

Division of Research
School of Business Administration

August 1987

NEW MEASURES FOR THREE USER
ACCEPTANCE CONSTRUCTS: ATTITUDE TOWARD USING,
PERCEIVED USEFULNESS, AND PERCEIVED EASE OF USE

Working Paper #528

Fred Davis
University of Michigan

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or
reproduced without the expressed permission
of the Division of Research.

Copyright 1987
University of Michigan
School of Business Administration
Ann Arbor Michigan 48109

Abstract

As the lack of user acceptance continues to obstruct the effective implementation of new information systems, MIS researchers seek a deeper understanding on the underlying cognitive and affective determinants of user behavior. Investigators increasingly emphasize the need to identify key attitude and belief constructs, and to develop valid, reliable instruments to measure them. Three constructs: attitude toward using, perceived usefulness, and perceived ease of use, are particularly relevant to user acceptance. The purpose of this research is to develop and validate improved measures of these constructs. A standard measure of "attitude toward the act" from social psychology, which has been found highly reliable and valid in various non-MIS contexts, is adapted for measuring attitude toward using. Since existing scales for perceived usefulness and perceived ease of use lack the desired levels of reliability and validity, new multi-item measurement scales for these constructs are systematically developed and validated according to established psychometric principles. First, theoretical definitions for usefulness and ease of use are stated. Fourteen candidate scale items are then generated from prior literature for both usefulness and ease of use. A pre-test is performed to refine the item pools and pare them down to 10 items per construct. A field experiment of 112 users of two different end-user systems is conducted to assess reliability and validity. Multitrait-multimethod analyses indicate very high levels of convergent and discriminant validity for the new scales. Cronbach alpha reliabilities are also quite high, with values of .96 for attitude toward using, .97 for perceived usefulness, and .91 for perceived ease of use. Thus, the new scales possess psychometric properties that far exceed those of existing measurement scales for these constructs. Implications for research and practice are discussed.

1. Introduction

Management Information Systems hold great promise for enhancing the performance of those who use them. All too often, unfortunately, the effective deployment of information technology is impeded by users' unwillingness to accept and use new systems. What causes users to accept or reject new systems? The importance of this question has stimulated considerable research into the underlying cognitive and affective determinants of user behavior. Research on user beliefs and attitudes has long been a fundamental avenue of inquiry for MIS research, driven by a desire to understand what determines user behavior and how user acceptance can be measured and influenced. Subjective measure of user beliefs and attitudes have been studied both as dependent variables in laboratory studies on the impact of differing system design characteristics (e.g., Benbasat & Dexter, 1986, Dickson, DeSanctis & McBride, 1986; Ghani & Lusk, 1982; Malone, 1981), and as independent variables in field studies of the determinants of implementation success (e.g., Ginzberg, 1981; Lucas, 1975; Robey, 1979; Schultz & Slevin, 1975; Swanson, 1974).

Belief and attitude measures are valuable both to MIS researchers, for testing hypotheses and theories about user acceptance, and to MIS practitioners, for evaluating user acceptance of specific systems and reducing the risk of wasting considerable development and implementation resources. In both contexts, the quality of conclusions drawn are only as good as the quality of measures employed. Faulty measures increase the likelihood of drawing incorrect conclusions from observed data. The MIS field is increasingly emphasizing the need to (1) identify the specific beliefs and attitudes which are most relevant to user acceptance (Ginzberg, 1981; Swanson, 1982) and (2) develop valid and reliable measures for these constructs.

Within this overall context, the goal of the present research is to develop and test improved measurement scales for three key MIS constructs: attitude toward using, perceived usefulness, and perceived ease of use. Although certainly not the only variables of interest to MIS research and practice, previous work suggest that these variables play a particularly central role in user acceptance processes. Moreover, the psychometric properties of existing measurement scales for these variables fall well short of ideal.

Attitude Toward Using

MIS attitudes have been heavily researched over the past several years. Early studies by Swanson (1974), Lucas (1975) and Schultz and Slevin (1975) established the basic paradigm, and many investigators have followed this line of inquiry (e.g., Fuerst & Cheney, 1982; Ginzberg, 1981; Maish, 1979; Robey, 1979; Robey & Zeller, 1978; Schewe, 1976). The importance of the attitude construct stems from the longstanding research tradition within psychology concerned with predicting and explaining behavior from peoples' attitudes. Indeed, the attitude-behavior relationship is one of the most widely studied areas in psychology (for review, see Cooper &

Croyle, 1984). A wide range of MIS attitude measures have been employed, although, by and large, investigators have not paid attention to the psychometric properties of their scales. While MIS attitudes are often significantly correlated with usage (e.g., Baroudi, Olson & Ives, 1986; Fuerst & Cheney, 1981; Ginzberg, 1981; Lucas, 1975; Maish, 1979; Swanson, 1974), this is not always the case (e.g., Schewe, 1976; Srinivasan, 1985). Reflecting on this mixed pattern of findings, Swanson points out in his 1982 review: "Variations in the attitude concept itself appear to explain much of the variation in research results reported to date." (Swanson, 1982, p. 161).

The mixed findings concerning the MIS attitude-usage relationship are consistent with past findings on the attitude-behavior linkage within psychology more generally (Wicker, 1969). Marking a significant step toward accounting for these anomalous results, Ajzen and Fishbein (1977) observed that the strength of the attitude-behavior relationship depends a great deal on exactly how the attitude measure is worded. For instance, a person's attitude toward an attitude object is less related to a particular behavior involving the object than is their attitude toward the behavior itself. This suggests, for example, that usage of a particular system would be better predicted by a measure of users' attitudes toward the act of using the system than by their attitudes toward the system per se. Attitude toward using corresponds in specificity to the behavior to be predicted with respect to action. In addition, according to Ajzen and Fishbein (1977), the attitude measure will predict better if it corresponds in specificity to the behavior with respect to the target (e.g., the specific system as opposed to computer systems in general), context (e.g., at work vs. at home), and time frame (e.g., this week, next year, etc.). More general attitudes are appropriate for predicting more general behaviors, and more specific attitudes are appropriate for predicting more specific behaviors. The key point is that the ability of an attitude measure to predict behavior depends on the degree to which the attitude and behavior under investigation are defined at parallel levels of specificity.

Based on Fishbein and Ajzen's (1975) general definition of "attitude toward the act," attitude toward using is defined as "the degree of evaluative affect an individual associates with using the target system in his or her job." This definition of attitude is specific with respect to target (the designated system), action (using), and context (in the person's job) and non-specific with respect to time frame. Fishbein and Ajzen (1975) recommend the use of measurement scales with adjective pairs (such as 'good-bad') that have been shown to load on the evaluative dimension of the semantic differential (Osgood, Suci & Tannenbaum, 1957). These recommended measures employ 7-point rating scale formats and typically exhibit reliability values above .8 (e.g., Bagozzi, 1981; Fishbein & Raven, 1962). The standard scales are readily adapted to the present context by indicating the desired target (system), behavior (using the system), context (in your job), and time frame (non-specific). Four to five items are typically employed in order to assure the desired psychometric properties.

Why should these standard attitude measures be any better than existing MIS attitude measures? Consider some of the most carefully validated MIS attitude instruments to date—those developed for measuring MIS user satisfaction (Bailey and Pearson, 1983; Ives, Olson & Baroudi, 1983; Srinivasan, 1985). Although satisfaction is clearly an affective construct, its theoretical relationship to attitude toward using remains unclear. Much of the difficulty can be traced to the psychological ancestry of these two MIS constructs. Whereas attitude toward using derives from attitude theory (e.g., Fishbein & Ajzen, 1975), user satisfaction (e.g., Bailey & Pearson, 1983) derives primarily from the job satisfaction tradition within organizational behavior (e.g., Lawler, 1973). Unfortunately, these two reference paradigms are almost completely disjoint. Two recent review papers in the same issue of the *Annual Review of Psychology* illustrate this lack of integration. In their review of attitude theory, Cooper and Croyle (1984) never mention the satisfaction construct, while Staw (1984) fails to address any relationship between job satisfaction (also called 'job attitudes') and attitude theory, pointing out that "While much effort has historically been placed on developing reliable measures of satisfaction, little work has focused on the construct of satisfaction itself...The field's current usage of satisfaction is as a theory-free affective variable..." (p. 631).

Despite this lack of integration within the base disciplines, Baroudi, Olson and Ives (1986) conducted a study of the relationships between user involvement, system usage, and information satisfaction in which they regarded Bailey and Pearson's satisfaction instrument as a measure of Fishbein and Ajzen's (1975) "attitude toward the object," observing a significant attitude-behavior correlation ($r = .28$). If Bailey and Pearson's UIS index does indeed tap attitude toward the object (system), Ajzen and Fishbein's (1977) admonition about correspondence in specificity suggests that attitude toward the act should relate even more strongly to behavior. Moreover, in contrast to the comparatively long UIS instrument (156 items), Fishbein and Ajzen (1975), and others have been able to achieve high levels of reliability and validity using 4 semantic differential items. Thus, both theoretical and methodological considerations favor the recommended Fishbein and Ajzen scales as the basis for a standard measure of the "attitude toward using" construct.

Perceived Usefulness and Perceived Ease of Use

In much of the previous research on MIS attitudes, investigators have not drawn the distinction between beliefs and attitudes (Swanson, 1982). However, in contemporary psychology, this distinction is widely recognized. Whereas a person's attitude (also called affect) refers to his or her feelings about or evaluation of an object or action, beliefs (also called cognitions, judgments and perceptions) refer to the degree to which an individual believes in or agrees with a statement which is capable of objectively being true or false (e.g., Bagozzi, 1983;

Triandis, 1977; Fishbein & Ajzen, 1975). The so-called "tri-partite" model of attitudes shows that cognitions, affect, and behavioral tendencies are statistically distinct but related psychological phenomena (Ostrom, 1969). Several theoretical models in psychology are concerned with the relationships between beliefs, attitudes and behavior (e.g., Bagozzi, 1981; Fishbein & Ajzen, 1975; Triandis, 1977).

Among the various beliefs that have been studied in MIS research, perceived usefulness, defined here as "the degree to which an individual expects that using a particular system would enhance his or her job performance," has considerable theoretical and empirical appeal. The powerful role of this construct for user acceptance was suggested by the work of Schultz and Slevin (1975). In an instrument development study, they administered a questionnaire regarding a new computer forecasting model to 94 management personnel. A factor analysis of the 67 questionnaire items yielded 7 factors, which they labeled: performance; interpersonal; changes; goals; support/resistance; client/researcher; and urgency. Schultz and Slevin interpreted the performance factor as the perceived "effect of the model on manager's job performance," which is almost exactly the same as the definition of perceived usefulness proposed above. Also, they measured a dependent variable called "perceived worth" on a scale with anchoring adjectives: "not useful at all"; "moderately useful"; and "excellent", which also appears to tap the "perceived usefulness" construct. The correlation between perceived worth and the performance factor ($r = .59$) is consistent with this possibility. Moreover, of the 7 factors revealed by the factor analysis, the performance dimension correlated most highly with a measure of self-predicted use ($r = .61$).

Independently, but at about the same time, Gallagher (1974) reported results of a comparison between two alternative approaches for assessing the value of information reports: estimated annual dollar value and semantic differential opinions. In a field study of 52 managers, the correlation between these two approaches was .29 ($p < .05$). Many of the 15 adjective pairs were similar in meaning to perceived usefulness: useful-useless; relevant-irrelevant; important-unimportant; applicable-inapplicable; necessary-unnecessary.

Robey (1979) and Robey and Zeller (1978) conducted further research using Schultz and Slevin's instrument, both observing that "performance" was the dimension most strongly related to system usage. Robey (1979) and Vertinsky, Barth and Mitchell (1975) called for an interpretation of the performance-usage relationship within the work motivation theory of Vroom (1964) and Porter and Lawler (1968). In those models, the individual's expectation that job effort will lead to performance is a major determinant of effort. In the proposed adaptations of the work motivation theory, the effort-performance expectancy is replaced by the usage-

performance expectancy. This provides the theoretical underpinning for the definition of perceived usefulness proposed in the present research. A more recent adaptation of Vroom's (1964) work motivation model by DeSanctis (1983) proved unsuccessful in explaining DSS usage in a laboratory simulation study, which is surprising given the previously observed strong linkage between expected performance and usage. Possible problems concerning the novel measure she used to tap expectancies and values may be partly at fault.

Franz and Robey (1986) recently introduced a new perceived usefulness scale. Although the construct validity of the 12-item scale was not assessed, using multitrait-multimethod analysis for example (Campbell & Fiske, 1959), content validity was assessed via phone interviews with 10 users. In a survey of 118 users, the scale obtained a Cronbach alpha reliability of .84. However, the scale measures a much broader domain of content than just the individual's expectation regarding the use-performance relationship. Three of the items tap self-reported use of the system (#1, 5 & 12) and two of the items ask about ease of use (#9 & 10). The present research regards usage and perceived ease of use as distinct constructs, separate from, but related to, perceived usefulness.

Perceived ease of use, defined here as "the degree to which an individual believes that using a particular system would be free of physical and mental effort," represents a second key belief underlying user acceptance. When seemingly useful systems are rejected by intended users, difficulty of use is a frequently cited explanation. Practicing systems designers are increasingly concerned with designing systems for usability (Gould & Lewis, 1985; Bewley et al., 1983), often impelled by mandates from senior technical management (e.g., Branscomb & Thomas, 1984). In many vendor organizations, usability testing has become a standard phase in the product development cycle, with large investments in sophisticated usability test facilities, advanced monitoring instrumentation, and human factors design specialists. Entire bodies of psychological theory are devoted to the modeling and prediction of computer usability (e.g., Card, Moran & Newell, 1983).

For the most part, human factors efforts have focused on "objective" ease of use, as measured by such performance criteria as task completion time, number of keystrokes, number of errors, and the like (e.g., Brosey & Shneiderman, 1978; Card, Moran & Newell, 1980; Roberts & Moran, 1983; Thomas & Gould, 1975). Although objective ease of use is clearly relevant to performance assuming the system is used, subjective ease of use should be more relevant in the users' decision whether or not to use the system. Human factors researchers have begun to pay more attention to perceived ease of use recently (e.g., Bewley et al., 1983; Good, 1982; Poller & Garter, 1983).

Currently, however, the instruments used to assess perceived ease of use tend to be ad hoc, non-validated measures.

Is perceived ease of use distinct from perceived usefulness? The question can be approached from both conceptual and empirical perspectives. Conceptually, perceived usefulness and perceived ease of use can be regarded as the subjective counterparts of a system's objective functionality and usability (Goodwin, 1987). Functionality refers to whether the system is capable of performing the tasks users need to carry out, whereas usability is concerned with how easy or hard it is to carry out those tasks using the system. Empirically, several factor analytic studies suggest that perceived usefulness and perceived ease of use are distinct variables. Zmud (1978) factor analyzed the ratings of three information reports to derive the dimensionality of the concept of information. Of the eight dimensions derived, one related to perceived usefulness ("relevant") and one related to perceived ease of use ("readable"). Larcker and Lessig (1980) factor analyzed six items used to rate four information sets. Three items loaded on each of two distinct factors: (1) "perceived importance", which Larcker and Lessig define as "the quality that causes a particular information set to acquire relevance to the decision maker," and the extent to which the information elements "are a necessary input for task accomplishment," and (2) "perceived usability" which is defined as the degree to which "the information format is unambiguous, clear, or readable." Clearly, these two dimensions are similar to what the present research terms perceived usefulness and perceived ease of use, respectively. Hauser and Simmie (1981) derived two perceptual dimensions, ease of use and effectiveness, by factor analyzing ratings of various communications technologies. Effectiveness is similar to usefulness. Swanson (1982), within the context of his channel disposition model, factor analyzed information report ratings, yielding a 5-dimension solution. Factor #2, "accessability," contained items such as "convenient," "controllable," "easy," "flexible," and "untroublesome." Hence it is similar to perceived ease of use. Factor #3, "value," contained items such as "important," "meaningful," "relevant," "useful," and "valuable; it was therefore tapping perceived usefulness. Across four independent factor analyses, perceived usefulness and perceived ease of use consistently emerged as distinct factors. Both conceptually and empirically, therefore, perceived usefulness and perceived ease of use appear to be distinct constructs.

Existing Scales

No existing validated, multi-item scales with the minimally desired reliability of .80 were found for perceived usefulness or perceived ease of use. The Fishbein paradigm (e.g., Ajzen & Fishbein, 1980, Appendix A) provides a recommended format for beliefs once they have been specified, but it does not furnish complete scales for specific belief variables. The majority of

studies measuring usefulness or ease of use either employed single-item scales or failed to report the psychometric characteristics of the multi-item scales used. The remaining candidate multi-item scales exhibited reliability below the standard "rule of thumb" minimal reliability of .80, were unvalidated, or both. Robey (1979) employed Schultz and Slevin's (1975) instrument, which contains a factor called "performance" that is similar to perceived usefulness. He found a Cronbach Alpha reliability of .81 for the scale, although the original instrument was non-validated, having been developed via exploratory factor analysis. Moreover, the performance scale contained items relating to "performance visibility" as well. Larcker and Lessig (1980) did perform a content analysis validation of their 3-item scales of "usableness" and "importance" but the reliabilities (.64-.77) fell short of our desired level of .80. Ginzberg's (1981) 2-item "importance" scale achieved a reliability of .59. Bailey and Pearson's (1983) instrument contained three 4-item semantic differential scales of usefulness-like factors ("relevance", "perceived utility" and "job effects") and four 4-item scales of factors that are similar to ease of use ("error recovery", "understanding of systems", "feeling of control", and "flexibility of systems"). However, the definitions given to respondents for each of these factors depart considerably from the conceptual definitions of usefulness and ease of use in the present research. Bailey and Pearson (1983) performed a content analysis validation, although they did it from the standpoint of these factors as measures of "computer user satisfaction" as opposed to usefulness and ease of use per se. Miller (1977) did not present evidence of reliability or validity for his 3-item ease of use and usefulness scales, nor did Schewe (1976) give such evidence related to his 3-item ease of use scale.

Given the lack of sufficiently reliable and valid scales in the existing literature, new scales for perceived usefulness and perceived ease of use will be developed. As discussed below, the scales found in the existing literature will be used as a source of items for constructing the new scales. Consistent with Ajzen and Fishbein (1980), beliefs will be measured using Likert-type ('agree-disagree') rating formats.

2. Measure Development Approach

The psychometric literature is vast, with multiple definitions of reliability and validity. Before proceeding, reliability and two aspects of construct validity, content validity and common method variance, will be defined as they will be used in this paper. The following true-score model is frequently used to conceptualize the role of error in a measure:

$$X_i = T_i + e_i$$

X_i = observed score from subject i

T_i = true score for subject i

e_i = random error for subject i

The observed score X_i , measured as a subject's response to a belief or attitude scale, is jointly determined by the true score that the measure is intended to tap, T_i , which is not directly observable, and a random error e_i which is independent of T_i and has zero variance. The reliability and construct validity of a measure can be described in terms of this true score model.

Reliability

Reliability refers to the extent to which a measurement scale is free of random error (e.g., Nunnally, 1978, p. 191), and is generally defined as the proportion of variance in the observed score X_i that is due to the true score T_i . Assuming statistical independence of T_i and e_i , the variance of the observed score equals the variance of the true score plus the error variance.

$$\sigma^2_x = \sigma^2_t + \sigma^2_e$$

Reliability, then, is defined as follows:

$$R = \sigma^2_t / \sigma^2_x = 1 - \sigma^2_e / \sigma^2_x$$

As the amount of random error in a measure increases, its reliability diminishes. Unreliable measures create various difficulties for statistical analyses in which they are used. For comparisons between mean values, unreliability inflates the standard errors of the estimated means. Further, unreliability leads to underestimation of correlation and regression coefficients relative to what their true value would be with error-free measures. To the extent that the measures employed are high in random error (i.e., are unreliable) the statistical conclusion validity (Cook & Campbell, 1979) of a study is threatened. Observed empirical relationships may understate true underlying relationships between constructs, increasing the possibility of false "no-effect" conclusions (type II error).

Although it is generally impossible to completely eliminate random measurement error, it is possible to substantially reduce the error so as to minimize its effect on statistical tests. We will employ a target reliability level of .80 based on Nunnally's (1978) suggestion that: "For basic research, it can be argued that increasing reliabilities much beyond .80 is often wasteful of time and funds. At that level correlations are attenuated very little by measurement error" (p. 245).

Construct Validity

Although the construct validity of a measurement scale has been defined in various ways in the psychometric literature (e.g., Bohrnstedt, 1970; Cronbach & Meehl, 1955; Nunnally, 1978), a predominant perspective views a measure's construct validity as the degree to which the measure's true score corresponds to the theoretical construct that the measure is intended to operationalize (e.g., Cook and Campbell, 1979, p. 59; Fishbein & Ajzen, 1975, p. 108). Whereas reliability is concerned with the amount of random variance in an observed score, construct validity is concerned with the degree to which the systematic variance in a score corresponds to the target construct.

Invalidity may come about in two conceptually distinct ways. First, a measure's true score component may tap some alternative theoretical variable other than the one intended. The correspondence between a measurement scale and the theoretical variable of interest is sometimes referred to as "content validity" (e.g., Bohrnstedt, 1970; Nunnally, 1978). If a measure lacking content validity is employed, researchers may incorrectly interpret the resulting data in terms of the theoretical variable that was intended by the measure, rather than the variable that was actually measured. Thus the observed relationship may either overstate or understate the true relationship between variables. Second, methodological artifacts unrelated to the target theoretical variable, such as individual differences in response set (e.g., Campbell, Siegman, & Rees, 1967; Silk, 1971), may comprise part of the systematic variance in a measure. This type of invalidity may be a source of spurious covariation between variables whose measures are affected similarly by such methodological artifacts. The resulting data may overstate the magnitude of the true underlying relationship. This source of invalidity is sometimes referred to as "shared method variance" (Campbell & Fiske, 1959, p. 85). In the present research, different techniques will be employed to deal specifically with each of these potential sources of measure invalidity.

Multi-item Scales

The present research will employ multi-item measurement scales. Whereas single-item scales tend to be less valid and reliable, possessing higher degree of irrelevant content along with the target content (Cook & Campbell, 1979, p. 65), the use of multi-item summed scales tends to allow the irrelevancies of individual items to cancel out, increasing reliability and validity. A primary method for increasing the reliability of a scale is to increase the number of items (Nunnally, 1978, p. 243). The individual items will use semantic differential and Likert-type rating formats. These types of items have traditionally been used in attitude scaling, and are the ones recommended by Ajzen and Fishbein (1980) for operationalizing beliefs and attitudes.

Substantial experience in applying items such as these has shown that they are generally capable of attaining high levels of reliability and validity (Fishbein & Raven, 1962; Jaccard, Weber & Lundmark, 1975; Ostrom, 1969; Robinson & Shaver, 1969; Shaw & Wright, 1967). Moreover, they are quite easy to use by non-experts, making them suitable for use by MIS practitioners in applied contexts.

Measure Development Process

The process used in the present research for developing usefulness and ease of use scales was designed to address the three key psychometric properties identified above: reliability, content validity and common method variance. The process follows contemporary views concerning psychological test construction. The following quote by noted psychometrician Anne Anastasia (1986) provides a perspective on the process used:

“More and more we recognize that the development of a valid test requires multiple procedures, which are employed sequentially at different stages of test construction...Validity is thus built into the test from the outset rather than being limited to the last stages of test development, as in traditional criterion-related validation. The validation process begins with the formulation of detailed trait or construct definitions, derived from psychological theory, prior research, or systematic observation and analysis of the relevant behavioral domain. Test items are then prepared to fit the construct definitions. Empirical item analyses follow, with the selection of the most effective (i.e., valid) items from the initial item pools.”

Thus, in the present research, the conceptual definitions of perceived usefulness and perceived ease of use are used to generate initial pools of 14 candidate items for each construct based on existing MIS and Human Factors literature. Next, pretest interviews were conducted in order to perform a content analysis of the items. The item generation and pretest steps were performed in order to increase the content validity of the measures, and are motivated by the “domain sampling” model of test construction, which is discussed in the following section. This pre-test is used to prune the number of items from 14 down to 10 for each belief construct. Next, a field experiment of 112 users of 2 computer systems is performed to empirically evaluate the reliability and construct validity of the resulting scales.

3. Measurement Scale Item Generation

The first step in the measure development process is to identify an initial set of measurement items as candidates for the final usefulness and ease of use scales. The candidate items will be derived from published articles that have discussed or attempted to measure the target constructs. As discussed above, the process used for generating items aims to ensure that the items possess content validity, which is defined as “the degree that the score or scale being used represents the concept about which generalizations are to be made” (Bohrnstedt, 1970, p. 91). As Nunnally (1978, 9. 258) points out in his discussion of content validity: “Rather than test the

validity of measures after they have been constructed, one should ensure the validity by the plan and procedures for construction.” In order to explain why generating items from the existing literature is expected to increase the content validity of the resulting measures, we now introduce the “domain sampling model” of measure construction, which psychometricians frequently employ as a conceptual tool to guide the measure development process (e.g., Bohrnstedt, 1970, p. 92; Nunnally, 1978, p. 193). Under the domain sampling model, there is assumed to be a universe or domain of content corresponding to each variable one is interested in measuring. Under this model, the optimal way to develop a scale would be to specify the domain of content corresponding to the target construct and then randomly sample items from the domain. The mean value of a summative scale composed of such a randomly sampled subset would theoretically provide an unbiased estimate of the mean of all the items in the entire domain (i.e. the universe score), which in turn corresponds to the magnitude of the true underlying construct. However, it is ordinarily not possible to rigorously specify the domain of content corresponding to psychological constructs (Bohrnstedt, 1970; Cronbach & Meehl, 1955; Nunnally, 1978).

Since it is impossible to completely define the domain of content corresponding to the target constructs and exhaustively identify all the items in the domain, domain sampling in its pure form cannot be achieved. However, there are a number of steps which can be taken to enable us to approximate domain sampling. First, since the conceptual definitions of the variables serve as a rough specification of the appropriate domains of content, they should be employed as a guide to measure development. This is what Cook and Campbell (1979, p. 64) refer to as “preoperational explication of constructs” in their discussion of construct validity, wherein they suggest that measures be tailored to fit the conceptual meaning of the target construct. Second, existing literature may be used as a source of domain content. Repeated attempts to define, theorize about, and measure a given construct gradually reveal the nature of its underlying domain of content (Bohrnstedt, 1970; Cook and Campbell, 1979; Nunnally, 1978). Given the existence of numerous published articles dealing with perceived usefulness and perceived ease of use, prior literature represents an important source of content for measure development. To obtain elements of content from the existing literature, guided by the conceptual definitions of the target constructs, represents an approximation to domain sampling. Third, as will be discussed later, a content analysis may be performed for the purpose of improving this approximation to domain sampling.

An alternative to the present item generation approach frequently employed is to elicit items from subjects in qualitative individual or focus group interviews (e.g., Calder, 1977; Churchill, 1979; McKennell, 1974). Generating items from the literature has two of advantages over direct

elicitation in the present context, however. First, there is a rich set of existing articles available to draw from, many of which have themselves employed a variety of qualitative elicitation as well as quantitative analysis techniques to understand how subjects think about these constructs. Second, these existing articles cut across a wide range of target systems, user populations and usage environments. Given the objective of creating a general model that is applicable across many contexts, the existing literature is likely to provide a more generalized representation of the desired content domains. In-depth interviews would, by necessity, be restricted to a limited user population and range of systems, which may result in highly context-specific content domains.

Thus, the measurement item candidates were generated by drawing item content from existing published studies in the Management Information Systems and Human Factors Fields. The following definitions of perceived usefulness and perceived ease of use were used as a guide for selecting which items from the literature to include in the initial pools:

Perceived Usefulness: The degree to which an individual believes that using a particular system would enhance his or her job performance.

Perceived Ease of Use: The degree to which an individual believes that using a particular system would be free of physical and mental effort.

The next step is to determine the number of items to be generated for the initial item pools. This is approached by first estimating the number of items required to achieve the desired level of reliability in the final scales, and then adding 4 additional items to account for the plan to eliminate 4 of the items based on the subsequent interviews and associated content analysis. The anticipated scale length required to achieve a Cronbach alpha reliability of .80 was estimated using the Spearman-Brown Prophecy formula:

$$\alpha = k \alpha' / [1 + (k-1) \alpha']$$

where α = desired reliability level

α' = reliability of comparable scale with n items

kn = number of items needed to achieve desired reliability

Twelve existing scales of various constructs from three published MIS studies (Ginzberg, 1981; Larcker & Lessig, 1980; Robey, 1979) were analyzed using the Spearman-Brown formula. This analysis suggested that 10 items would be required for each perceptual variable to achieve the target reliability level of .80. Thus, adding the 4-item margin needed for item elimination, it was decided to generate 14 candidate items for each variable.

The item pools for perceived usefulness and perceived ease of use are given in Tables 2 and 3

Table 2. Perceived Usefulness Item Pools

Item #	Item Wording
1	My job would be difficult to perform without electronic mail.
2	Using electronic mail gives me greater control over my work.
3	Using electronic mail improves my job performance.
4	The electronic mail system addresses my job-related needs.
5	Using electronic mail saves me time.
6	Electronic mail enables me to accomplish tasks more quickly.
7	Electronic mail supports critical aspects of my job.
8	Using electronic mail allows me to accomplish more work than would otherwise be possible.
9	Using electronic mail reduces the time I spend on unproductive activities.
10	Using electronic mail enhances my effectiveness on the job.
11	Using electronic mail improves the quality of the work I do.
12	Using electronic mail increases my productivity.
13	Using electronic mail makes it easier to do my job.
14	Overall, I find the electronic mail system useful in my job.

respectively. They are worded in terms of "electronic mail" as an example system. A wide range of published literature was drawn upon in generating the items. In addition to the empirical studies reviewed in Chapter 3, theoretical papers and reports of in-depth qualitative studies were used. Table A.1 in the chapter appendix specifies the articles used for abstracting

Table 3. Perceived Ease of Use Item Pools

Item #	Item Wording
1	I often become confused when I use the electronic mail system.
2	I make errors frequently when using electronic mail.
3	Interacting with the electronic mail system is often frustrating.
4	I need to consult the user manual often when using electronic mail.
5	Interacting with the electronic mail system requires a lot of my mental effort.
6	I find it easy to recover from errors encountered while using electronic mail.
7	The electronic mail system is rigid and inflexible to interact with.
8	I find it easy to get the electronic mail system to do what I want it to do.
9	The electronic mail system often behaves in unexpected ways.
10	I find it cumbersome to use the electronic mail system.
11	My interaction with the electronic mail system is easy for me to understand.
12	It is easy for me to remember how to perform tasks using the electronic mail system.
13	The electronic mail system provides helpful guidance in performing tasks.
14	Overall, I find the electronic mail system easy to use.

the items, and Table A.2 gives the correspondence between these articles and specific ease of use and usefulness items.

The items within each item pool tend to have a lot of overlap in their meaning. This is expected since they are intended to be measures of the same underlying construct. Though individuals may attribute slightly different meaning to particular item statements, the goal of the multi-item approach is to downplay the effects of individual items, allowing idiosyncrasies to be cancelled out by other items, yielding a more pure indicant of the underlying construct.

4. Item Pre-test Interviews

Purpose

The purpose of the pretest interviews is to further assure content validity by empirically assessing the semantic correspondence between the measurement items contained in the item pools and the underlying variables they are intended to measure. By deriving the item pools from numerous existing studies attempting to measure the perceptual variables, we have some assurance that they provide a broad coverage spanning the domains of their respective constructs. However, we must regard this as only an approximation of what we would have obtained had we actually been able to draw sample items from their underlying content domains according to the domain sampling model.

The pretest interview's aim is to improve this approximation. Let us consider the potential deficiencies of this approximation and how it may be improved. First, although the items selected for the item pools were initially assumed to reside within the domain, it is possible that some of the items do not really belong to the domain. We can attempt to identify and remove these items by asking participants to rate the degree to which a statement corresponds in meaning to the definitions of usefulness (or ease of use) and eliminating items receiving low ratings. Recall that four additional items for each perceptual construct were added during item generation to provide for this elimination process.

Second, our selection process lacked the randomness of item selection employed by the idealized domain sampling. As a consequence, our item pools may have too much coverage in some areas of meaning, or sub-strata (Bohrstedt, 1970, p. 92) within the domain and not enough in others. We can gather data to assess and improve the approximation to random sampling. In this case we ask subjects to rate the similarity of items to one another (using a categorization process). Based upon such data, we can infer the nature and structure of domain sub-strata, remove items in sub-strata where excess overlap exists, and add items to sub-strata where inadequate coverage is revealed.

Method

Subjects. A convenience sample of 15 subjects from the Sloan School of Management, MIT, participated in the pretest interviews. The sample included 5 secretaries, 5 graduate students and 5 members of the professional staff. All were experienced computer users.

Materials. Materials for the interviews were twenty-six 4 by 6 inch index cards. Each card had one Likert statement printed on it. The twenty-six statements corresponded to thirteen of the Likert items for each of the two perceptual variables: perceived usefulness and perceived ease of use. The fourteenth, or "overall" item for each construct was omitted since its wording was similar to the label given to the definitions of the constructs against which subjects were asked to compare the remaining items, as discussed below. Electronic mail was used as the example target system in the item wordings. A random identification number was printed on the back of each of the cards.

Procedure. The procedure was administered via face-to-face interviews and consisted of two tasks, prioritization and categorization, which were each repeated separately for the 13-card decks corresponding to usefulness and ease of use. For prioritization, subjects were first given a card upon which the label and definition of the target construct was printed and asked to read it. Next, they were asked to "rank order these 13 statements according to how well each statement's meaning matches the definition of usefulness (ease of use). Put the statement that most closely matches the meaning of usefulness (ease of use) on the top of the deck, put the statement that least matches the meaning of usefulness (ease of use) on the bottom, and so on. Electronic mail was selected as an example system only; our interest is in the meaning of the statements themselves." Extensive experience with card sorting as a data collection technique suggests that subjects find it enjoyable and exhibit high interest and concentration in the task. In the present interviews subjects appeared to find the card sorting task easy, interesting and involving to perform.

For the categorization task, subjects were asked to "put these 13 statements into categories so that items in a category are most similar in meaning to each other, and different from those in other categories. Use about 3 to 5 categories." This approach an adaptation of the "Own Categories" procedure of Sherif and Sherif (1967). Whereas Sherif and Sherif were concerned with mapping items into categories ordered along an evaluative continuum, in the present research we are concerned with assessing the similarity in meaning of items. That a subject places one item into the same category as another item provides a simple indicant of similarity, and requires less time and subject effort to obtain than other similarity measurement procedures such as diadic or triadic judgements.

Results and Discussion

The procedure yielded data which are summarized in 4 data matrices. Two of these contain the rankings assigned by subjects to the perceived usefulness (Table A.3) and perceived ease of use (Table A.4) items. These ranking matrices give the frequency with which the 15 subjects

placed each item in a particular position in priority. The other two data matrices contain subject ratings of similarity between items for perceived usefulness (Table A.5) and perceived ease of use (Table A.6). Each cell of these symmetric matrices gives the number of subjects who put an item in the same category with some other item during the categorization task. This serves as a measure of the degree of similarity between the items as perceived by the group of subjects as a whole.

The ranking matrices (Tables A.3 and A.4) were used to derive a priority index for each item. The median rank was used as the basis for establishing priority for an item. The median was chosen in preference to the mean because of its robustness to the skewed distribution of the priority ratings. The mean was used to break ties, however. Table A.7 shows the medians, means and resulting priorities for perceived usefulness and perceived ease of use.

A simple cluster analysis was performed on the two similarity matrices (Tables A.5 and A.6). Items that 7 or more of the 15 subjects placed into the same category were assigned to the same cluster. For example, usefulness items 1 and 4 were coded as belonging to the same cluster. Although the derived clusters were unique in the present context, the simple clustering algorithm used may not yield unique clusters in all contexts. If not, more advanced techniques are available which do yield unique clusters (e.g., Johnson, 1967). The results of this cluster analysis are summarized in Tables 4 (usefulness) and 5 (ease of use), which gives the clusters, item numbers, item names, and item priorities. These clusters are viewed as manifestations of the underlying domain substrata, and as such serve as a basis for assessing the smoothness of domain coverage. For perceived usefulness, notice that items fall into 3 major clusters. The first cluster contains items relating to job effectiveness, the second to productivity and the third to the importance of the system to the job. If we eliminate the lowest ranked items (items 1, 4, 5 and 9), the remaining items exhibit desirable characteristics relative to the objectives of this process. Namely, important clusters (A and B), have neither too much nor too little representation of items, whereas less important clusters (C and D) do not have excess coverage.

Looking now at perceived ease of use (Table 5), we again find 3 major clusters. The first relates to physical effort, while the second relates to mental effort. Selecting the six highest priority items and eliminating the seventh provides solid coverage of the first two clusters. Item # 11 ("understandable"), was reworded slightly to become "clear and understandable" in an effort to pick up some of the meaning of item 1 ("confusing") which has been eliminated. The third cluster is somewhat more difficult to interpret, but appears to be tapping in to perceptions regarding how easy the system is to learn. Remembering how to perform tasks, using the manual, and relying on system guidance are phenomena which are associated with the process of learning to use a new system. Unfortunately, items 4 and 13 provide a rather indirect

Table 4. Perceived Usefulness Item Clusters

Cluster	Item		Priority	Item # for Survey
	#	Name		
A	10	Effectiveness	1	8
	3	Job Performance	2	6
	11	Quality of Work	3	1
B	12	Increase Productivity	4	5
	8	Accomplish More Work	6	7
	6	Work More Quickly	7	3
	9	Reduces Unproductive Time	10	
	5	Saves Me Time	11	
C	7	Critical to My Job	5	4
	13	Makes Job Easier	8	9
	4	Addresses My Needs	12	
	1	Job Difficult Without	13	
D	2	Control Over Work	9	2
Add	14	Overall Usefulness		10

assessment of ease of learning. In order to correct for this, items 4 and 13 were replaced with two items that target ease of learning more directly: "ease of learning" and "effort to become skillful." Items 6, 9 and 2 were eliminated because: a) they did not cluster with other items, and b) they received low priorities, and were therefore regarded as not residing within the content domain for ease of use.

In addition to the nine items remaining from this pretest Interview process, an "overall" item for each construct, generated in the Item Generation process but not included in the pretest interviews, was included to provide a total of ten items per variable for the final scale.

Table 5. Perceived Ease of Use Item Clusters

Cluster	Item		Priority	Keep
	#	Name		
A	8	Controllable	1	4
	10	Cumbersome	2	1
	7	Rigid & Inflexible	6	5
B	3	Frustrating	3	3
	11	Understandable	4	8
	5	Mental Effort	5	7
	1	Confusing	7	
C	12	Ease of Remembering	8	6
	4	Dependence on Manual	9	
	13	Provides Guidance	12	
D	6	Error Recovery	10	
E	9	Unexpected Behavior	11	
F	2	Error Proneness	13	
Add	14	Overall Ease of Use		10
Add		Ease of Learning		2
Add		Effort to Become Skillful		9

5. Measure Validation

The field experiment reported below provides data needed to assess reliability and convergent and discriminant validity. Cronbach's (1951) alpha reliability coefficient will be computed. Alpha was chosen over the alternative available reliability coefficients for a variety of reasons, including: (1) alpha provides a measure of the internal consistency of the items forming a multi-item scale, which is consistent with the domain sampling model by which the scales were developed; (2) it is a generalization of split-half and parallel forms coefficients; (3) compared to test-retest coefficients, it does not require 2-waves of measurement nor does it confound true

fluctuations in the variable with measurement error; and (4) alpha provides a lower bound estimate of the proportion of variance in the observed measurement scale that is attributable to the variance of the true underlying construct (Bohrnstedt, 1970; Cronbach, 1951; Nunnally, 1978). As indicated previously, we set out to achieve a minimum reliability of .80 for our measurements of attitude toward using, perceived usefulness and perceived ease of use.

Campbell and Fiske's (1959) multitrait-multimethod technique will also be applied to the field experiment data, which provides circumstantial evidence of content validity and will permit an assessment of the extent of common method variance in the measures. The multitrait-multimethod technique has been widely used for the purposes of construct validation (Jaccard, Weber, & Lundmark, 1975; Ostrom, 1969; Silk, 1971). Convergent and discriminant validation using this technique provides evidence pertinent to both content validity and common method variance. The multitrait-multimethod approach provide useful circumstantial evidence of content validity- the failure of scales to achieve convergent and discriminant validity would cast doubt on the assumption that the scales correspond to distinct well-defined content domains. The Campbell and Fiske procedure also enables the researcher to gauge the degree of method variance in the items composing scales. To the extent that an item used to measure a trait is high in method variance, it should exhibit attenuated correlation with other items of the same trait, and increased correlation with the same items applied to a different trait, which would be reflected in reduced convergent and discriminant validity.

Method

Subjects and Procedure

Subjects were 112 professionals and managers working for a large North American corporation. A questionnaire was circulated to 120 users on one day and collected from 112 on the following day, yielding a response rate of 93.3 %.

Questionnaire

The questionnaire contained questions regarding two systems that are widely used in the laboratory: electronic mail and the XEDIT file editor. In order to ensure respondent familiarity with the target system, instructions in the questionnaire asked subjects not to fill out the section regarding a given system if they don't use it. Of these 112 participants, 109 completed the section of the questionnaire pertaining to electronic mail and 76 completed the section pertaining to XEDIT. For each system, respondents were asked to rate their perceived ease of use (EOU), perceived usefulness (USEF) and attitude toward using (ATT).

Attitude toward using was measured using standard 7-point semantic differential rating scales as suggested by Ajzen & Fishbein (1980):

All things considered, my using electronic mail in my job is:

Neutral

Good: ____:____:____:____:____:____:____: Bad

In addition, the adjective pairs Wise-Foolish, Favourable-Unfavourable, Beneficial-Harmful and Positive-Negative were used, for a total of five items making up the attitude scale. These are all adjective pairs found to load on the evaluative dimension of the semantic differential (Osgood, Suci & Tannenbaum, 1957).

Perceived ease of use and perceived usefulness were measured using the 10-item measurement scales described in the previous section. Subjects were instructed to circle the number corresponding to their responses on rating scales having the following format:

	Strongly Agree		Neutral		Strongly Disagree		
I find the electronic mail system cumbersome to use.	1	2	3	4	5	6	7

Reliability

As Table 6 shows, the target reliability of .80 was surpassed for attitude, usefulness and ease of use, with reliabilities generally exceeding .90.

Table 6. Cronbach Alpha Reliability of Measurement Scales

Variable	Label	# Items	Cronbach Alpha Reliability		
			E. Mail	XEDIT	Pooled
Perceived Ease of Use	EOU	10	.86	.93	.91
Perceived Usefulness	USEF	10	.97	.97	.97
Attitude Toward Using	ATT	5	.94	.97	.96

Construct Validity

The major source of data used to assess convergent and discriminant validity is the multitrait-multimethod matrix (Campbell & Fiske, 1959), which contains the intercorrelations of the items (methods) making up a scale applied to the two different target systems, electronic mail and XEDIT (traits). For example, ease of use for one system is regarded as a distinct trait from ease of use of XEDIT. Separate multitrait-multimethod matrices were computed for each of our constructs: attitude toward using (Table A.8), perceived usefulness (Table A.9), and perceived ease of use (Table A.10).

Convergent validity

In order to demonstrate convergent validity, items that measure the same trait should correlate highly with one another (Campbell & Fiske, 1959). That is, the elements of the monotrait triangles (the submatrix of intercorrelations between the items intended to measure the given construct for the same system) within the multitrait-multimethod matrices should be large. The 20 monotrait-heteromethod correlations for attitude toward using were all significant, ranging from .57 to .96. Similarly for usefulness, the 90 monotrait-heteromethod correlations were all significant, ranging from .54 to .93. The monotrait-heteromethod correlations for ease of use were generally lower, falling in the range from .06 to .84, with 4 of the 90 correlations (4.4%) being nonsignificant at the .05 level ($r_{12} = .14$, $r_{25} = .06$, $r_{36} = .19$, $r_{56} = .09$). These were all for electronic mail items, which parallels our finding that the ease of use scale applied to electronic mail exhibited the lowest reliability. A possible explanation of why ease of use had some lower item correlations is because, unlike the other two motivational constructs, ease of use items were worded in both positive (e.g., "controllable") and negative (e.g., "cumbersome") directions. A separate analysis of positively and negatively worded items will be discussed later.

Discriminant validity

The multitrait-multimethod matrices (Tables A.8, A.9 & A.10) are also used to assess discriminant validity. The criterion is that an item should correlate more highly with other items that are intended to measure the same trait than it correlates with either the same item used to measure a different trait or different items used to measure a different trait (Campbell & Fiske, 1959). Formally, this comparison may be specified as:

$$r(X_{1i}, X_{1j}), i \neq j > r(X_{1i}, X_{2k}), \text{ where } X_{1i} \text{ and } X_{2i} \text{ refer to item } i \text{ used to measure traits 1 and 2 respectively.}$$

For example, within the multitrait-multimethod matrix, the correlation between items 1 and 2 measuring usefulness for electronic mail should be larger than the individual correlations between all 10 usefulness items applied to XEDIT and items 1 and 2 applied to electronic mail. For attitude toward using, the monotrait-heteromethod correlations exceeded their corresponding heterotrait-heteromethod and heterotrait-monomethod correlations for all 200 comparisons without exception. Similarly for usefulness, 1800 such comparisons were confirmed without exception. Of the 1800 comparisons for ease of use there were 58 exceptions (3%). These exception were associated with ease of use items applied to electronic mail, and involved the following items (broken down by monotrait-heteromethod vs. heterotrait-heteromethod and monotrait-heteromethod vs. heterotrait-monomethod comparisons):

<u>Item#</u>	<u>MTHM vs. HTHM</u>	<u>MTHM vs. HTMM</u>
1	6	0
2	10	1
4	0	1
5	16	7
6	5	1
7	4	4
9	3	0

The large number of MTHM vs. HTHM disconfirmations associated with items 1, 2 and 5 are probably due in part to the low monotrait correlations associated with these items, as discussed in the context of convergent validity above. The large number of MTHM vs HTMM exceptions for items 5 and 7 is related to the high heterotrait-monomethod correlations for these items (.33 and .30 respectively) coupled with the generally low pattern of monotrait correlations.

Table A.11 gives a summary frequency table of the correlations comprising the multi-method matrices for attitude, usefulness and ease of use. From this table it is possible to see the clear separation in magnitude between monotrait and heterotrait correlations for attitude and usefulness, and the relatively low monotrait correlations for ease of use applied to electronic mail, resulting in an overlap with the heterotrait correlations. Also, notice that the monotrait correlations tend to be higher for XEDIT than electronic mail. This increase in convergence may have resulted from the fact that the XEDIT scales were filled out after the electronic mail scales, and the greater familiarity with the scales may have reduced random error. The frequency tables also show that the heterotrait-heteromethod correlations do not appear to be substantially

elevated above the heterotrait-monomethod correlations. This is an additional diagnostic suggested by Campbell and Fiske (1959, p. 85) to detect the presence of method variance.

Direction-of-wording Effect

The multitrait-multimethod analysis found a small proportion of exceptions to the convergent (4.4% exceptions) and discriminant (3% exceptions) validity criteria. While this would typically be strong evidence in favor of the validity of the ease of use scale (e.g., Campbell & Fiske, 1959; Silk, 1971), it is worthwhile pursuing why these exceptions occurred, and examining whether the scale can be improved. One characteristic differentiating the ease of use scale from the attitude and usefulness scales is the use of a mixture of positively and negatively worded items for the ease of use scale. The odd numbered ease of use items are framed negatively. Examination of the multitrait-multimethod matrix shows that the low monotrait correlations for ease of use for electronic mail tend to be associated with odd numbered (negative) items, and that the highest heterotrait-monomethod correlations are associated with odd numbered items (5 and 7). This suggests that convergent and discriminant validity may be improved by employing just the positive items.

Table A.12 gives separate frequencies of the multitrait-multimethod correlations for the positive and negative ease of use items. The frequencies show that the monotrait-heteromethod correlations for the positive items are higher than those for the negative items, especially for electronic mail. The magnitude of these correlations is evidence of convergent validity for the positive items, with all correlations being significant, and 9 out of 10 falling in the .50-.79 range. Moreover, it may be the case that the presence of the negative items exerted a downward influence on the correlations of the positive items. Cronbach alpha for the positive items was found to be .92 for electronic mail and .94 for XEDIT, compared to .73 and .89 respectively for the negative items. This implies greater random error for the negative items, with the error being less for XEDIT possibly due to practice effects. Two of the heterotrait-monomethod correlations were especially high (in the .30-.39 range) for the negative items, suggesting the presence of method variance. A separate assessment of discriminant validity on the positive ease of use items found no exceptions out of 200 comparisons. Cronbach alpha reliability for the positive items, pooled across systems is .92.

Another way to assess the effect of the negative items on method variance is to compare the correlations between systems on the entire ease of use scale before and after omitting the negative items. The correlation across systems was significant for the original ease of use scale ($r = .22, p < .05$), although not for usefulness ($r = .18, n.s.$) or attitude ($r = .09, n.s.$). The observed correlation between scores on the same scale applied to different systems may be due to a

combination of "true" correlation of the underlying traits and "artifactual" correlation due to shared method variance. On theoretical grounds, we may expect there to be a "true" correlation for each of the three constructs. Ease of use may be jointly determined by the characteristics of the system as well individual characteristics such as general computer experience and intelligence. Such individual characteristics may have a similar effect on a person's perceived ease of using two different systems, producing a true trait correlation across systems. Similarly, attitude towards using computers in general may influence attitudes toward using two specific systems, and the characteristics of an individual's job may have a simultaneous influence on the perceived usefulness of two similar systems. Both of these variables may vary across subjects, producing true covariation across systems. To remove the negative ease of use items is expected to only reduce artifactual common method variance and not true variance, since the remaining scale composed of positive items has greater reliability (i.e. greater true score variance) than the original 10-item scale. When the negative items were removed from the ease of use scale, the correlation between ease of use scores across systems fell from .22 to .10. The drop in correlation is attributable to a reduction in common method variance, which was detected in the original scale by the multitrait-multimethod analysis.

Thus, using only the positive items brings the convergent and discriminant validity of the ease of use scale in line with that of the usefulness and attitude scales. The negatively worded items have a higher degree of random error and method variance. Reversing the direction of wording of items making up a scale is often advised in order to reduce the effect of methods variance (e.g. Cook & Campbell, 1979, p. 66). Ironically, just the opposite occurred here, with reversed items adding substantial method variance. Our finding parallels a finding made by Silk (1971, 393) that: "the 'reversed' item appears to be affected by method factors more than any of the other items except item 1." Evidence suggests that direction-of-wording effects are typically much smaller than trait variance, however (Campbell, Siegman, & Rees, 1967), which is consistent with the present pattern of results. Given the substantial disadvantages of the negative ease of use items in the present context, it was decided to omit them from the ease of use scale for the analysis of the survey data.

Scale Refinement

For purposes model evaluation, a refined ease of use scale was formed by taking the five positive items and adding a sixth positive ease of use item formed by reversing one of the existing negative items. The item which read: "The electronic mail system is rigid and inflexible to interact with" was translated into: "I find the electronic mail system to be flexible to interact with." Correspondingly, the usefulness scale was reduced in length from 10 to 6 items for the experiment. To select six of the original ten items, item analysis was performed by examining

corrected item-total correlations which were calculated by removing the item from the rest of the items in the scale prior to computing the correlation. The items having the five highest correlations were selected and combined with the "overall usefulness" item (# 10) for the final usefulness scale. These were items 3, 5, 6, 8, and 9 (with corrected item-total correlations of 87.5, 93.0, 93.0, 93.0, and 87.0 respectively). The Spearman-Brown formula estimates that this should yield a reliability of .94 for the revised usefulness scale. In order to ensure that the reduced scales still represented the appropriate domains of content, the clusters corresponding to the selected items, found via pretesting above, were examined. The 6-item ease of use scale contained 2 items associated with cluster A, 1 with cluster B, and 2 with cluster C. Neither of the 2 non-included B items could have been converted easily from negative to positive in wording. For the usefulness scale, 2 items corresponded to cluster A, 2 to cluster B and 1 to cluster C. For both scales, the "overall" item was included, but was not part of the cluster analysis. Thus, the revised scales continue to span the inferred content domain substrata identified in the pretest.

6. Conclusion

The multitrait-multimethod analysis found very high levels of convergent and discriminant validity for the scales used in the present research. After eliminating the negatively worded ease of use items, there were no exceptions found to the convergent and discriminant validity criteria. All monotrait correlations were significant and high, and were all greater in value than their corresponding heterotrait correlations. This is an unusually high level of validity, and many scales are considered to be quite healthy despite minor departures from the criteria (e.g., Campbell & Fiske, 1959; Silk, 1971). We regard this as evidence that the scales are not materially invalidated by method variance, and as circumstantial evidence of content validity of the scales. In addition, the scales exhibited reliabilities in excess of .90.

Thus, according to the analyses presented above, the scales developed herein are highly valid and reliable. However, this does not imply that the scales cannot be improved in the future. Measure development is most effective when approached as an on-going process of refinement. Each time the scales are used, data bearing on their reliability and validity is generated. The quality of individual items can be assessed. The scales may be streamlined further for methodological convenience. Over time, as we learn more about the theoretical constructs these measures operationalize, the scales may need to be modified to remain in correspondence with the appropriate domains of content. Thus, a key direction for future research is to further assess the psychometric properties of these scales.

The availability of these improved scales creates the opportunity for researchers to examine theoretical models of how attitude toward using, perceived usefulness and perceived ease of use

relate to each other, to other psychological variables, to external influences such as system characteristics and communication, and to actual system usage. Clearly, the quality of one's theories depends to a great extent on the quality of the measures one employs in assessing them. To the extent that the measures used are flawed, observed relationships provide a distorted view of the underlying theoretical processes, reducing the likelihood of correct inferences. Future research should use the above scales to empirically assess MIS theories, contrast competing theories against one another, and gain a richer understanding of the determinants and consequences of user behavior.

The development of good theory often contributes to practical issues as well. Many MIS theories are applied in nature, being concerned with how various managerially controllable phenomena, such as system design characteristics, training, consulting, hardware acquisition policies, and so on influence key outcomes such as user acceptance and performance impacts. Improved measures and theories should ultimately provide the basis for new applied techniques practitioners can use to gather decision-making information. If, for example, through improved theory and measurement, MIS management were able to forecast future user acceptance of a system reasonably well, key MIS decisions such as choosing which design features to include in a given system, and whether or not even to develop a proposed system, could be made with less uncertainty about how the proposed system will be received by the users. Based on the measurements, the proposed system could be modified (or abandoned) if significant problems were indicated. Research is needed to assess the practical viability of such methodologies.

In summary, future research is recommended to (1) further validate and refine the reported measurement scales, (2) use the derived scales to test theoretical models of user behavior, and (3) test the practical application of the models and measures in applied design and evaluation contexts.

References

- Ajzen, I. & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin, 84, 888-918.
- Ajzen, I. & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall.
- Anastasi, A. (1986). Evolving concepts of test validation. Annual Review of Psychology, 37, 1-15.
- Bagozzi, R.P. (1981). Attitudes, intentions and behaviors: A test of some key hypotheses. Journal of Personality and Social Psychology, 41, 607-627.
- Bagozzi, R.P. (1983). A holistic methodology for modeling consumer response to innovation. Operations Research, 31, 128-176.
- Bailey, J.E. & Pearson, S.W. (1983). Development of a tool for measuring and analyzing computer user satisfaction. Management Science, 29, 530-545.
- Baroudi, J.J., Olson, M.H. & Ives, B. (1986). An empirical study of the impact of user involvement on system usage and information satisfaction. Communications of the ACM, 29, 232-238.
- Barrett, G.V., Thornton, C.L., & Cabe, P.A. (1968). Human factors evaluation of a computer based storage and retrieval system. Human Factors, 10, 431-436.
- Benbasat, I. & Dexter, A.S. (1986). An investigation of the effectiveness of color and graphical presentation under varying time constraints. MIS Quarterly, March, 59-84.
- Bewley, W.L., Roberts, T.L., Schoit, D., & Verplank, W.L. (1983). Human factors testing in the design of Xerox's 8010 "star" office workstation. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 72-77.
- Bohrstedt, G.W. (1970). Reliability and validity assessment in attitude measurement. In G.F. Summers (Ed.), Attitude measurement. Chicago: Rand-McNally, Ch. 3, 80-99.
- Branscomb, L.M. & Thomas, J.C. (1984). Ease of use: A system design challenge. IBM Systems Journal, 23, 224-235.
- Brose, M. & Shneiderman, B. (1978). Two experimental comparisons of relational and hierarchical database models. International Journal of Man-machine Studies, 10, 625-637.
- Butler, T.W. (1983). Computer response time and user performance. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 58-62.
- Calder, B. J. (1977). Focus groups and the nature of qualitative marketing research. Journal of Marketing Research, 14, 353-364.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Campbell, D.T., Siegman, C.R. & Rees, M.B. (1967). Direction-of-wording effects in the relationships between scales. Psychological Bulletin, 68, 293-303.
- Card, S.K., Moran, T.P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. Communications of the ACM, 23, 396-410.
- Card, S.K., Moran, T.P. & Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, NJ: Erlbaum.
- Carroll, J.M. & Carrithers, C. (1984). Training wheels in a user interface. Communications of the ACM, 27, 800-806.
- Churchill, G.A. (1979). A paradigm for developing better measures of marketing constructs. Journal of Marketing Research, 16, 64-73.

- Cook, T.D. & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.
- Cooper, J. & Croyle, R.T. (1984). Attitudes and attitude change. Annual Review of Psychology, 35, 395-426.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.
- DeSanctis, G. (1983). Expectancy theory as an explanation of voluntary use of a decision support system. Psychological Reports, 52, 247-260.
- Dickson, G.W., DeSanctis, G. & McBride, D.J. (1986). Understanding the effectiveness of computer graphics for decision support: A cumulative experimental approach. Communications of the ACM, 29, 40-47.
- Dzida, W., Herda, S. & Itzfeldt, W.D. (1978). User-perceived quality of interactive systems. IEEE Transactions on Software Engineering, SE-4, 270-276.
- Fishbein, M. & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.
- Fishbein, M. & Raven, B.H. (1962). The AB scales: An operational definition of belief and attitude. Human Relations, 15, 35-44.
- Franz, C.R. & Robey, D. (1986). Organizational context, user involvement, and the usefulness of information systems. Decision Sciences, 17, 329-356.
- Fuerst, W.L. & Cheney, P.H. (1982). Factors affecting the perceived utilization of computer-based decision support systems in the oil industry. Decision Sciences, 13, 554-569.
- Gallagher, C.A. (1974). Perceptions of the value of a management information system. Academy of Management Journal, 17, 46-55.
- Ghani, J.A. & Lusk, E.J. (1982). The impact of a change in representation and a change in the amount of information on decision performance. Human Systems Management, 3, 270-278.
- Ginzberg, M.J. (1981). Early diagnosis of MIS implementation failure: Promising results and unanswered questions. Management Science, 27, 459-478.
- Good, M. (1982). An ease of use evaluation of an integrated document processing system. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 142-147.
- Goodwin, N.C. (1987). Functionality and usability. Communications of the ACM, 30, 229-233.
- Gould, J.D. & Boies, S.J. (1984). Speech filing- An office system for principals. IBM Systems Journal, 23, 65-81.
- Gould, J.D. & Lewis, C. (1985). Designing for usability- key principles and what designers think. Communication of the ACM, 28, 300-311.
- Guthrie, A. (1973). Middle managers and MIS: An attitude survey. Journal of Economics and Business, 26, 59-66.
- Hauser, J.R. & Simmie, P. (1981). Profit maximizing perceptual positions: An integrated theory for the selection of product features and price. Management Science, 27, 33-56.
- Ives, B., Olson, M.H., & Baroudi, J.J. (1983). The measurement of user information satisfaction. Communications of the ACM, 26, 785-793.

- Jaccard, J., Weber, J., & Lundmark, J. (1975). A multitrait-multimethod analysis of four attitude assessment procedures. Journal of Experimental Social Psychology, *11*, 149-154.
- Johnson, S.C. (1967). Hierarchical clustering schemes. Psychometrika, *32*, 241-254.
- Kaiser, K. & Srinivasan, A. (1982). User-analyst differences: An empirical investigation of attitudes related to system development. Academy of Management Journal, *25*, 630-646.
- Keen, P.G.W. (1981). Value analysis: Justifying decision support systems. MIS Quarterly, *5*, 1-15.
- King, W.R. & Epstein, B.J. (1983). Assessing information system value: An experimental study. Decision Sciences, *14*, 34-45.
- Larcker, D.F. & Lessig, V.P. (1980). Perceived usefulness of information: A psychometric examination. Decision Sciences, *11*, 121-134.
- Lawler, E.E. (1973). Motivation in work and organizations. Monterrey, CA: Brooks/Cole.
- Lucas, H.C. (1975). Performance and the use of an information system. Management Science, *21*, 908-919.
- Lucas, H.C. (1978). Unsuccessful implementation: The case of a computer-based order entry system. Decision Sciences, *9*, 68-79.
- McKennell, A.C. (1974). Surveying attitude structures: A discussion of principles and procedures. Quantity and Quality, *7*, 203-293.
- Magers, C.S. (1983). An experimental evaluation of on-line help for non-programmers. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 277-281.
- Maish, A.M. (1979). A user's behavior toward his MIS. MIS Quarterly, *3*, 39-52.
- Malone, T.W. (1981). Toward a theory of intrinsically motivating instruction. Cognitive Science, *4*, 333-369.
- Mantei, M. & Haskell, N. (1983). Autobiography of a first-time discretionary microcomputer user. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 286-290.
- Miller, L.H. (1977). A study in man-machine interaction. National Computer Conference, 409-421.
- Neal, A.S. & Simons, R.N. (1983). Playback: A method for evaluating the usability of software and its documentation. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 78-82.
- Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). The measurement of meaning. Urbana, University of Illinois Press.
- Ostrom, T.M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. Journal of Experimental Social Psychology, *5*, 12-30.
- Pindyck, R.S., & Rubinfeld, D.L. (1981). Econometric models and economic forecasts. New York: McGraw-Hill.
- Poller, M.F. & Garter, S.K. (1983). A comparative study of moded and modeless text editing by experienced editor users. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 166-170.
- Poppel, H.L. (1982). Who needs the office of the future? Harvard Business Review, November-December, 146-155.
- Porter, L.W. & Lawler, E.E. (1968). Managerial attitudes and performance. Homewood, IL: Dorsey.

- Rice, R.E. (1980). The impact of computer mediated organizational and interpersonal communication. In Williams, M. (Ed.) Annual Review of Information Science and Technology, 15, NY: Knowledge Industries, 221-249.
- Roberts, T.L. & Moran, T.P. (1983). The evaluation of text editors: Methodology and empirical results. Communications of the ACM, 26, 265-283.
- Robey, D. (1979). User attitudes and management information system use. Academy of Management Journal, 22, 527-538.
- Robey, D. & Zeller, R.L. (1978). Factors affecting the success and failure of an information system for product quality. Interfaces, 8, 70-75.
- Robinson, J.P. & Shaver, P.R. (1969). Measures of social psychological attitudes. Ann Arbor, MI: Institute for Social Research.
- Rossen, M.B. (1983). Patterns of experience in text editing. CHI '83 Human Factors in Computing Systems (Boston, December 12-15, 1983), ACM, New York, 171-175.
- Schewe, C.D. (1976). The management information system user: An exploratory behavioral analysis. Academy of Management Journal, 19, 577-590.
- Schultz, R.L. & Slevin, D.P. (1975). In Schultz, R.L. & Slevin, D.P. (Eds.) Implementing operations research/ management science. New York: American Elsevier, 153-182.
- Shaw, M.E. & Wright, J.M. (1967). Scales for the measurement of attitudes. New York: McGraw-Hill.
- Sherif, M. & Sherif, C.W. (1967). The own categories procedure in attitude research. In M. Fishbein (Ed.) Readings in attitude theory and measurement. New York: Wiley, p. 190-198.
- Silk, A.J. (1971). Response set and measurement of self-designated opinion leadership. Public Opinion Quarterly, 35, 383-397.
- Srinivasan, A. (1985). Alternative measures of system effectiveness: Associations and implications. MIS Quarterly, September, 243-253.
- Staw, B.M. (1984). Organizational behavior: A review and reformulation of the field's outcome variables. Annual Review of Psychology, 35, 627-666.
- Swanson, E.B. (1974). Management information systems: Appreciation and Involvement. Management Science, 21, 178-188.
- Swanson, E.B. (1982). Measuring user attitudes in MIS research: A review. OMEGA, 10, 157-165.
- Thomas, J.C. & Gould, J.D. (1975). A psychological study of Query by Example. Proceedings of the National Computer Conference. Montvale, NJ: AFIPS Press, 439-445.
- Triandis, H.C. (1977). Interpersonal behavior. Monterey, CA: Brooks/Cole.
- Vertinsky, I., Barth, R.T. & Mitchell, V.F. (1975). A study of OR/MS implementation as a social change process. In R.L. Schultz & D.P. Slevin (Eds.) Implementing operations research/management science. New York: American Elsevier, 253-272.
- Vroom, V.H. (1964). Work and motivation. New York: Wiley.
- Wicker, A.W. (1969). Attitudes vs. actions: The relationship of verbal and overt behavioral responses to attitude objects. Journal of Social Issues, 25, 41-78..
- Zmud, R.W. (1978). An empirical investigation of the dimensionality of the concept of information. Decision Sciences, 9, 187-195.

Appendices

TableA.1. Articles Used for Usefulness and Ease of Use Item Generation

#	Author (s) & Year	#	Author (s) & Year
1	Bailey & Pearson, 1983	21	Lucas, 1978
2	Barrett, et al., 1968	22	Magers, 1983
3	Bewley, et al., 1983	23	Maish, 1979
4	Brosey & Shneiderman, 1978	24	Malone, 1981
5	Butler, 1983	25	Mantei & Haskell, 1983
6	Card, et al., 1980	26	Miller, 1977
7	Carroll & Carrithers, 1984	27	Neal & Simons, 1984
8	DeSanctis, 1983	28	Poller & Garter, 1983
9	Dzida, et al., 1978	29	Poppel, 1982
10	Fuerst & Cheney, 1982	30	Rice, 1980
11	Gallagher, 1974	31	Roberts & Moran, 1983
12	Ginzberg, 1981	32	Rossen, 1983
13	Good, 1982	33	Schewe, 1976
14	Gould & Boies, 1984	34	Smith, et al., 1982
15	Gould & Lewis, 1985	35	Schultz & Slevin, 1975
16	Guthrie, 1973	36	Swanson, 1974
17	Kaiser & Srinivasan, 1982	37	Zmud, 1978
18	Keen, 1981		
19	King & Epstein, 1983		
20	Larcker & Lessig, 1980		

Table A.2. Correspondence Between Usefulness and Ease of Use Scale Items and Articles from Which They Were Generated

Item #	Article # from Table A.1.	
	Usefulness	Ease of Use
1	11, 20	9, 11, 22
2	23, 30, 35	3, 4, 6, 7, 9
3	1, 8, 12, 35	7, 22, 27
4	16, 17, 35	1, 9, 22
5	18, 29, 30	23, 32
6	18, 30	1, 9, 22, 23
7	11, 20, 35	1, 9, 22, 23
8	14, 18, 29	1, 9, 22
9	21, 35	9, 11, 24, 25
10	10, 23	6, 31
11	8, 21, 18	3, 11, 19, 36
12	23	3, 9, 28
13	35	9, 22, 34
14	1, 11, 15, 23, 35, 37	1, 2, 15, 20, 23, 26

Table A.3 Ranking of Item Meaning for Perceived Usefulness:
Frequency by Item

Item #	Ranked Correspondence with Construct Meaning (1 = highest)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1			2		2			1	1	1	3	4
2	1		1			2	2	3		3	2		1
3	4	1	4	1	1				1	1		2	
4		1	1	2	1		1		1		1	5	2
5	1			1	1	1	1	1	2	3	2	2	
6				3	2	3	2	2			1	1	1
7	3	2		1	2		1	1		1	2		2
8			2	1	1	2	3	4			2		
9						2	1	1	5		2		4
10	2	3	4	1		1		1	1	1	1		
11	2	4	1	1	3		1		1		1	1	
12	2	1		1	2	1	3	1	1	2	1		
13		3	1	1	2	1		1	2	2		1	1

Table A.4 Ranking of Item Meaning for Perceived Ease of Use:
Frequency by Item

Item #	Ranked Correspondence with Construct Meaning (1 = highest)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1			1	1	1	2	5	1		1		3	
2				1			1	3	3	1	3	2	1
3	2	1	2	3	2				1	1	1		2
4	2				2	2	1	1	1	2	1	2	1
5	1	2	1		2	2	2				4		1
6			2	1	1	2		1	2	2	1	1	2
7		2	2	2	1	1	1	1		1	1	1	2
8	6	2	1	2	1		1			1		1	
9				1	1	2	1	3	1	2		3	1
10	1	2	2	2	3	1	1	2		1			
11	2	2	1	1	1	3		1	1				3
12		4	1		1	1	1		3	2	2		
13	1		2	1			1	1	2	2	2	2	1

Table A.5. Similarity Matrix for Perceived Usefulness Items: Frequency With Which Items Were Assigned to Same Category

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	1												
3	3	3											
4	7	5	1										
5	2	0	0	0									
6	1	0	0	0	13								
7	7	4	2	11	0	0							
8	1	1	2	2	7	9	0						
9	1	1	2	0	9	11	0	8					
10	1	2	11	0	1	1	2	1	0				
11	1	3	11	0	0	0	2	0	1	11			
12	0	0	3	0	6	8	0	11	7	3	2		
13	7	4	2	2	3	2	2	2	1	4	2	1	

Table A.6 Similarity Matrix of Perceived Ease of Use Items: Frequency With Which Items Were Assigned to Same Category

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2	3												
3	7	3											
4	3	3	2										
5	9	3	5	2									
6	0	6	1	1	1								
7	2	2	5	1	2	1							
8	2	1	2	1	1	2	7						
9	2	5	3	2	1	2	4	2					
10	3	2	4	0	4	2	9	5	5				
11	8	1	4	2	6	3	1	6	3	2			
12	1	1	0	8	4	3	1	4	1	1	6		
13	4	1	1	8	3	1	0	2	1	2	3	6	

Table A.7 Determination of Item Priorities for Perceived Usefulness and Ease of Use

ITEM	PERCEIVED USEFULNESS			PERCEIVED EASE OF USE		
	Median Rank	Mean Rank	Priority	Median Rank	Mean Rank	Priority
1	12	9.3	13	7	7.5	7
2	8	7.9	9	10	9.5	13
3	3	4.7	2	5	5.9	3
4	12	8.7	12	9	7.7	9
5	10	8.3	11	7	6.7	5
6	7	7.1	7	9	8.1	10
7	7	6.3	5	7	6.9	6
8	7	6.8	6	3	3.7	1
9	9	9.7	10	9	8.7	11
10	3	4.5	1	5	4.9	2
11	5	4.7	3	6	6.1	4
12	7	5.7	4	9	6.5	8
13	8	6.7	8	10	8.2	12

Table A.8. Multitrait-multimethod Matrix of Item Intercorrelations - Attitude Toward Using

ITEM	Attitude Toward Using- Electronic Mail (MAU)					Attitude Toward Using - Xedit (XAU)				
	1	2	3	4	5	1	2	3	4	5
MAU1										
MAU2	.72									
MAU3	.70	.72								
MAU4	.62	.75	.75							
MAU5	.57	.78	.71	.82						
XAU1	-.10	-.04	.14	-.01	.11					
XAU2	-.04	-.01	.15	.03	.13	.85				
XAU3	-.01	.02	.15	.02	.13	.84	.95			
XAU4	.02	.05	.18	.08	.16	.80	.92	.94		
XAU5	.00	.02	.21	.04	.15	.84	.94	.95	.96	

Table A.9 Multitrait-multimethod Matrix of Item Intercorrelations- Perceived Usefulness

ITEM	Perceived Usefulness- Electronic Mail (MUF)										Perceived Usefulness- Xedit (XUF)												
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10			
MUF1																							
MUF2	.77																						
MUF3	.68	.60																					
MUF4	.57	.59	.54																				
MUF5	.70	.62	.80	.67																			
MUF6	.69	.77	.69	.65	.76																		
MUF7	.68	.65	.77	.67	.82	.76																	
MUF8	.73	.73	.76	.76	.81	.79	.87																
MUF9	.77	.71	.75	.72	.79	.76	.85	.87															
MUF10	.65	.58	.68	.62	.80	.62	.71	.77	.74														
XUF1	.24	.31	.26	.25	.18	.34	.34	.38	.38	.23													
XUF2	.18	.27	.15	.21	.12	.32	.26	.33	.29	.20	.85												
XUF3	.09	.17	.05	.22	.00	.20	.15	.23	.20	.11	.85	.90											
XUF4	-.07	.00	-.03	.24	-.05	.13	.12	.13	.07	.01	.68	.65	.73										
XUF5	.08	.20	.12	.23	.07	.27	.24	.29	.26	.17	.85	.86	.90	.80									
XUF6	.07	.18	.05	.27	.04	.25	.23	.29	.21	.12	.82	.84	.86	.80	.92								
XUF7	-.07	.05	.00	.19	-.04	.12	.11	.17	.11	.00	.67	.71	.80	.74	.85	.86							
XUF8	.08	.16	.04	.23	.03	.24	.22	.27	.22	.14	.77	.83	.86	.75	.89	.93	.85						
XUF9	.04	.13	.03	.01	-.03	.05	.02	.09	.09	.13	.73	.80	.83	.60	.82	.79	.76	.80					
XUF10	.00	.10	.05	.08	.00	.16	.17	.16	.16	.13	.80	.79	.83	.76	.88	.86	.79	.83	.86				

Table A.10. Multitrait-multimethod Matrix of Item Intercorrelations - Perceived Ease of Use

ITEM	Perceived Ease of Use - Electronic Mail (MEU)										Perceived Ease of Use - Xedit (XEU)									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
MEU1																				
MEU2	.14																			
MEU3	.35	.23																		
MEU4	.43	.51	.25																	
MEU5	.38	.06	.49	.29																
MEU6	.26	.47	.19	.57	.09															
MEU7	.24	.22	.43	.30	.32	.27														
MEU8	.41	.57	.35	.77	.29	.68	.34													
MEU9	.37	.32	.42	.30	.30	.30	.48	.35												
MEU10	.40	.61	.32	.75	.25	.62	.35	.78	.47											
XEU1	.19	.12	.12	.21	.11	.15	.19	.16	.09	.21										
XEU2	.08	.19	.06	.15	.11	.02	.13	.01	.12	-.03	.51									
XEU3	.10	.11	.17	.20	.11	.07	.25	.15	.18	.20	.73	.40								
XEU4	.24	.10	.05	.29	.01	.14	.14	.18	.07	.14	.78	.62	.67							
XEU5	.15	.06	.13	.18	.33	.04	.27	.04	.31	.10	.63	.46	.63	.55						
XEU6	.06	.20	.11	.11	.09	.10	.04	.05	.08	.01	.57	.69	.44	.67	.42					
XEU7	.23	.10	.16	.22	.15	.10	.30	.15	.17	.15	.65	.37	.68	.59	.72	.42				
XEU8	.10	.11	.06	.09	.10	-.01	.06	-.04	.07	-.04	.64	.78	.52	.73	.54	.82	.46			
XEU9	.09	.02	.09	.09	-.10	.20	.07	.07	.15	.09	.63	.37	.57	.52	.52	.44	.64	.42		
XEU10	.14	.12	.07	.06	.16	-.05	.05	-.05	.11	.01	.70	.68	.66	.77	.58	.78	.50	.84	.49	

Table A.11. Multitrait-multimethod Correlations by Construct, Type and Size:

Correlation Size	Construct													
	Attitude Toward Using				Perceived Usefulness				Perceived Ease of Use					
	Same Trait/ Diff. Meth.		Different Trait		Same Trait/ Diff. Meth.		Different Trait		Same Trait/ Diff. Method		Different Trait		Same Meth.	
	Elec. Mail	Xedit	Same Meth.	Diff. Meth.	Elec. Mail	Xedit	Same Meth.	Diff. Meth.	Elec. Mail	Xedit	Elec. Mail	Xedit	Same Meth.	Diff. Meth.
-.20 to -.11														1
-.10 to -.01			2	4				6					1	5
.00 to .09			1	8			3	25	2				1	32
.10 to .19			2	7			2	27	2				5	40
.20 to .29				1			5	25	9				1	11
.30 to .39								7	14				2	1
.40 to .49									9					
.50 to .59	1								4					
.60 to .69	1								14	4				
.70 to .79	7								20	11				
.80 to .89	1	4							7	26				
.90 to .99		6								4				
# Correlat'ns	10	10	5	20	45	45	10	90	45	45	45	10	45	90

Table A.12. Multitrait-multimethod Correlations for Positive and Negative Ease of Use Items

Correlation Size	Positive Items						Negative Items						
	Monotrait-heteromethod			Heterotrait			Monotrait-heteromethod			Heterotrait			
	E-mail	XEDIT	Hetero method	Mono method	Hetero method	E-mail	XEDIT	Mono method	Hetero method	E-mail	XEDIT	Mono method	Hetero method
-.20 to -.11													
-.10 to -.01			5	1	5								1
.00 to .09			6	1	6								4
.10 to .19			8	2	8						3		11
.20 to .29			1	1	1				1				4
.30 to .39									5		2		
.40 to .49	1								4				
.50 to .59	3									2			
.60 to .69	3	4								6			
.70 to .79	3	4								2			
.80 to .89		2											
.90 to .99													
# Correlat'ns	10	10	20	5	20	10	10	20	10	10	5	20	20

Overall Evaluation of Electronic Mail

All things considered, my using electronic mail in my job is:
(place X mark on each of the five scales)

Neutral

1. Good | ___ | ___ | ___ | ___ | ___ | ___ | ___ | Bad
2. Wise | ___ | ___ | ___ | ___ | ___ | ___ | ___ | Foolish
3. Favourable | ___ | ___ | ___ | ___ | ___ | ___ | ___ | Unfavourable
4. Beneficial | ___ | ___ | ___ | ___ | ___ | ___ | ___ | Harmful
5. Positive | ___ | ___ | ___ | ___ | ___ | ___ | ___ | Negative

Perceived Usefulness of Electronic Mail

	Strongly Agree		Neutral			Strongly Disagree	
	1	2	3	4	5	6	7
1. Using electronic mail improves the quality of the work I do.	1	2	3	4	5	6	7
2. Using electronic mail gives me greater control over my work.	1	2	3	4	5	6	7
3. Electronic mail enables me to accomplish tasks more quickly.	1	2	3	4	5	6	7
4. Electronic mail supports critical aspects of my job.	1	2	3	4	5	6	7
5. Using electronic mail increases my productivity.	1	2	3	4	5	6	7
6. Using electronic mail improves my job performance.	1	2	3	4	5	6	7
7. Using electronic mail allows me to accomplish more work than would otherwise be possible.	1	2	3	4	5	6	7
8. Using electronic mail enhances my effectiveness on the job.	1	2	3	4	5	6	7
9. Using electronic mail makes it easier to do my job.	1	2	3	4	5	6	7
10. Overall, I find the electronic mail system useful in my job.	1	2	3	4	5	6	7

Perceived Ease of Use of Electronic Mail

	Strongly Agree		Neutral			Strongly Disagree	
	1	2	3	4	5	6	7
1. I find the electronic mail system cumbersome to use.	1	2	3	4	5	6	7
2. Learning to operate the electronic mail system is easy for me.	1	2	3	4	5	6	7
3. Interacting with the electronic mail system is often frustrating.	1	2	3	4	5	6	7
4. I find it easy to get the electronic mail system to do what I want it to do.	1	2	3	4	5	6	7
5. The electronic mail system is rigid and inflexible to interact with.	1	2	3	4	5	6	7
6. It is easy for me to remember how to perform tasks using the electronic mail system.	1	2	3	4	5	6	7
7. Interacting with the electronic mail system requires a lot of mental effort.	1	2	3	4	5	6	7
8. My interaction with the electronic mail system is clear and understandable.	1	2	3	4	5	6	7
9. I find it takes a lot of effort to become skillful at using electronic mail.	1	2	3	4	5	6	7
10. Overall, I find the electronic mail system easy to use.	1	2	3	4	5	6	7