to include items with slightly different shades of meaning because the original list will be refined to produce the final measure. Experienced researchers can attest that seemingly identical statements produce widely different answers. By incorporating slightly different nuances of meaning in statements in the item pool, the researcher provides a better foundation for the eventual measure.

Etezadi-Amoli and Farhoomand contend that the procedures used to identify the underlying factors or components that comprise the domain of the end-user computing satisfaction construct and to eliminate items that were not factorially pure: (1) contradict the original assumption that there is only one common factor underlying the EUCS construct, and (2) result in the elimination of six potentially important variables (items). It was impossible to know a priori whether the end-user computing satisfaction construct was simple (one dimension) or complex (i.e., consisting of several components or dimensions). One of the goals of the research was to "identify underlying factors or components of end-user computing satisfaction" (Doll and Torkzadeh, 1988, p. 260). Given the lack of prior knowledge, it seemed inappropriate to assume that EUCS was a simple construct. On the second point, we have stated in our previous article that many of the items with multiple loadings "may be excellent measures of overall end-user computing satisfaction" (Doll and Torkzadeh, 1988, p. 266). Including multiple loading items in the scale "blurs the distinction between factors" and does not tap additional dimensions of the EUCs construct. Thus, including these six items would unnecessarily lengthen the questionnaire, making it more difficult to use.

Etezadi-Amoli and Farhoomand are quite correct in pointing out that there are relatively high correlations between the various items of the five factors. There is a great deal of common factor variance in all the items. Thus, the four content items have significant correlations with the two accuracy items even though their factor scores are orthogonal. When we suggested that the five components (content, format, accuracy, ease of use, and timeliness) are distinct, we meant that they meet the condition for discriminant validity (i.e., items of a component are more correlated with items of the same component than they are with items that measure other components).

We are not surprised that a confirmatory factor analysis using LISREL and the actual correlation matrix of the 12-item instrument did not indicate an adequate fit to the five-factor model. As we reported, there were only three factors with eigenvalues greater than 1. The five-factor solution was forced. A confirmatory factor analysis would be helpful. However, another large (600 +) multi-organizational sample should be gathered to confirm factor structure and discriminant validity. A large sample is necessary because the items could be expected to have considerable common variance and relatively large error variance compared to their unique variance.

Etezadi-Amoli and Farhoomand suggest that one should select at least three items for each factor. This appears to be a legitimate concern when one knows the factors ahead of time and is selecting items to measure each factor. In our case, we did factor analysis to find out how many factors were present.

Conclusions

Prior to the development of the end-user computing satisfaction instrument, the instruments used to measure user satisfaction (e.g., Bailey and Pearson 1983; Ives, et al., 1983) primarily measured affect through semantic differential scales (Melone, 1990). Several studies have reported problems with these instruments (Galletta and Lederer, 1989; Treacy, 1985). The end-user computing satisfaction instrument is different in that it emphasizes the cognitive or belief aspects of attitudes in a short, easy-to-use, application-specific instrument using Likert-type scales.

No instrument should be used without questioning the procedures used to develop it and the appropriateness of the measure for the research questions being examined. We appreciate this opportunity to respond to Etezadi-Amoli and Farhoomand's concerns. We hope our response has helped clarify theoretical and methodological issues related to the instrument.

We are gratified that the end-user computing satisfaction instrument is being used by so many researchers. We have found it to be a sensitive instrument that has enabled us to detect contingency relationships that might not have been discovered using a less-refined instrument (Doll and Torkzadeh, 1989). Our efforts to develop additional evidence related to the instrument's validity, internal consistency, and stability (Torkzadeh and Doll, 1991) are continuing.

References

Ajzen, I. and Fishbein, M. "Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research," *Psychological Bulletin* (84:5), September 1977, pp. 888-918.

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

Chaiken S. and Stangor, C. "Attitudes and Attitude Change," *Annual Review of Psychology* (38), 1987, pp. 575-630.

Churchill, G.A., Jr. "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* (16:1), February 1979, pp. 64-73.

Dawes, R.M. and Smith, T.L. "Attitude and Opinion Measurement," in *Handbook of Social Psychology* (3rd edition), G. Lindzey and E. Aronson (eds.), Random House, New York, NY, 1985, pp. 509-566.

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

Doll W.J. and Trokzadeh, G. "A Discrepancy Model of End-User Computing Involvement," *Management Science* (35:10), October 1989, pp. 1151-1171.

Fishbein, M. "An Investigation of the Relationships Between Beliefs About an Object and the Attitude Towards that Object," *Human Relations* (16:3), August 1963, pp. 233-240.

Fishbein, M. and Ajzen, I. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research,* Addison-Wesley Publishing, Reading, MA, 1975.

Galletta, D.F. and Lederer, A.L. "Some Cautions on the Measurement of User Information Satisfaction," *Decision Sciences* (20:3), Summer 1989, pp. 419-439.

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

McGuire, W.J. "Attitude and Attitude Change," in *Handbook of Social Psychology* (3rd edition), G. Lindzey and E. Aronson (eds.), Random House, New York, NY, 1985, pp. 233-346.

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

Ostrom, T.M. "The Emergence of Attitude Theory: 1930-1950," in *Psychological Foundations of Attitudes,* A.G. Greenwald, T.C. Brock, and T.M. Olson (eds.), Academic Press, New York, NY, 1968, pp. 1-32.

Torkzadeh, G. and Doll, W.J. "Test-Retest Reliability of the End-User Computing Satisfaction Instrument," *Decisions Sciences* (22:1), Winter 1991, forthcoming.

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

Zajonc, R.B. "On the Primacy of Affect," *American Psychologist* (39:2), February 1984, pp. 117-123.

William J. Doll
University of Toledo

Gholamreza Torkzadeh
University of Toledo