THE MEASUREMENT OF USER INFORMATION SATISFACTION

Blake Ives
Dartmouth College

Margrethe H. Olson
New York University

Jack Joseph Baroudi
New York University

The Measurement of User Information Satisfaction

BLAKE IVES *Dartmouth College*
MARGRETHE H. OLSON and JACK J. BAROUDI *New York University*

Blake Ives' current research interests include knowledge utilization, expert systems, the design of user interfaces, and the management of the computing resource.

Margrethe Olson is currently involved in the long-term evaluation of companies experimenting with pilot work-at-home programs, as part of a longstanding interest in the impact of office automation on the nature of work.

Jack Baroudi's research interests include the study of information systems personnel turnover, career paths, and management. He is also interested in the impacts of computer technology on individuals, organizations, and society.

Authors' Present Addresses: Blake Ives, Program in Computer and Information Science, Nathan Smith Building, Dartmouth College, Hanover, NH 03755. Margrethe H. Olson and Jack J. Baroudi, Computer Applications and Information Systems Area, Graduate School of Business Administration, 100 Trinity Place, New York, NY 10006.

INTRODUCTION

Information systems are expensive. The decision to install an information system necessitates a choice of mechanisms to determine whether an information system is needed, and once implemented, whether it is functioning properly. User information satisfaction (UIS) is one such evaluation mechanism. UIS is defined as the extent to which users believe the information system available to them meets their information requirements. UIS provides a meaningful "surrogate" for the critical but unmeasurable result of an information system, namely, changes in organizational effectiveness. In this paper past attempts to develop and validate a measure of UIS are reviewed. Instruments that measure UIS or components of UIS are compared, and one instrument with a strong empirical foundation [22] is subjected to additional psychometric examination. The results of a replication and extension of previous work with this instrument, based on a two-phase survey of production managers, are presented.

THE USER INFORMATION SATISFACTION CONSTRUCT

The concept of UIS can be traced to the work of Cyert and March [6, p. 126], who suggest that an information system which meets the needs of its user will reinforce satisfaction with that system. If the system does not provide the needed information, the user will become dissatisfied and look elsewhere. Although ideally one would like to evaluate the effectiveness of an information system based on its degree of use in decision making and the resultant productivity benefits, this "decision analysis" approach is generally not feasible [19]. Satisfaction of users with their information systems is a potentially measurable, and generally acceptable, surrogate for utility in decision making.

UIS is a perceptual or subjective measure of system success; it serves as a substitute for objective determinants of information system effectiveness which are frequently not available. Theoretically, the determination of information system value is a matter of economics: the costs of system operations and development are subtracted from the actual benefits (in improved organizational effectiveness) to obtain the net value of the system to the organization. In practice, however, this may not be a simple determination because (1) intangible costs and especially benefits of information systems are difficult to recognize and to convert to their monetary equivalent; (2) some

ABSTRACT: This paper critically reviews measures of user information satisfaction and selects one for replication and extension. A survey of production managers is used to provide additional support for the instrument, eliminate scales that are psychometrically unsound, and develop a standard short form for use when only an overall assessment of information satisfaction is required and survey time is limited.

decision support systems (e.g., a database supported by a user query facility) are used for disparate, relatively unstructured, ad hoc decisions; objectively assessing the benefits of such systems may be nearly impossible; (3) data on system success may be determinable but not recorded by the organization and, therefore, unavailable for research purposes [7].

System usage can be a surrogate indicator of system success under certain conditions. If users consider the system to be unreliable or its data inaccurate, their usage will reflect those doubts. If usage is voluntary, the system will be avoided. Since there are motivations for using the system other than its objective utility in decision making (e.g., mandate from management, political motivation, self-protection for justifying "poor" decisions), either or both objective and perceptual measures may be appropriate depending on the situation. Also, UIS can be more than a substitute for an objective measure of system success. UIS measures how users view their information system rather than the technical quality of the system. A "good" information system perceived by its users as a "poor" system is a poor system.

User information satisfaction has been used in a number of research efforts as a surrogate measure of system effectiveness [2, 4, 8, 16, 17, 22, 24]. Each investigation is based on a measure which is unique to that particular study. In some studies, other terms have been used to refer to essentially the same concept: "felt need" [9], "system acceptance" [11], "perceived usefulness" [15], "feelings about the information system" [18], and "MIS appreciation" [23] are a few examples.

User involvement in information system development has frequently been hypothesized to be related to UIS although the results to date have been mixed (see [12] for a review of this research). UIS has also been analyzed in relation to system usage [2, 16, 18, 23] and estimates of system value [8, 16].

MEASURING USER INFORMATION SATISFACTION

The construct of UIS has been operationalized in many different ways. Several studies employed single-item rating scales [2, 17]; such scales have been criticized as unreliable [15, 20]. Single-item scales also provide little information as to what the user finds dissatisfying (or satisfying) and are thus of limited utility outside a research setting. Multiple-item UIS measures have become increasingly common. Generally, they are of two types. The first focuses on the information system product. With such diverse names as "system acceptance" [11], "output quality" [16], and "MIS appreciation" [23], these scales focus on the content of the information system (e.g., accuracy, relevance) and the manner in which the information is presented (e.g., format, mode). The second type of multiple-item scale includes the organizational support for developing and maintaining the system as well as the system product itself. This type of instrument contains items concerned with training, documentation, development procedures, systems maintenance, etc., as well as items related to system content. Thus it provides an indicator of the overall quality of information services provided by an information service function.

Generally, UIS measures have not been carefully validated [15]. Recently, however, several rigorous attempts have been made to develop valid and reliable UIS measures. These efforts are discussed below.

Gallagher

Gallagher's [8] questionnaire focused on user perceptions of the information value of reports provided by an information system. The questionnaire had two types of questions: requests for managers to estimate the dollar value of a report,

and semantic differential adjectives on which the managers rated the reports.

The questionnaire results were based on responses from 75 managers utilizing the same information system in a single company. Gallagher concluded from his results that both the estimated dollar value and the semantic differential measures had potential for analyzing information value. However, the correlation between the two measures was too low to conclude that they were measuring the same phenomenon.

Several other problems exist with Gallagher's measures. Both measures focus only on the product (in this case, a report) and not on the quality of service provided by the information services function. Moreover, the scales could not be easily generalized to other information system products. The dollar estimates have no anchor point and there was no attempt to validate the relationships between the estimated and real dollar value of a report. Although Gallagher interviewed respondents to verify their beliefs in their estimates, the standard deviation was extremely high and the distribution quite skewed. Moreover, 30 percent of the respondents did not respond to the dollar value question, citing lack of familiarity with information system costs as the reason. Finally, no validation of the semantic differential scales was reported.

Jenkins and Ricketts

Jenkins and Ricketts developed a 20-item measure of "user satisfaction" on the basis of "a survey of existing literature and structured interviews with leading researchers in the field" [13, p. 2]. Eighteen of the 20 items were chosen as representative of each of 5 factors defined a priori as constituting user satisfaction (i.e., input procedures, systems processing, report content, report form, report value). The other two items were overall measures of UIS. Each item consisted of a 7-point, semantic differential scale anchored at each end by bipolar adjectives (for example, 1 (very untimely), 2, 3, 4, 5, 6, 7 (very timely)).

The instrument was psychometrically tested in five laboratory experiments involving 197 participants. Analysis showed each item to be normally distributed and demonstrated an acceptable overall interitem reliability (i.e., coefficient alpha) of 0.85. The scores from 5 factors derived by factor analysis were used in a regression equation with one of the two global satisfaction measures serving as a criterion variable. These factors significantly accounted for approximately 30 percent of the variance in global satisfaction scores. The factor analysis, however, failed to substantiate the factor structure originally proposed.

The work of Jenkins and Ricketts has several limitations. The procedure used to generate the original items is not rigorously described. The instrument was designed to focus on the information system product; like Gallagher's scale, it does not cover information systems service. The factor structure originally proposed did not hold up in factor analysis. Finally, the two global measures of information satisfaction were included in the same instrument with the items to be validated against them, suggesting possible method variance [3].

Jenkins and Ricketts have extensively redesigned their questionnaire based on subsequent investigations. No data has yet been reported on this new instrument.

Larcker and Lessig

Larcker and Lessig [15] developed two 3-item scales that together constitute "perceived usefulness." The first scale measures "perceived importance," an indicator of "whether the information is relevant, informative, meaningful, important.

helpful. or significant [15, p. 123]. The "perceived usableness" scale indicates "whether the information format is unambiguous, clear, or readable" [15, p. 123].

Items were initially de   d by faculty and students who proposed characteristics or information associated with "importance" and "usableness"; these dimensions had been selected as "two aspects that seem to be common to prior measurement instruments" [15, p. 123]. The list of suggested characteristics was reduced to six items by another panel of faculty and graduate students. The items were experimentally tested in a study of decision making involving 29 faculty and graduate students. Factor analysis of the six items verified the independence of the two scales.

Larcker and Lessig analyzed the convergent (between measures) and discriminant (across settings) validity of the two dimensions using the multitrait–multimethod procedure of Campbell and Fisk [3]. They found acceptable interitem correlations within each evaluation setting and acceptable differences between correlations across settings and concluded that convergent and discriminant validity were established. The reported reliabilities (Cronbach's alpha) for the two dimensions ranged between 0.64 and 0.77.

Larcker and Lessig's measures have several critical weaknesses. The original two dimensions, importance and usableness, are not empirically derived and, as the authors note, may be "ignoring additional dimensions of perceived usefulness, such as information accuracy or timeliness" [15, footnote. p. 123]. Like the other two measures, the instrument relates specifically to the information system product and not to factors related to the quality of service.

The reliabilities reported for the two scales are relatively low for applied research [20, p. 226], although Larcker and Lessig point out that they are acceptable for exploratory work. The instrument was developed and the study conducted in an artificial setting involving faculty and graduate students using a capital-budgeting decision. The validity of generalizing the measures to more realistic settings and other problem types is unproven. Finally, their application of the multitrait–multimethod procedure to establish validity may be questioned; the authors interpret different measures of the same construct to be different measurement methods. They also interpret the different evaluation settings (variations on the capital-budgeting decision) to represent traits.

Pearson

Pearson [1, 22] developed a list of "factors" that contribute to information satisfaction. A list was derived from the existing research on computer-user interactions and was then reviewed for completeness and accuracy by three data processing professionals. Next it was compared to an analysis of critical incidents collected in interviews with 32 user managers. As a result, 39 distinct factors were identified, which were the basis for an instrument which utilized the semantic differential technique. Four adjective pairs were provided for each factor, plus a "satisfied–dissatisfied" pair and an impor-

TABLE I. An Example Information Satisfaction Factor (from Pearson [22])

19 Reliability of Output Information		
Consistent	:-:-:-:-:-:-:	Inconsistent
High	:-:-:-:-:-:-:	Low
Superior	:-:-:-:-:-:-:	Inferior
Sufficient	:-:-:-:-:-:-:	Insufficient
Satisfied	:-:-:-:-:-:-:	Dissatisfied
Important	:-:-:-:-:-:-:	Unimportant

Note: The first four adjective pairs are the actual items that make up the factor. The satisfied/dissatisfied and important/unimportant pairs are used for validating the factor.

tance rating. The resulting instrument was completed by the same managers who had previously been interviewed.

Pearson originally proposed a scoring method which used the "importance" rating as a weighting factor when calculating the overall satisfaction score; a description of this scoring method is found in [1]. In Pearson's sample, the weighted and unweighted scores were highly correlated, making the additional information provided by the importance rating unnecessary. A sample factor, "reliability of output information," and its associated items are shown in Table I.

Reliability, calculated for each factor based on the variance in responses between the four adjective pairs, was found to be at an acceptable level. Content validity was claimed based on the method used to develop the instrument. Predictive validity was established by comparing the total score on the instrument with an overall satisfaction rating which had been collected during the interview: the correlation was 0.79. The scores on each factor were also compared with the "satisfied–dissatisfied" adjective pair; although the independence of these measures was subject to question, the correlations were all very high. Finally, construct validity was established by examining the responses for each factor on the importance scale and comparing them with the rankings of importance obtained earlier, and by correlating the individual factors against the total score; the correlations (Spearman) were at acceptable levels.

There are several problems with Pearson's procedures. The sample on which the instrument was tested was relatively small (29 user managers) and may have been biased by their prior participation in the development of the instrument. As Pearson noted, the construction of the instrument did not assure independence of responses. This may have unduly affected the reliability scores and overemphasized the claims for construct validity.

Comparison of Measures

Table II contains a summary of the four measures reviewed on the following criteria: derivation (empirical or otherwise). amount of empirical support, level of coverage (product, system services), and number of indicators in the measure. In order to choose the most appropriate measure for further

TABLE II. Evaluation of Alternative UIS Measures

Measure	Derived From	Empirical Support	Level of Coverage	Number of Indicators
Gallagher	Empirical	Adequate	Product	18
Jenkins and Ricketts	Literature and interviews	Inadequate	Product	5
Larcker and Lessig	Interviews	Adequate	Product	2
Pearson	Literature, interviews, and empirical	Adequate	Product and support	39

study. it was assumed that an empirically derived measure. with adequate empirical support. which covers both the information system product and general system services and provides multiple indicators, would be favored. On this basis, Pearson's measure was chosen for further investigation as to its potential to become a standard measure of UIS.

A STUDY TO ASSESS A UIS MEASURE

Although Pearson's study represents an important first step toward the development of a valid UIS measure. further investigation is required to assess the validity and reliability of Pearson's measure and to refine it for use in research and practice. The authors chose to undertake such an investigation and the results are reported here. The explicit goals of this investigation were to

(1) replicate Pearson's findings concerning the validity of the instrument:
(2) reinforce the validity of the instrument through further tests:
(3) reduce the length of the overall measure while maintaining reasonable levels of reliability and the existing structure of scales:
(4) develop a standard "short-form" of the instrument for research requiring only a global indicator of user information satisfaction.

The measure was administered as part of a survey of 800 production managers in U.S. manufacturing organizations. The participants were selected from a commercially obtained mailing list of production managers: their participation was solicited by mail without their prior knowledge of the study or its purpose.

Two separate mailings were made over a 2-month period. In one mailing the managers received a copy of Pearson's UIS measure. The second mailing included a separate 4-item measure of information satisfaction (UIS4). which is discussed in [21] and reproduced in the appendix. Other questionnaires measuring constructs hypothesized to relate to UIS were also included in each mailing. The ordering of the two mailings was reversed for one-half of the sample.

A total of 280 managers completed the UIS measure. a response rate of 35 percent. The number completing both sets of questionnaires was 200 (25 percent). Although the response rates are somewhat low. this was expected for the following reasons:

(1) The mailing list was about a year old. so that a number of managers on the list had changed jobs or companies.
(2) The survey was mailed, unsolicited, without prior knowledge on the part of the participants.
(3) Managers were requested to complete several relatively lengthy forms on two separate occasions.

RESULTS

In the following discussion. each of the variables related to UIS will be referred to as *scales* rather than factors (the term used in Pearson's work). The 39 scales are listed in the appendix. The four semantic differential measures which make up each scale are referred to as items.

Reliability

The reliability of a measure refers to its stability over a variety of conditions [20, p. 191]. Generally. reliability is of two types: test–retest, and amount of error in the measurement. Neither Pearson nor the current study attempted to assess the test–retest reliability of the instrument. The amount of error

in a measure is determined by Cronbach's alpha test applied to interitem scores and to the overall measure. This study's interitem reliability scores for the individual scales ranged between 0.82 and 0.97 with 30 being greater than 0.90. An overall measure of UIS (the sum of the 39 scales composed of 156 items) results in an overall questionnaire reliability of 0.97. Pearson did not provide an overall measure of reliability but reported individual scale reliabilities of between 0.75 and 0.98 with 32 having reliabilities greater than 0.90. They are shown in column 2 of Table IV.

The reliability scores are acceptable at the 0.80 level required for basic research. However. the four items composing each scale were located together on the instrument and were scored in the same direction (i.e., positive on the right, negative on the left); a tendency for respondents to simply mark straight down a column for the four items composing a particular scale may have artificially inflated the reliability data.

Content Validity

The content validity of a questionnaire refers to the "representativeness" or sampling adequacy of the content. In general. it is the manner by which the questionnaire and its items are built that ensures the reasonableness of the claims of content validity [16, 20]. The careful and systematic way that Pearson generated his scales lends credence to his claims of content validity.

Two types of circumstantial evidence for content validity can be generated statistically: internal consistency and the correlation of scales with other measures of the construct. In his research Pearson found all the interitem correlations to be positive and all but one significant at the 0.05 level: this is indicative of internal consistency. In this study all the interitem correlations were found to be positive and significant at the 0.001 level.

As already noted. the authors obtained an independent measure of user information satisfaction at a different time from Pearson's measure. To further test content validity. each of the 39 scales was correlated against this measure for the 200 subjects who responded on both measures. The correlations ranged from 0.22 to 0.54 with 21 scales correlating at levels above 0.40. All correlations were significant at the 0.001 level. They are shown in column 3 (UIS4) of Table IV.

The correlations. though significant. indicate only limited evidence for content validity by themselves. However. when the measures of internal consistency and the carefulness with which the questionnaire was constructed are considered. strong supportive evidence for content validity exists.

Predictive Validity

Predictive validity is demonstrated by correlating a measure against other measures of the same construct. Pearson correlated his overall satisfaction score with a self-report satisfaction measure provided by the user during an interview. The correlation between these two measures was 0.79. significant at the 0.001 level. In this study the overall score from the separate 4-item measure (UIS4. shown in the appendix) was used to analyze predictive validity by correlating it with the overall score obtained from the Pearson questionnaire. A correlation of 0.55 was obtained (significant at the 0.001 level). The results obtained by Pearson and this study are consistent and thus indicate predictive validity for the questionnaire.

Construct Validity

The construct validity of a measure is demonstrated by validating the theory behind the instrument [14]. A final claim of construct validity cannot be made until the questionnaire and

TABLE III Results of Factor Analysis

Scales	EDP Staff and Services	Factor Loadings Information Product	Vendor Support	Information Product	Knowledge or Involvement
Relationship with EDP staff	0.62				
Processing of requests for system changes	0.69				
Confidence in system		0.61			
Timeliness of output		(0.47)		0.66	
Vendor support			0.61		
Training provided users					0.57
User's understanding of systems					0.71
User's participation					0.55
Currency of output		0.54		0.50	
Attitude of EDP staff	0.73				
Reliability of output		0.74			
Allocation priorities for EDP resources	0.55				
Convenience of access	0.55				
Relevancy of output		0.74			
Volume of output		0.56			
Accuracy of output		0.80			
Precision of output		0.80			
Communication with EDP staff	0.71				
Time required for systems development	0.65				
Personal control of EDP services	0.56				
Completeness of output		0.70			
Flexibility of systems	0.56				

theory have been subjected to several alternative forms of testing with consistent findings. The extent to which most of Pearson's data have been replicated by this study suggests construct validity of the measure.

Two methods of construct validation cited by Kerlinger [14] are examination of the correlations between total scores and item scores and factor analysis. The first approach assumes the total score to be valid; thus the extent to which the item correlates with the total score is indicative of construct validity for the item. In this study each scale score was subtracted from the total score in order to avoid a spurious part–whole correlation [5]; the result is a "new total score" (UIS38) which was then correlated with the scale score. The correlations between the "new total score" and the scales ranged from 0.42 to 0.83 with 34 scales correlating with the new total score at levels above 0.60. All 39 correlations are significant at the 0.001 level. They are shown in column 4 (UIS38) of Table IV. If it can be assumed that the total score does measure user satisfaction, these results support construct validity.

The second method of construct validation, factor analysis, is considered one of the most powerful methods of construct validation as it allows the examination of the underlying structure of the overall measure [14]. Pearson performed a factor analysis and discovered 8 dimensions, but his sample size to item ratio was so small as to be generally unacceptable for statistical use. In this study a factor analysis was also conducted. While the ratio of sample size to number of scales in the study (7 : 1) must also be regarded with some caution, it is substantially better than Pearson's 1 : 1 ratio. Employing a cutoff level of 0.50, a 5-factor structure resulted with 22 scales loading at that level.

Those scales loading at the 0.50 level in the factor analysis are shown in Table III. The first factor carries with it scales related to the information services function. The second factor draws from scales related to the information system product. The third factor includes only one scale, vendor support. Vendor support was judged to be one of the least important scales by Pearson's subjects; the authors also questioned the inclu-

sion of this scale in the questionnaire as not all facilities have information products which require vendor support (or support which is visible to the user). This scale was considered for elimination, as discussed in the next section. The fourth factor contained two scales related to the information system product. Close examination showed these two scales both loaded heavily with factor two, suggesting they are part of that structure rather than a separate dimension. The fifth factor contained those scales related to knowledge or involvement in the design of the information system. While no a priori loadings were hypothesized, the factor analysis indicates that a logical structure of scales does exist.

IMPROVING INSTRUMENT QUALITY

Two approaches were taken to improve the quality of the original Pearson measure and, as a side benefit, to reduce its length. The first approach involved permanently eliminating scales which showed undesirable psychometric qualities. Results reported on the basis of such scales can be spurious. The second approach was to eliminate items within scales in order to reduce the time to complete the instrument without losing any of its positive psychometric qualities. If time is a major consideration, a "short-form" of the instrument, which provides only an overall measure of the construct, may be desirable. In the next section the development of a short form of this UIS measure is discussed.

Elimination of Scales

As no well-established minimum value exists for validity correlations and since all reliabilities were at least at the 0.80 level, considered acceptable for research purposes [20], scales could not be eliminated based on any one criterion. Therefore the elimination of scales was based on careful empirical analysis and judgment. Each scale was ranked on the following criteria:

(1) reliability (column 2 in Table IV);
(2) content validity, i.e., correlation with independent UIS measure (UIS4, column 3 in Table IV);

TABLE IV. Evidence   :limination of Scales

Scales	Reli~~	Correlations with		Factor Loaded
		UIS4[a]	UIS38[a]	
Relationship with EDP staff	0.90	0.54	0.61	Yes
Processing of requests for system changes	0.84	0.46	0.62	Yes
Means of input/output with EDP center	0.87	0.48	0.66	No
Competition with EDP unit[***]	0.82	0.32	0.59	No
Confidence in systems	0.97	0.45	0.73	Yes
Timeliness of output	0.91	0.37	0.71	Yes
Chargeback method[***]	0.85	0.33	0.57	No
Perceived utility	0.92	0.42	0.67	No
Vendor support[***]	0.91	0.22	0.42	[b]
Computer language used[***]	0.87	0.21	0.46	No
Expectations of computer support	0.90	0.48	0.71	Yes
Correction of errors	0.87	0.36	0.79	Yes
Security of data[***]	0.94	0.22	0.42	No
Training provided users	0.97	0.32	0.61	Yes
User's understanding of systems	0.92	0.30	0.63	Yes
User's participation	0.94	0.32	0.67	Yes
Currency of output	0.95	0.41	0.73	Yes
Attitude of EDP staff	0.92	0.50	0.77	Yes
Reliability of output	0.95	0.43	0.83	Yes
Top management involvement	0.92	0.29	0.70	No
Format of output[***]	0.90	0.37	0.63	No
Response/turnaround time	0.94	0.40	0.75	No
Allocation priorities for EDP resources	0.92	0.36	0.67	Yes
Convenience of access	0.96	0.36	0.69	Yes
Relevancy of output	0.94	0.41	0.77	Yes
Volume of output	0.86	0.39	0.69	Yes
Job effects of computer support	0.91	0.48	0.74	No
Accuracy of output	0.94	0.40	0.73	Yes
Precision of output	0.96	0.41	0.72	Yes
Communication with EDP staff	0.94	0.47	0.80	Yes
EDP organizational position	0.90	0.44	0.66	No
Time required for systems development	0.91	0.47	0.74	Yes
Personal control of EDP systems	0.94	0.37	0.68	Yes
Scheduling of EDP products and services	0.96	0.37	0.66	No
Documentation	0.94	0.36	0.66	Yes
Completeness of output	0.95	0.41	0.77	Yes
Technical competence	0.95	0.43	0.73	No
Flexibility of system	0.95	0.47	0.77	Yes
Integration of database	0.97	0.46	0.72	No

[a] All correlations significant at $p = 0.001$.
[b] Single scale loading on factor.
[***] Selected for elimination.

(3) construct validity. i.e., correlation with overall measure minus the score for that scale (UIS38. column 4 in Table IV).

Next, the lowest ten values in each category were examined under the assumption that a low ranking in a category indicated only weak signs of the desired property. In the case of construct validity the scale had to possess both poor ranking and either not load or load separately in the factor analysis. These rankings were then compared and any scale which was found to be low in two of the three desired properties was eliminated. Table IV shows the results for each of the scales on the three criteria.

In the table. scales marked with a triple asterisk were selected for elimination. The rationale for eliminating scales 4, 7, 9, and 10 is further supported by the responses Pearson received as to the importance of each scale. His subjects ranked scales 4, 7, and 10 as being the least important items for user information satisfaction, while item 9 was ranked twenty-seventh in importance. Scale 13 (security) was ranked

by Pearson's subjects as moderately important but ranked so low on the other criteria that its elimination was easily justified. None of the items recommended for elimination were ranked by Pearson's subjects as being extremely important.

Reducing the Number of Items per Scale
Presently the instrument requires 20 to 30 minutes for a respondent to complete, with four items for each scale. It was felt that it would be advantageous to reduce the number of items per scale if reliability were not affected. To identify candidate items for elimination without biasing evidence of reliability for the new measure, a 100-person "holdback" sample was removed from the original group of respondents. The remainder of the sample ($n = 100$) were used to determine which items could be safely dropped. Finally, the holdback sample data was tested to determine both the reliability and validity of the new measure.

It was recognized that it would be possible to improve internal consistency and reliability, or at least to minimize the effects of reducing the length of the instrument, by removing

TABLE V. Comparison of Reliability Data for 4 Items Versus 2 Items per Scale

Scales	4 Items*	2 Items*
Relationship with EDP staff	0.90	0.94
Processing of requests for system changes	0.85	0.90
Means of input/output with EDP center	0.90	0.83
Competition with EDP unit	0.86	0.87
Confidence in systems	0.97	0.97
Timeliness of output	0.93	0.94
Chargeback method	0.84	0.96
Perceived utility	0.93	0.94
Vendor support	0.91	0.93
Computer language used	0.88	0.95
Expectations of computer support	0.92	0.91
Correction of errors	0.86	0.84
Security of data	0.94	0.95
Training provided users	0.97	0.97
User's understanding of systems	0.93	0.92
User's participation	0.96	0.92
Currency of output	0.95	0.95
Attitude of EDP staff	0.92	0.88
Reliability of output	0.96	0.95
Top management involvement	0.92	0.88
Format of output	0.90	0.84
Response/turnaround time	0.94	0.96
Allocation priorities for EDP resources	0.95	0.93
Convenience of access	0.96	0.97
Relevance of output	0.94	0.95
Volume of output	0.89	0.92
Job effects of computer support	0.90	0.88
Accuracy of output	0.96	0.95
Precision of output	0.95	0.94
Communication with EDP staff	0.94	0.90
EDP organizational position	0.89	0.81
Time required for systems development	0.91	0.90
Personal control of EDP systems	0.93	0.84
Scheduling of EDP products and services	0.98	0.96
Documentation	0.92	0.90
Completeness of output	0.93	0.96
Technical competence	0.96	0.97
Flexibility of system	0.96	0.92
Integration of database	0.97	0.95

* n = 100.

those items within a scale that had the lowest correlations with the other items. Interitem correlations for each scale of the nonholdback sample were examined to determine which items were candidates for elimination. Items were eliminated from the scale on the basis of their low correlations with the two items which remained. A version of the questionnaire containing two items per scale was then tested on the holdback sample. Table V displays reliability data for the proposed measure; Table VI contains validity data. For every scale, the reliability and validity data for the two-item measures were considered adequate. On the basis of interitem correlations, two items were then eliminated from each scale.

PRODUCING A SHORT FORM
Although the improvements discussed above reduced the length of the instrument, it is still rather long. It was felt that it would be worthwhile to develop a "short form" of the instrument for use when time is limited and/or only an overall assessment of UIS (rather than analysis of particular symptoms of a problem) is adequate. Development of such a form entails selecting representative scales such that the overall score on the short form has an acceptably high correlation with the full instrument.

The procedure employed to develop the short form was as follows. First, the scales that had demonstrated undesirable psychometric characteristics [5, 8, 10, 11, 14, 22] were not considered. Second, only those scales with factor loadings of 0.50 or better, as shown in Table III, were included. Third, each remaining scale utilized only the two items derived from the analysis shown in Table VI. The scales retained for the short form are indicated in the appendix.

To determine whether the resulting short form adequately measured Pearson's original concept of UIS, the scales on the short form were removed from the original Pearson measure and the short form total correlated with the remaining scales. The correlation was 0.90 (significant at $p = 0.001$). The short form total was also correlated with the independent four-item measure of UIS (UIS4); a correlation of 0.54 (significant at $p = 0.001$) was found. These correlations provide substantial evidence that the short form questionnaire is a sound general measure of Pearson's original UIS concept. A copy of the short form questionnaire is available from the authors.

DISCUSSION
The authors feel that the instrument developed by Pearson, with the extensions and improvements presented here, represent substantial progress toward establishment of a standard measurement technique for UIS. However, continuing efforts should be made to validate, extend, and disseminate the instrument. In this section some suggestions are made for application and improvement of the instrument in practice and in further research.

Application in Practice
The improved long form may be utilized to evaluate an information system or general systems development effort in an organization. It provides not only an overall assessment but also the capability to analyze (based on individual scales) what aspects of the systems effort are most problematical. The general form has the advantage that results can be compared across systems or even across organizations. On the other hand, it may be more advantageous in certain situations to customize the instrument to a certain system and/or systems function. Although the ability to generalize is lessened, customization has the advantage of clarity to the users. Another possible approach is to utilize the short form for an overall measure augmented by items of particular interest from the original and/or specific items tailored to the situation by the instrument user.

Application to Research
The primary significance of this work to the research community is the establishment of a generalized, standardized instrument for use across studies. However, the instrument could be significantly improved by further testing. The following are suggested:

(1) Establish test–retest reliability.
(2) Change the instrument format to eliminate biased responses and retest the interitem reliability. In its present form, the adjective pairs for each scale are grouped together immediately following the scale description; also all items are scaled in the same direction from positive to negative. It would appear that mixing up the scale responses and changing the direction of some of the adjective pairs would eliminate bias and lessen the potential for interitem reliability scores to be artificially inflated. The cost is that for each item the scale description would have to be repeated, thus increasing significantly the overall length of the instrument.

TABLE VI. Comparison of Validity Data for 4 Items Versus 2 Items per Scale (correlations with independ. iS measure, overall score)

Scales	4 Items with UIS4	2 Items with UIS4		4 Items with UIS38	2 Items with UIS38
Relationship with EDP staff	0.51	0 48		0.05	0.63
Processing of requests for system changes	0.43	0.40		0.69	0.66
Means of input/output with EDP center	0.52	0.48		0.70	0.68
Competition with EDP unit			Dropped from measure		
Confidence in systems	0.43	0.43		0.68	0.68
Timeliness of output	0.37	0.38		0.68	0.71
Chargeback method			Dropped from measure		
Perceived utility	0.41	0.39		0.73	0.69
Vendor support			Dropped from measure		
Computer language used			Dropped from measure		
Expectations of computer support	0.52	0.48		0.84	0.80
Correction of errors	0.37	0.38		0.78	0.74
Security of data			Dropped from measure		
Training provided users	0 32	0.29		0.61	0.56
User's understanding of systems	0.29	0.28		0.62	0.58
User's participation	0.38	0.36		0.77	0.79
Currency of output	0.39	0.43		0.76	0.74
Attitude of EDP staff	0.52	0.51		0.74	0.73
Reliability of output	0.41	0.38		0.80	0.80
Top management involvement	0.17[a]	0.20[b]		0.62	0.62
Format of output			Dropped from measure		
Response/turnaround time	0.42	0.39		0.77	0.72
Allocation priorities for EDP resources	0.37	0.36		0.62	0.63
Convenience of access	0.31	0.29		0.59	0.55
Relevancy of output	0.39	0.35		0.72	0.68
Volume of output	0.38	0.33		0.67	0.60
Job effects of computer support	0.43	0.43		0.72	0.73
Accuracy of output	0.36	0.30		0.72	0.69
Precision of output	0.38	0.40		0.69	0.69
Communication with EDP staff	0.48	0.43		0.83	0.78
EDP organizational position	0.44	0.46		0.73	0.73
Time required for systems development	0.44	0.40		0.68	0.67
Personal control of EDP systems	0.35	0.34		0.70	0.66
Scheduling of EDP products and services	0.35	0.35		0.70	0.68
Documentation	0.37	0.39		0.62	0.58
Completeness of output	0.49	0.45		0.74	0.74
Technical competence	0.40	0.42		0.48	0.71
Flexibility of system	0.50	0.49		0.75	0.74
Integration of database	0.39	0.40		0.71	0.69

[a] $p = 0.094$.
[b] $p = 0.041$.
Note: All other correlations significant at $p = 0.001$.

(3) Test the instrument in other samples besides production managers, representing different organizational functions and different management levels.

(4) Test the validity of the short form. The short form presented here is based on the data from the original sample. Although its correlation with the remaining items (long form minus short form items) is encouraging, further validation tests in other settings are required.

(5) Perform further construct validation. Testing the instrument against objective measures of system success, where possible, will greatly contribute to claims for its construct validity. Further understanding can be gained by comparing the results with objective measures of system usage in cases where usage is voluntary.

Development of a Central Data Bank
The real value of establishing a generalized instrument for UIS cannot be realized unless a mechanism is established for provision of a centralized data bank of results. Such a data bank would permit comparison of results across organizations and across other variables of interest. Given a larger sample, further analysis could be performed, for instance, by functional area or by management level. The authors call on fellow researchers in information systems to encourage the development of such a data bank.

CONCLUSION
This article presents significant progress toward development of a standard measure of user information satisfaction. Whether or not this instrument is chosen, the authors encourage the MIS research community to choose a standard instrument for measuring UIS.

The authors also encourage efforts to disseminate an instrument for evaluating UIS to the practicing information systems community and to encourage its use. The lack of adequate mechanisms to evaluate information systems effectiveness has long been apparent; any further efforts to develop or validate a UIS instrument should be oriented to the goal of aiding the practitioner. The instrument presented here is a potential candidate for adoption by both researchers and practitioners in information systems.

APPENDIX

I. Scales composing the original UIS instrument (Pearson)[1,2]:

1. Relationship with the EDP staff[+*]
2. Processing of requests for changes to existing systems[+*]
3. Means of input/output with the EDP center[*]
4. Interdepartmental competition with the EDP unit
5. Confidence in systems[*]
6. Timeliness of output information[*]
7. Chargeback method of payment for services
8. Perceived utility (worth versus cost)[*]
9. Vendor support of hardware and software
10. Computer language used to interact with systems
11. Expectation (expected versus actual level of computer-based support)[*]
12. Correction of errors[*]
13. Security of data
14. Degree of EDP training provided to users[+*]
15. Users' understanding of systems[+*]
16. Users' feelings of participation[+*]
17. Currency (up-to-dateness) of the output information[*]
18. Attitude of the EDP staff[+*]
19. Reliability of output information[+*]
20. Top management involvement in EDP activities[*]
21. Format of output
22. Response/turnaround time[*]
23. Determination of priorities for allocation of EDP resources[*]
24. Convenience of access (to utilize the computer capability)[*]
25. Relevancy of output information (to intended function)[+*]
26. Volume of output information[*]
27. Personal job effects resulting from the computer-based support[*]
28. Accuracy of output information[+*]
29. Precision of output information[+*]
30. Communication with the EDP staff[+*]
31. Organizational position of the EDP function[*]
32. Time required for new systems development[+*]
33. Personal control of EDP service received[*]
34. Schedule of recurring output products and services[*]
35. Documentation[*]
36. Completeness of the output information[+*]
37. Technical competence of the EDP staff[*]
38. Flexibility of systems[*]
39. Integration (automated sharing of information) of system database[*]

[1] A superscript plus sign indicates that the scale was retained in the improved instrument.

[2] An asterisk indicates that the scale was retained in the improved instrument.

II. Four-item General UIS Measure

How adequately do you feel the data processing group meets the information processing needs of your area of responsibility?

Very well   Adequately   Marginally   Poorly

How adequately do you feel the data processing group meets the needs of the broader class of users they serve?

Very well   Adequately   Marginally   Poorly

Data processing support can be judged on two criteria: efficiency and effectiveness. *Efficiency* deals with how well they do what they do. Are reports on time? Are projects developed within budget? *Effectiveness* takes a broader focus. Are they

doing the right things? Are critical "life-blood" applicati being developed? Are new computer technologies being __c-cessfully integrated into the organization?

How *efficient* do you feel the data processing group is?
Very efficient   Fairly efficient
Somewhat inefficient   Very inefficient

How *effective* do you feel the data processing group is?
Very effective   Fairly effective
Somewhat ineffective   Very ineffective

REFERENCES

1. Bailey, J.E., and Pearson, S.W. Development of a tool for measuring and analyzing computer user satisfaction. *Manage. Sci. 29*, 6 (May 1983), 519–529.
2. Barrett, G.V., Thornton, C.L., and Cabe, P.A. Human factors evaluation of a computer-based information storage and retrieval system. *Human Factors 10* (1968).
3. Campbell, D.T., and Fiske, D.W. Convergent and discriminant validation by the multitrait, multimethod matrix. *Psychol. Bull. 56,* (1959).
4. Cheney, P. Organization characteristics and information systems: An investigation. Ph.D. dissertation. University of Minnesota, Minneapolis, 1977.
5. Cohen, J., and Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Lawrence Erlbaum Assoc., Hillsdale, N.J., 1975.
6. Cyert, R.M., and March, J.G. *A Behavioral Theory of the Firm.* Prentice-Hall, Englewood Cliffs, N.J., 1963.
7. Edstrom, A. User influence and the success of MIS projects. *Human Relations 30,* 1977.
8. Gallagher, C.A. Perceptions of the value of a management information system. *Acad. Manage. J. 17,* 1 (1974).
9. Guthrie, A. Attitudes of user-managers towards MIS. *Manage. Informatics 3,* 5 (1974).
10. Hackman, J.R. Behavioral processes in organizations. In M.D. Dunnette (Ed.), *Handbook of Industrial and Organizational Psychology.* Rand-McNally, Chicago, 1976.
11. Igersheim, R.H. Management response to an information system. *AFIPS Conference Proceedings,* National Computer Conference, 1976.
12. Ives, B., and Olson, M.H. User involvement and MIS success: A review of research. *Manage. Sci.,* to appear.
13. 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, Nov. 1979.
14. Kerlinger, F.N. *Foundations of Behavioral Research,* 2nd ed. Holt, Rinehart, and Winston, New York, 1973.
15. Larcker, D.F., and Lessig, V.P. Perceived usefulness of information: A psychometric examination. *Decision Sci. 11,* 1 (1980), 121–134.
16. Lucas, H.C., Jr. *Why Information Systems Fail.* Columbia University Press, New York, 1975.
17. Lucas, H.C., Jr. *The Implementation of Computer-Based Models.* National Association of Accountants, New York, 1976.
18. Maish, A.M. A user's behavior toward his MIS. *MIS Quart. 3,* 1 (March 1979).
19. Nolan, R.L., and Seward, H.H. Measuring information user satisfaction to evaluate systems. In R.L. Nolan (Ed.), *Managing the Data Resource Function.* West Publishing Co., 1974.
20. Nunnally, J.C. *Psychometric Theory.* McGraw-Hill, New York, 1978.
21. Olson, M.H., and Ives, B. User involvement in system design: An empirical test of alternative approaches. *Inform. Manage. 4* (1981).
22. Pearson, S. Measurement of computer user satisfaction. Ph.D. dissertation. Arizona State University, Tempe, 1977.
23. Swanson, E.B. Management information systems: appreciation and involvement. *Manage. Sci. 21* (1974).
24. Zmud, R.W. An empirical investigation of the dimensionality of the concept of information. *Decision Sciences 9,* 2 (April 1978), 187–196.