Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation

Gary C. Moore
Faculty of Management
The University of Calgary
Calgary, Alberta, Canada T2N 1N4

Izak Benbasat
Faculty of Commerce and Business Administration
University of British Columbia
Vancouver, British Columbia, Canada V6T 1Y8

This paper reports on the development of an instrument designed to measure the various perceptions that an individual may have of adopting an information technology (IT) innovation. This instrument is intended to be a tool for the study of the initial adoption and eventual diffusion of IT innovations within organizations. While the adoption of information technologies by individuals and organizations has been an area of substantial research interest since the early days of computerization, research efforts to date have led to mixed and inconclusive outcomes. The lack of a theoretical foundation for such research and inadequate definition and measurement of constructs have been identified as major causes for such outcomes. In a recent study examining the diffusion of new end-user IT, we decided to focus on measuring the potential adopters' perceptions of the technology. Measuring such perceptions has been termed a "classic issue" in the innovation diffusion literature, and a key to integrating the various findings of diffusion research. The perceptions of adopting were initially based on the five characteristics of innovations derived by Rogers (1983) from the diffusion of innovations literature, plus two developed specifically within this study. Of the existing scales for measuring these characteristics, very few had the requisite levels of validity and reliability. For this study, both newly created and existing items were placed in a common pool and subjected to four rounds of sorting by judges to establish which items should be in the various scales. The objective was to verify the convergent and discriminant validity of the scales by examining how the items were sorted into various construct categories. Analysis of inter-judge agreement about item placement identified both bad items as well as weaknesses in some of the constructs' original definitions. These were subsequently redefined. Scales for the resulting constructs were subjected to three separate field tests. Following the final test, the scales all demonstrated acceptable levels of reliability. Their validity was further checked using factor analysis, as well as conducting discriminant analysis comparing responses between adopters and nonadopters of the innovation. The result is a parsimonious, 38-item instrument comprising eight scales which provides a useful tool for

1047-7047/91/0203/0192/$01.25

the study of the initial adoption and diffusion of innovations. A short, 25-item, version of the instrument is also suggested.

Instrument development—Innovation diffusion—Information technology adoption—Research methodology

1. Introduction

This paper describes the development of an instrument designed to measure users' perceptions of adopting an information technology (IT) innovation. This instrument is intended to be a tool for the study of the initial adoption of IT by individuals in organizations, and the technology's diffusion within the organization. The adoption of information technologies by individuals and organizations is part of the process of information systems (IS) implementation, a research area that has received substantial attention during the last 25 years (Kwon and Zmud 1987). Understanding how to implement IT successfully, however, is still one of the challenging issues facing the IS field (Swanson 1988; Brancheau and Wetherbe 1990), and the research efforts to date have been mixed and inconclusive. The lack of a theoretical base and the wide array of measures used by implementation researchers without adequate theoretical and psychometric justification have been identified as major causes of this incomplete state of knowledge in IS implementation (Davis et al. 1989; Kwon and Zmud 1987).

It is worth noting that poor theory development (for a discussion see Dickson et al. 1980, and Keen 1980) and inadequate measurement of constructs have plagued IS research in a wide variety of topics. Among the authors who have noted the problems with poor operationalization of constructs are: Taylor and Benbasat (1980) and Huber (1983) for cognitive styles, Ives and Olson (1984) for user involvement, Jarvenpaa et al. (1985) for the study of information presentation, Treacy (1986) for IT and competitive advantage, and Benbasat (1990) for laboratory studies in IS in general. As Keen (1980) has argued, the lack of a cumulative tradition in IS is one of the serious issues facing the field. Yet, without a shared set of definitions, topics and concepts, there can be no cumulative tradition (Keen 1980). It is also clear that well-defined constructs are based on theory, and the operationalization of these constructs through measures with high degrees of validity and reliability is a prerequisite for the beginning of a cumulative tradition. It is an encouraging sign that some of the leading journals in IS have within the last few years published papers that deal with instrument development (e.g., Bailey and Pearson 1983; Ives et al. 1983; Doll and Torkzadeh 1988; and Davis 1989).

Recently, researchers in IS have begun to rely on the theories of innovation diffusion to study implementation problems (Alexander 1989; Brancheau and Wetherbe 1990; Moore 1987; Johnson and Rice 1987). A major focus in these studies has been how potential users' **perceptions** of the information technology innovation influence its adoption.

Measuring potential adopters' perceptions of innovations has been termed a "classic issue in the innovation literature" and a "potential key" for integrating various findings within diffusion research (Tornatzky and Klein 1982). One of the most often cited reviews of the perceived characteristics literature is that of Rogers (1983), who, in a survey of several thousand innovation studies, identified five characteristics of an innovation which affect the rate of diffusion of an innovation. They are relative

advantage, compatibility, complexity, observability, and trialability. Therefore, in a recent study examining the diffusion of new information technology, we proposed to test this set of characteristics in the context of the adoption of Personal Work Stations (PWS) by individuals.

In spite of the importance of perceived characteristics in diffusion research, a literature search indicated that most existing instruments designed to tap these characteristics lacked reliability and validity. Furthermore, no comprehensive instrument to measure the variety of perceptions of innovations existed. Such an instrument should be vital to diffusion researchers, and thus the research described in this paper was undertaken to develop one. Understanding the development history of an instrument is important in assessing its validity, and therefore this paper describes in some detail the various steps undertaken in instrument development. The short-run focus of this research was on developing scales to measure the various perceptions of using the PWS in particular. It was our intention, however, that any scales developed should also be generally applicable to a wide variety of innovations, especially other types of information technologies. Consequently, items applicable **only** to using PWS were excluded from the scales. The resulting instrument is therefore general enough to be used, with slight modifications, in most diffusion studies. It should also be noted that while some researchers are investigating organizational level innovations and individual level responses to them (e.g., Leonard-Barton 1988; Leonard-Barton and Deschamps 1988), the intent of this research is to tap respondents' reactions in an "initial adoption" environment where the individual adoption decision is voluntary. As discussed below, we felt that the issue of compulsory versus voluntary adoption was significant, and we therefore also undertook development of a scale to directly measure this factor.

2. The Perceived Characteristics of Using an Innovation

The main constructs of interest in this research were the various **perceived characteristics of using an innovation.** This section describes why the focus is on perceptions, how the initial set of characteristics were chosen, and why the perceptions of **using** the innovation rather than perceptions of the innovation itself are of interest.

The reason for focusing on the **perceived** characteristics of innovations is that the findings of many studies which have examined the **primary** characteristics of innovations have been inconsistent. According to Downs and Mohr (1976, p. 702), the focus on the primary attributes of innovations has created this inconsistency. Primary attributes are intrinsic to an innovation independent of their perception by potential adopters. The behaviour of individuals, however, is predicated by how they **perceive** these primary attributes. Because different adopters might perceive primary characteristics in different ways, their eventual behaviours might differ. This is the root of the problem of using primary characteristics as research variables. Furthermore, studying the interaction among the **perceived attributes** of innovations helps the establishment of a general theory. As an example of the effect of this division between primary and secondary characteristics, consider the attribute of **cost** and its influence on buying behaviour. Actual **cost price** is a primary attribute, whereas the **perception** of cost is a secondary attribute. In the latter case, the adopter could consider price relative to disposable income, and thus, what might appear "costly" to one potential adopter, could be "inexpensive" to another, depending on their relative levels of

income. It is argued that it is **relative** cost which has the greatest effect on buying behaviour.

In determining what characteristics to examine, we relied primarily on the extensive work of Rogers (1983). He identified five general attributes of innovations that a variety of diffusion studies had shown to consistently influence adoption. He defined them as follows:

Relative Advantage: the degree to which an innovation is perceived as being better than its precursor;

Compatibility: the degree to which an innovation is perceived as being consistent with the existing values, needs, and past experiences of potential adopters;

Complexity: the degree to which an innovation is perceived as being difficult to use;

Observability: the degree to which the results of an innovation are observable to others; and

Trialability: the degree to which an innovation may be experimented with before adoption.

In a separate analysis, Tornatzky and Klein (1982) identified 10 characteristics which had been addressed the most frequently in the 105 articles they reviewed. These included the five characteristics identified above by Rogers, plus cost, communicability, divisibility, profitability, and social approval. In their discussion, Tornatzky and Klein noted that communicability was closely related to observability, and divisibility to trialability. In addition, in studying the **individual level** adoption of PWS within organizations, cost and profitability did not seem to be appropriate characteristics to study. These are issues most likely addressed at the organization level when the decision is taken to make the PWS available. Therefore, it was decided not to consider these additional characteristics in this project.

Two further constructs were identified beyond Rogers' classification which were thought important in the decision to adopt an innovation. The first was **Image,** defined as "the degree to which use of an innovation is perceived to enhance one's image or status in one's social system". Some researchers, including Rogers, included **Image** as an aspect of **Relative Advantage.** Nevertheless, Rogers also argued that "undoubtedly one of the most important motivations for almost any individual to adopt an innovation is the desire to gain social status" (1983, p. 215). Furthermore, as discussed by Tornatzky and Klein (1982), some researchers have found the effect of **Image** (social approval) to be different enough from **Relative Advantage** to be considered a separate factor (e.g., Holloway 1977). Indeed some researchers have been examining the effects of "structural equivalence," or their perceptions of the action proper for an occupant of their position in the social structure (Burt 1987). For these reasons, it was decided to also develop a scale to measure the image enhancing effects (assumed to be a positive effect) of PWS usage.

In addition to the perceived characteristics, another construct deemed to be necessary for this particular study was **voluntariness of use,** defined as "the degree to which use of the innovation is perceived as being voluntary, or of free will". When examining the diffusion of innovations, consideration must be also given to whether individuals are free to implement personal adoption or rejection decisions. For example, use of a particular innovation within organizations may be either mandated or discouraged by corporate policy. Such policies take the freedom of choice of rejection or

adoption out of individuals' hands. Furthermore, common sense and experience indicate that there tend to be degrees of voluntariness with respect to behaviour in organizations. While many studies assume that they have "voluntary" adopters of innovations because adoption is not strictly mandatory, some adopters may in fact feel a degree of compulsion. As with most factors, it is often not **actual** voluntariness which will influence behaviour, but rather the **perception** of voluntariness.

It should be noted that Rogers' definitions are based on perceptions of the innovation **itself,** and not on perceptions of actually **using** the innovation. As argued by Ajzen and Fishbein (1980, p. 8), however, attitudes towards an **object** (A_o) can frequently differ from attitudes toward a particular **behaviour** (A_b) concerning that object. For example, a difference may exist between any employer's attitude towards a particular individual (the object) and his attitude towards hiring that individual (the behaviour). An employer may dislike the individual but may nevertheless believe that hiring him will bring positive results. Thus, his attitude toward **hiring** that individual will be positive.

The relevance of the above argument to the study of the diffusion of innovations is clear. Innovations diffuse because of the cumulative decisions of individuals to **adopt** them. Thus, it is not the potential adopters' perceptions of the innovation itself, but rather their perceptions of **using** the innovation that are key to whether the innovation diffuses. Unfortunately, however, within diffusion research, perceptions have traditionally been defined with respect to the innovation itself, as evidenced by Rogers' definitions. Nevertheless, the perceived characteristics of innovations can easily be recast in terms of using the innovation (Moore 1987). For example, the definition of **Relative Advantage** needs only simple rewording to be defined as "the degree to which **using** the innovation is perceived as being better than **using** its precursor". Thus, all characteristics were redefined in terms of the potential adopters' use, trial or observation of the innovation, and labelled the **Perceived Characteristics of Innovating** (PCI).

3. Existing Measurement Instruments

Prior to developing measurement instruments for the seven PCI discussed in the previous section, the literature was searched for tests or scales that were already developed, which were then evaluated in terms of their validity and reliability. Among the first to explicitly use Rogers' taxonomy was Ostlund (1969). He included Rogers' five PCI's in a study of the adoption of a variety of new consumer goods. The scales in his study, however, used only two items per perceived characteristic, with no reliability coefficients reported. Work on extending Ostlund's instrument was later carried out by Bolton (1981), who examined perceptions of videotext technology. By adding new items, he eventually had approximately five items for each PCI scale. Once again, however, the psychometric properties of the instrument remained below the desired levels. Of 18 Cronbach's Alpha (ALPHA) coefficients from three different replications of the survey, only four were above 0.80. Of these, three were for the same PCI, **Compatibility,** from the three replications. Because of the generally low reliability coefficients, it was concluded that neither Ostlund's instrument, nor Bolton's extension of it, could be used without modification.

Holloway (1977) also used Rogers' taxonomy, and, after a review of PCI studies to that date, decided to develop his own scales. Nevertheless, operationalization of the constructs in this study still appeared to be inadequate. Factor analysis revealed that

several items which ought to have loaded on different PCI's loaded within the same construct. Therefore, because of apparent construct validity problems, it was concluded that Holloway's instrument could not be used without modification.

As a check on the exhaustiveness of the search for existing validated instruments, Rogers and his colleagues were contacted directly (Rice 1987). Discussion revealed that they were aware of very few studies which had explicitly measured the five characteristics as defined by Rogers, although an additional instrument developed by Hurt and Hubbard (1987) was identified. In discussing the development of their instrument, Hurt and Hubbard argued that the measurement of the PCI to date had been "troublesome", and that the "relationships among innovation characteristics and innovativeness have not been systematically investigated" (1987, p. 2). Thus, they attempted to create an instrument based directly on Rogers' classification. Although high reliability figures were obtained for **Complexity** and **Compatibility,** with ALPHA's of 0.93 and 0.86 respectively, factor analysis of the results did not support the overall classification. First, **Trialability** and **Observability** did not emerge as separate factors. Hurt and Hubbard concluded that either the items written "did not clearly discriminate between the two theoretical attributes" or that "those two allegedly independent characteristics are treated as a single concept by potential adopters" (p. 8). Our examination of the items did not help determine which explanation was more likely. Secondly, **Relative Advantage** did not emerge as a factor at all. Again, it is not known whether this was an artifact of the instrument development process or a reflection of attitudes about the innovation being studied, microcomputers. Thus, it was concluded that this instrument also required further development.

Finally, Davis (1986) developed a Technology Acceptance Model, which was quite similar to a Diffusion of Innovations model. Davis included two constructs, "perceived usefulness" and "perceived ease of use", which he defined as follows (p. 82):

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 similarities between these constructs and Rogers' perceived **Relative Advantage** and perceived **Complexity** are clear. Furthermore, Davis also provided evidence of the relative paucity of scales to measure these two PCI. He concluded, after a search for appropriate scales, that no validated scales with the desired reliability existed for either construct. He therefore undertook an instrument development process resulting in two scales with eventual reliabilities in excess of 0.90 for each.

As a result of the analysis of existing instruments, no valid and reliable scales had been identified to measure **Observability** or **Trialability.** New scales thus had to be developed for these constructs. Furthermore, in spite of the promise of Davis' scales for **Relative Advantage** and **Ease of Use,** and of Hurt's and Hubbard's for **Compatibility,** it was concluded that they should also go through a validation process to ensure that they fit within a general instrument to measure all PCI.

In that Davis chose to call his scale "perceived usefulness," and that Tornatzky and Klein (1982) discussed *measurement* problems with the use of the term **Relative Advantage,** the question arises as to the appropriate terminology to use for this construct. Traditionally, the construct has not been well explicated nor measured, which has led to it becoming a "garbage can" into which a variety of advantages can be

tossed. While Davis' "usefulness" term might seem to be a better name for this construct, it also suffers the same problem as **relative advantage,** being rather broadly based. One's job can be enhanced in many ways by the use of IT, which are all reflected in his scale items. Furthermore, the term "perceived usefulness" was used by Larcker and Lessig (1980) to describe the attributes of *information* in the design of information systems. Thus, Davis' use of the term may be confounded with this approach.

Tornatzky and Klein state that if "relative advantage is measured in terms of profitability, or social benefits, or time saved, or hazards removed, why bother to refer to relative advantage at all?" (1982, p. 34). On the other hand, innovations typically are developed with certain purposes in mind, and they must be perceived to fulfill their intended purposes **better than** their precursors if they are to be adopted. Thus the concept of "relative advantage" has significant intuitive appeal as it is a very generalizable concept. Tornatzky's and Klein's concern about measurement issues could lead to a plethora of scales and terms, all of which would be subsumed under the idea of "relative advantage". For example, even though Davis does not use the term "relative" in his label, the *definition* of "perceived usefulness" is in relative terms. For all these reasons, given the intuitive appeal of the term **relative advantage** and its recognition across a variety of disciplines, it was decided to maintain the term for this study.

4. Instrument Development Process

Development of the instrument was carried out in three stages. The first stage was **item creation,** whose purpose was to create pools of items for each PCI by identifying items from existing scales, and by creating additional items, that appeared to fit the construct definitions. The next stage in the process was **scale development.** As will be described more fully in subsequent sections, the basic procedure was to have panels of judges sort the items from the first stage into separate categories, based on the similarities and differences among items. Based on their placement, the items could then be examined and any inappropriately worded or ambiguous items could be eliminated. The various scales were then combined into an overall instrument for the **instrument testing** stage, which included three separate steps. First, the instrument was distributed to a small sample of respondents and an analysis of the responses was conducted to get an initial indication of the scales' reliability. Items which did not contribute to the reliability of the scales were culled for the second round of testing, which was a full scale pilot test with a larger number of subjects. In the third step, the scales were further refined, and a field test of the instrument was carried out. The following sections describe each of the steps in detail.

4.A. *Stage 1: Item Creation*

The objective of this first step was to ensure **content validity.** First, all items identified in the existing instruments were categorized according to the various PCI which they were originally intended to address. This generated an initial item pool for each of the PCI. Then, items considered to be too narrow in focus and applicable only in particular situations or to particular innovations were culled. Once this was done, new items were created for those PCI categories with fewer than 10 items, or where it was felt that all dimensions of the construct had not been covered. The typical item in previous instruments tended to be a statement to which the respondent was asked to indicate a degree of agreement. This approach was retained for this study, with a

seven-point Likert scale ranging from "extremely disagree" to "extremely agree" chosen as the response format.

As well as focusing on behaviour towards the innovation, as discussed earlier, Azjen and Fishbein (1980) also point out that the various **elements of behaviour** must be delineated in order to develop an accurate indication of respondents' perceptions. In addition to **actual behaviour** (using a PWS), these elements include the **target** at which the behaviour is directed (the PWS), a **context** for the behaviour (in one's job), and a **time frame** (now and into the future). While both the target and the behaviour were defined in all items, the latter two elements did not have to be explicitly included as the general instructions for completing the eventual questionnaire made these elements implicit

Once the item pools were created, items for the various PCI were then re-evaluated to eliminate those which appeared redundant or ambiguous (i.e. which might load on more than one factor). The following items from the **Compatibility** pool are examples of potentially *ambiguous* items:

I really need a tool like the PWS.
Using a PWS would help me a lot with my work.

Each of these items also captures an aspect of **Relative Advantage.** In fact, examination of the scales used in previous research, and of factor loadings of items within them, indicated that much of the factorial complexity of **Compatibility** items may have been a result of the construct's original definition. Rogers defined **Compatibility** as "the degree to which the innovation is perceived as consistent with the existing values, past experiences, and **needs** [emphasis not in original] of the potential adopter" (1983, p. 223). The inclusion of "needs" is considered to be a source of confounding with **Relative Advantage,** as there can be no advantage to an innovation that does not reflect an adopter's needs. For that reason reference to "needs" in derivation of items to measure the **Compatibility** construct was eliminated. The culling process left the following number of items in each pool for a total of 94 items:

Voluntariness:	6	Ease of Use:	18
Image:	8	Observability:	14
Relative Advantage:	18	Trialability:	14
Compatibility:	16		

4.B. *Stage 2: Scale Development*

a. *General.* The goals of this stage were twofold: to assess the **construct validity** of the various scales being developed, and to attempt to identify any particular items which still may have been ambiguous. In order to achieve these goals, judges were asked to sort the various items into construct categories. This is similar to the technique used by Davis (1986, 1989), who was interested in assessing the **coverage** of the domain of his constructs Davis first asked judges to rank how well the items fit the construct definitions he had provided, and then asked the judges to sort items into **construct categories.** By comparing the categories developed, Davis was then able to assess the domain coverage of the particular construct. The second step in our research is similar to Davis' procedure in that definitions were provided to the judges. The major difference in our study was the inclusion of an **initial step** where the judges

were **not** told what the underlying constructs were, but were asked to provide their own labels for the constructs. It is felt that these procedures extend those of Davis', in that they attempt to verify the construct validity of the scales in a significantly different way by having judges supply their own construct labels and definitions. This procedure minimizes the potential of "interpretational confounding", which occurs "as the assignment of empirical meaning to an unobserved variable [e.g. factor] other than the meaning assigned to it by an individual a priori to estimating unknown parameters" (Burt 1976, p. 4). In this instance, we had several individuals define, a priori, the meaning of the constructs represented by the items. If these definitions matched the scale's intent, then our confidence in the construct validity of the scales increased.

A second indicator of construct validity was the convergence and divergence of items within categories. If an item was consistently placed within a particular category, then it was considered to demonstrate **convergent validity** with the related construct, and **discriminant validity** with the others. Secondly, in the sorting rounds, if the number of categories created by the various judges, the labels assigned to them, and the items included in them, were consistent, then scales based on these categories could also be said to demonstrate convergent and discriminant validity.

b. *Sorting Procedures.* Each item was printed on one 3 × 5-inch index card. The cards were then shuffled into random order for presentation to the judges. Each judge sorted the cards into categories and labelled the categories of items independently from the other judges. **In all sorting rounds, a different set of judges was used.** Each set included a secretary, administrative clerk, student and professor. This range of backgrounds was chosen to ensure that a range of perceptions would be included in the analysis.

Prior to sorting the cards, judges were read a standard set of instructions which had been previously tested with a separate judge to ensure their comprehensiveness and comprehensibility. Judges were allowed to ask as many questions as necessary to ensure they understood the procedure. Following this, a trial sort was done by each judge on ten sample items unrelated to the constructs of the study. In this case, ten statements were written about various aspects of an automobile, some of which were deliberately constructed to be ambiguous. Judges were asked to sort the "test" cards, following the instructions they had received. Any misunderstandings resulting from the instructions were clarified. Furthermore, after the judges had sorted the ten test items, and labelled the categories, they were shown how the ambiguous cards could be resorted to create different construct categories. This was done to ensure that the judges understood the idea of attempting to sort the items based on an **underlying construct** for each category, and to place items in categories which best reflected the underlying construct. It also cued the judges to be more aware of ambiguous or unclear items. In virtually all cases, the judges reported that they found the trial sort to be a very helpful exercise.

c. *Inter-Rater Reliabilities.* To assess the reliability of the sorting conducted by the judges, two different measurements were made. First, for each pair of judges in each sorting step, their **level of agreement** in categorizing items was measured using Cohen's Kappa (Cohen 1960). Once the Kappa scores were calculated, an assessment was made of the level of agreement across all pairs of judges. For Kappa, no general authority exists with respect to required scores, but recent studies have considered

scores greater than 0.65 to be acceptable (e.g. Vessey 1984; Jarvenpaa 1989; Todd and Benbasat 1989).

A second, overall, measure of both the reliability of the classification scheme and the validity of the items was developed for this research. The method required analysis of how many items were placed by the **panel** of judges for each round within the "target" construct. In other words, because each item was included in the pool explicitly to measure a particular underlying construct, a measurement was taken of the overall frequency with which all judges placed items within the intended theoretical construct. The higher the percentage of items placed in the target construct, the higher the degree of inter-judge agreement across the panel which must have occurred. Secondly, scales based on categories which have a high degree of "correct" placement of items within them can be considered to have a high degree of construct validity, with a high potential for good reliability scores. It must be emphasized that this procedure is more a qualitative analysis than a rigorous quantitative procedure. There are no established guidelines for determining "good" levels of placement, but the matrix can be used to highlight any potential problem areas. (An example of how this measure would work is provided in Appendix 1.)

d. *Results of First and Second Sorts.* Four judges were involved in each of the first two sorting rounds, which included items developed for the seven constructs. In the first round, two judges created seven categories, while the other two had six. The inter-judge raw agreement scores averaged 0.83, Kappa scores averaged 0.80 (Table 1), and the initial overall placement ratio of items within the target constructs was 78% (Appendix 1). Examination of the off-diagonal entries in the placement matrix revealed two significant clusters, which reflected those judges who had only six categories (Appendix 1). Nevertheless, this argued well for any potential internal consistency measurements because the off-diagonals showed clustering, rather than a scattering of items. Several items were also identified as either being too ambiguous (fitting in more than one category) or too indeterminate (fitting in no category), and were dropped from the item pool.

After the sorting had been completed, each judge independently labelled and defined each of their categories. They then met as a group and carried out the same task. Except for **Observability** (which had construct ambiguity problems as discussed below), both the independent judges' and the panel's labels and definitions very closely matched those of the original constructs (Appendix 2). All items remaining at this point were then passed to the next sorting round.

The four new judges in the second round were asked to sort the remaining items based on construct definitions which were provided. A "too ambiguous/doesn't fit" definition was also included to ensure that the judges did not force fit any item into a particular category. Examination of the resulting "factor structure" showed very high agreement among the judges, again with the exception of **Observability.** The raw agreement scores averaged 0.86, while Kappa averaged 0.83 (Table 1). The overall placement ratio of items within target constructs was 92%, with all constructs at or above 90%, except for **Observability** which was at 73% (Appendix 1). This indicated that items were generally being placed as they were intended. Thus, except for **Observability,** it was concluded that the development process had resulted in scales which demonstrated construct validity, with a high potential for very good reliability coefficients.

TABLE 1
Inter-Judge Agreements

Agreement Measure	Round 1	Round 2	Round 3	Round 4
Raw Agreement	0.86	0.88	0.64	0.74
	0.74	0.87	0.60	0.74
	0.83	0.85	0.82	0.78
	0.80	0.87	0.79	0.91
	0.85	0.86	0.74	0.96
	0.90	0.82	0.69	0.93
			0.94	
			0.76	
			0.78	
			0.75	
Average	0.83	0.86	0.75	0.85
Cohen's Kappa	0.84	0.86	0.58	0.70
	0.70	0.85	0.53	0.72
	0.76	0.82	0.79	0.74
	0.82	0.84	0.76	0.89
	0.89	0.83	0.70	0.96
	0.80	0.79	0.64	0.92
			0.94	
			0.70	
			0.74	
			0.71	
Average	0.80	0.83	0.71	0.82
Placement Ratios Summary				
Voluntariness	0.83	0.96	0.93	1.00
Image	0.91	1.00	0.72	0.93
Relative Advantage	0.99	0.90	0.98	0.87
Compatibility	0.66	0.91	0.53	0.98
Ease of Use	0.96	0.96	1.00	0.93
Trialability	0.68	0.96	0.83	0.84
Observability	0.43	0.73		
Result Demonstrability			0.94	0.91
Visibility			0.73	0 94
Average	0.78	0.92	0.85	0.92

Analysis of the **Observability** items indicated that the original construct was quite complex. It had been originally defined as the degree to which the **results** of an innovation are **visible** and **communicable** to others (Rogers 1983, p. 232), but it also included the idea of the innovation being **visible.** Rogers had indicated that software dominant innovations have "less observability and usually have slower rates of adoption" than hardware innovations (1983, p. 232). If results were the only focus, it would not matter whether they were related to software or hardware. Software, however, is less visible than hardware. Thus, it appears that the more a potential adopter can see an innovation, the more likely he is to adopt it. In fact, research has shown that "mere exposure [to objects] is capable of making an individual's attitude toward these objects more positive" (Zajonc and Markus 1982).

The items in the pool tapped both these dimensions, and had been drawn mainly from other studies with the expectation that their construct validity and internal cohesiveness would be high. This was not the case. As a result, it was decided to split the construct, and focus on each dimension independently. One dimension concentrated on the **tangibility** of the results of using the innovation, including their **Observability** and **Communicability,** and was labelled **Result Demonstrability.** This ties back to the term used by Zaltman et al. who indicated that the more "amenable to **demonstration** the innovation is, [and] the more visible its advantages are . . . the more likely it is to be adopted" (1973, p. 39). Items from the **Observability** pool which reflected the ability to measure, observe, and communicate the results of using the innovation were kept for this dimension, which left only seven items from the original pool. Because some items had also been placed within **Relative Advantage,** they were reworded to drop any references to "benefits" or "advantages" in order to concentrate on the "results" aspect of the construct. Two additional items were written to increase the coverage of the item pool. The other dimension was the actual **Visibility** of the PWS. An additional nine items were written for this construct and added to the overall item pool. Therefore, at the end of this step, eight scales were being developed: voluntariness, image, relative advantage, compatibility, ease of use, result demonstrability, trialability, and visibility.

e. *Results of Third Sort* The procedures of the third sort (Table 1) were a repeat of the first sorting step, this time using the item pool which had been refined after the first two sorting rounds. Again, five new judges were recruited, this time from the business faculty of a different university. The level of agreement among these judges was quite acceptable, although the range of scores was greater than in the first sorts. Furthermore, 85% of the items were placed within the target constructs, although there were some clusters outside the targets (Appendix 1). All labels supplied by the judges (Appendix 2) to the categories they had created reflected the original definition enough that no problems with construct validity could be identified. The placement percentages, however, had a wider range than in previous sorts, with problems indicated for **Image, Compatibility,** and **Visibility.** Nevertheless, because once again the items tended to be grouped together, their internal consistency remained high. Scales were refined, however, by eliminating items which had been placed fairly consistently outside the target construct.

f. *Results of Fourth Sort.* The aim of this final round was the same as for round two. Again, four new judges were recruited and given the definitions of the constructs before the sort. The resulting grouping of items showed a simple "factor structure",

with only one significant off-target cluster. In this case, one judge placed seven of the eleven **Trialability** items within the **Voluntariness** construct. Again, this is not surprising in that the perception of **Voluntariness** could also contribute to the perception of **Trialability.** The inter-judge raw agreement scores (Table 1) reflect these results. Except for the one judge with the off-target **Trialability** items, the agreement among the pairs of judges was above 0.90. Even for this judge, the agreement scores are in the 0.75 range. Kappa scores are also correspondingly high, with an average of 0.82. The three low scores in Table 1 are for the pairings which include the judge with the set of off-target items. The overall placement ratio of items within the target construct was 92%, with the lowest score for an individual construct being **Trialability** at 84% (Appendix 1). This placement of items within the target constructs shows that a high degree of construct validity and potential reliability had been achieved.

The final step of this round was to reduce the items for the various scales to approximately ten each. **Relative Advantage** had 18 items remaining after the various sorting rounds. It was therefore decided to keep the items from Davis' "perceived usefulness" scale (1986), and four items from two other sources to provide a potentially different perspective. Thus the scale for the next stage contained 14 items, reflecting the rather complex nature of the construct. Likewise, the 10 items retained for the **Ease of Use** scale were all taken directly from Davis' scale for "perceived ease of use". All items left for **Compatibility, Result Demonstrability,** and **Visibility,** and ten of the twelve items remaining for **Trialability** were kept in the item pool.

The final refinement of the scales for the initial pilot test was to reword items as necessary to include nonusers of the PWS. This created two sets of items which were grouped in two different sections of the questionnaire. The re-wording involved a slight change in tense and mood for the items, as illustrated by the italicized words in the following example:

User: Using a PWS *is not* similar to anything that I've done before.
Nonuser: Using a PWS *would not be* similar to anything that I've done before.

4.C. *Stage 3: Instrument Testing*

a. *First Pilot Test.* The next stage of the development process was an initial pilot test of the overall instrument, which had been created by randomly ordering items from the eight scales into a common group. Because this was an initial test, the sample size was kept quite small. Questionnaires were distributed to a convenient sample of 20 users and nonusers taken from the business faculties of two universities, including both voluntary users of the PWS, primarily faculty members, and nonvoluntary users, primarily secretaries.

The first aim of this test was to ensure that the mechanics of compiling the questionnaire had been adequate. This was accomplished by having respondents first complete the questionnaire, and then comment on its length, wording, and instructions. The second aim of the test was to make an initial reliability assessment of the scales. Because 81 items are too many for this type of instrument, as respondents pointed out, we wished to eliminate as many as possible while retaining desired reliability levels. The scales were analyzed using the six measures of reliability discussed by Guttman (1945). Of the six, he argued that the one with the highest rating establishes the lower bound of the true reliability of the scale (henceforth *Guttman's lower bound* or GLB). Cronbach's ALPHA (Cronbach 1970), one of Guttman's six measures, was highlighted in the analysis as is fairly standard in most discussions of

reliability. The accepted level of reliability depends on the purpose of the research project. For example, Nunnally argued that in early stages of research, reliabilities of 0.50 to 0.60 would suffice, and that "for basic research, it can be argued that increasing reliabilities beyond 0.80 is often wasteful" (1967, p. 226). Thus, for this study the target level of minimum reliability was set in the 0.70 to 0.80 range.

The correlation of items within each scale (henceforth item–item), the corrected item-to-total correlations (henceforth item–scale), the effects on ALPHA if the item were deleted, and the item standard deviation scores were used to determine which items were candidates for deletion from the scale. Items with low item–item and item–scale correlations, which would raise ALPHA if deleted, or which showed low variance (and hence would have low explanatory power in any model) were all candidates for elimination. Before any item was deleted, however, a check was made to ensure that the domain coverage (i.e. content validity) of the construct would not suffer. As a result, several items were dropped and the length of some scales significantly reduced. The remaining items are listed in Appendix 3. The following shows the original number of items in each scale, and the number after items were culled.

Scale	Original Length	Reduced Length
Voluntariness	5	4
Image	7	5
Relative Advantage	14	9
Compatibility	11	4
Ease of Use	10	8
Result Demonstrability	8	4
Trialability	11	5
Visibility	9	4
TOTAL	75	43

b. *Second Pilot Test* The next to last stage of the development process was a "full scale" pilot test of the questionnaire using respondents whose background was similar to the target population of the final study. The primary aim of this test was to ensure that the various scales demonstrated the appropriate levels of reliability. Again, checks were made for any difficulties that respondents might face in completing the questionnaire

The sample for the pilot test was drawn from the head office of a utility company. It included both users and nonusers of PWS from a variety of levels and departments. Questionnaires were distributed to 75 individuals, with 66 (88%) usable questionnaires returned. The same analysis was conducted as for the first pilot test, focusing on reliabilities and the item–item and item–scale correlations (see Table 2). As a result, no changes were made to the scales for **Image, Voluntariness, Relative Advantage, Compatibility,** and **Result Demonstrability**. For **Ease of Use** and **Trialability,** slight modifications were made to a few items to correct some complex wording. In addition, two items from **Ease of Use** and one from **Trialability** were dropped to improve ALPHA. Finally, the **Visibility** scale of four items demonstrated significantly reduced reliability from the first pilot test. It was decided to attempt to improve reliability by slightly reworking some of the items (for example by emphasizing the word "not" in one item by underlining it), and by retrieving one item which had been dropped during an earlier culling.

TABLE 2

Reliability Coefficients

	PILOT TESTS						FIELD TEST					
	FIRST (n = 20)			SECOND (n = 66)			SAMPLE #1 (n = 270)			SAMPLE #2 (n = 270)		
SCALE NAME	ITEMS	ALPHA	GLB	ITEMS	ALPHA	GLB	ITEMS	ALPHA	GLB	ITEMS	ALPHA	GLB
VOLUNTARINESS	5	0 93	0.96	4	0.87	0 88	4	0.82	0.83	4	0.87	0.86
IMAGE	7	0 71	0.89	5	0.84	0.88	5	0.79	0.80	4	0 80	0.83
RELATIVE ADVANTAGE	14	0.89	0.98	9	0.90	0.91	9	0 95	0.95	8	0 92	0 93
COMPATIBILITY	11	0.52	0.86	4	0.81	0 82	4	0 88	0.88	4	0 83	0.84
EASE OF USE	10	0.79	0.91	8	0.83	0.85	8	0.81	0.81	6	0 80	0.80
TRIALABILITY	11	0.77	0.94	5	0.72	0.73	5	0.73	0 74	5	0.71	0.72
RESULT DEMONSTRABILITY	8	0.20	0.64	4	0.72	0 74	4	0.81	0.81	3	0.77	0 78
VISIBILITY	9	0 83	0.96	4	0.37	0.46	5	0.72	0.75	4	0 73	0.81

Notes ITEMS Number of items in each scale; ALPHA: Cronbach's Alpha reliability coefficient, GLB: Guttman's Lower Bound for reliability

c. *Final Field Test.* For the final field test, 800 questionnaires were distributed to individuals in seven companies from a variety of industries, including two utilities, two resource-based companies, two government departments, and a natural grains pool. 540 usable questionnaires were returned for a response rate of 68%. The rate was fairly consistent across all organizations. Respondents were from a variety of departments, and represented all organization levels, including executive and middle management (21%), first-line supervisors (25%), nonmanagement professionals (24%), and technical and clerical staff (30%).

Analysis was conducted in two rounds. First, the sample was randomly split into two. One half was used to determine whether the scales could be further refined, with the second half held in reserve for testing any revisions. While previous testing concentrated on reliability measures, this test also included factor analysis of the data (Appendix 3 contains all items sorted by construct for ease of reference). As shown in Table 2, the reliabilities of the majority of the scales for the first half of the data are at or above the 0.80 level, with only two, **Trialability** and **Visibility,** near the 0.70 lower bound set for the study. These results indicate that amendments to the scales following the **second pilot test** succeeded in raising the reliabilities of the problematic scales to the desired levels, although there was still room for improvement. For example, by deleting the item indicated in Appendix 3, the ALPHA score for **Visibility** could be improved to 0.75. This item concerns the visibility of PWS **outside** the organization and thus apparently taps a different dimension.

Factor analysis was used as another assessment of construct validity. Fornell (1983) has argued that, in traditional factor analysis, the results are "indeterminate" because factor loadings can be rotated in numerous ways. Thus, data analysis where possible ought to be grounded in strong **a priori** notions. This fits the approach in this research where the constructs of interest are based on a substantial body of prior research (Rogers 1983; Tornatzky and Klein 1982) and have been explicated prior to any item

development. This also follows several prescriptions that items be **developed** to fit the construct's conceptual meaning as a method of ensuring construct validity (see, for example, Bohrnstedt 1983; Campbell and Fiske 1959; Kerlinger 1978).

Principal Components analysis was conducted with VARIMAX rotation. The results indicate that a seven-factor solution was the most likely as seven factors had eigenvalues greater than 1.0, while the scree plot also showed a break after the seventh factor. The seven factors accounted for approximately 63% of the variance in the data set. All expected factors emerged fairly "cleanly" except for **Compatibility,** whose items all loaded with the **Relative Advantage** items as one factor (the implications of which are discussed below). This was verified by confirmatory factor analysis on this data (using **ITAN,** Gerbing and Hunter 1988). When an eight-factor solution was investigated, **Relative Advantage** and **Compatibility** were correlated at the 0.99 level (i.e., were one factor).

Next, the rotated factor matrix was examined for items which either did not load strongly on any factor (<0.40), or were too complex (which loaded highly or relatively equally on more than one factor). As a result, five items, identified as being too complex, were candidates for possible deletion from the scales. These included the **Visibility** item mentioned above, as well as two **Ease of Use** items, and one item from each of the **Relative Advantage, Image,** and **Result Demonstrability** scales (all indicated in Appendix 3). These results were again confirmed using ITAN. Thus, these items were dropped from the scales (leaving a 38-item instrument), and a second factor analysis was conducted using the portion of the sample which had been withheld.

This analysis again used Principal Components with VARIMAX rotation, specifying a seven-factor solution. While the first analysis was exploratory (except for the ITAN confirmation), this analysis was more confirmatory as the solution was restricted to seven factors, although all parameters were freely estimated. The results show that the seven factors again captured 63% of the variance (Appendix 4 indicates the amount of variance captured by each factor). As can be seen from the factor pattern matrix in Appendix 4, a fairly simple factor structure emerged. No item loads highly on more than one factor. Furthermore, all items remaining in the various scales loaded together on the "target" factor, with the lowest loading being 0.45. Comrey (1973) indicated that loadings in excess of 0.45 could be considered fair, greater than 0.55 good, 0.63 very good, and 0.71 excellent. As the factor pattern shows, most loadings on the target factors are in the excellent range (25 out of 38), with only four in the "fair" range. Finally, as is shown in Table 2, all scales achieved the minimum reliability scores targeted for this study, with only **Trialability** near the lower bound (GLB = 0.72). The reliability score for **Result Demonstrability** (with one less item) also slipped slightly (GLB = 0.78). These results indicate that the various scales (except for **Compatibility**) achieved a high degree of unidimensionality.

The scales may obviously still be improved. For example, while **Trialability** had a relatively low reliability score (0.72), no particular item seemed to be pulling it down, as deletion of any one would lower ALPHA. This scale had received higher reliability scores in the initial test (GLB = 0.94), which led to the deletion of several items. The final results indicate that perhaps too many items were culled and thus Appendix 3 indicates items which are suggested for re-insertion in any future **Trialability** scale. The inter-item correlations for all scales are included in Appendix 5 and can be used to guide any further deletions of items from the scales.

The main area of concern for the final instrument is the fact that **Relative Advantage** and **Compatibility** did not emerge as separate factors, even though they had been consistently separated throughout the sorting procedures. This may mean that, while conceptually different, they are being viewed identically by respondents, or that there is a causal relationship between the two. For example, it is unlikely that respondents would perceive the various advantages of using the PWS, if its use were in fact not compatible with the respondents' experience or work style. Nevertheless, as pointed out by Bollen and Hoyle (1990), one must distinguish "conceptual from empirical dimensionality . . . High or even perfect correlation is not a sufficient condition to claim that a concept is unidimensional rather than bidimensional" (p. 497). The fact that the scales were separated in the card sorting indicates that the sorters clearly distinguished the conceptual differences between the two constructs. The empirical relationship between these two constructs is therefore definitely an area requiring further investigation.

The one scale which is not explicitly a PCI is **Perceived Voluntariness.** It was developed because it was initially assumed that this construct is not binary, or in other words, that voluntariness is not an "either-or" perception. The results support this assumption. The mean response to the scale was 3.7, with 4.0 being a neutral response. The scores were also distributed normally (skewness was 0.04).

d. *Adopters versus Nonadopters.* Diffusion theory specifies that adopters should have more positive perceptions of using the PWS than nonadopters, and thus should score higher on the scales developed. Therefore, as a final check on the validity of the instruments, the responses to the scales for a split sample of **adopters** versus **nonadopters** were compared. Adopters were determined by asking respondents to indicate whether they currently used a personal work station (PWS). The PWS was defined in the instructions accompanying the questionnaire as a set of computerized tools designed for an individual. It was indicated that the PWS "usually consists of a personal or microcomputer with one or more software packages, such as a word processing program or a spreadsheet, or a computer terminal hooked up to a central mainframe computer, again with the appropriate software". It was further emphasized that the "key aspect of a PWS is that it is computer technology directly that [one] would use directly". Pilot tests of the instructions indicated that respondents had no difficulty with this definition.

To construct the scales, the means of the items in each scale were used rather than their factor scores. Factor scores may be viewed as being more exact, in that the relative weight of an item in a scale is based on its loading on the factor. In some instances, this approach may be preferable as these scores may perform better than other approaches (Lastovicka and Thamodaran 1991). Nevertheless, such scores are often less interpretable and generalizable than using simpler approaches such as summing or averaging the items that load highly on the factor (Gorsuch 1988). For many research purposes, this simpler approach is considered "entirely adequate" (Tabachnik and Fidell 1989), and hence was followed in this study. Once the scales for the PCI were constructed, the scores of adopters versus nonadopters were compared using the nonparametric Mann-Whitney U-test. As can be seen from Table 3, the differences between the two groups are significant for all variables. These results fit the theory, providing further evidence of the validity of the scales.

TABLE 3
Variable Means for Adopters Versus Nonadopters

Perceived Characteristics	Adopters (n = 418)	Nonadopters (n = 122)	U-Test Z-Scores	Significance
Voluntariness	3.3	5.0	−10.01	0.0000
Image	4.3	3.9	3.08	0.0021
Relative Advantage	5.9	4.8	−9.39	0.0000
Compatibility	5.7	4.4	−9.68	0.0000
Ease of Use	5.1	4.5	−5.54	0.0000
Trialability	4.6	4.3	−2.23	0.0257
Result Demonstrability	5.8	4.9	7.81	0.0000
Visibility	5.9	5.3	5.13	0.0000

Finally, while the relative effects of the different perceptions was not an issue in this study, a direct discriminant function analysis was conducted using all scales except that for **Voluntariness.** This was done to determine how well they predicted membership in the adopter/nonadopter categories, and their relative weights. Of the original 532 responses, 25 were dropped from the analysis because they reported themselves as "former users" of the PWS, and hence did not fit the "adopter/nonadopter" classification scheme. The discriminant function including all variables showed a strong association between the categories and the predictors [$\chi^2(7) = 132$, $p < 0.001$], correctly classifying 85% of the sample. When **Compatibility** is dropped from the

TABLE 4
Discriminant Analysis
Standardized Canonical Discriminant Function Coefficients

	Equation	
Constructs Included	1	2
Compatibility	0.59	N/A
Relative Advantage	0.27	0.70
Result Demonstrability	0.43	0.44
Visibility	0.35	0.39
Ease of Use	−0.29	−0.17
Trialability	−0.22	−0.20
Image	−0.08	−0.03

analysis, the significance levels and relative ordering and magnitudes of the effects of the perceptions undergo some changes, with **Ease of Use** showing a less significant effect in the second equation. On the other hand, **Relative Advantage** shows a greater effect, which reflects its correlation with **Compatibility.** Table 4 suggests that the best predictors for distinguishing between the categories are **Relative Advantage, Result Demonstrability,** and **Visibility.** On the other hand, both **Trialability** and **Image** seem to be weak predictors of adoption for this particular study. These results support the notion that the perceptions of using an innovation do affect one's decision to adopt or reject it, and suggest further areas of investigation into the relative effects of these perceptions.

e. *Instrument Reduction.* An ongoing concern to survey researchers is the length of survey instruments. While the length of the instrument is relatively short given the number of constructs being measured, further refinement is always possible. First, depending on one's objectives, the results of the discriminant analysis indicate that one perception, **trialability,** had significantly less weight than the others *in an organizational context.* This may be a result of organizations making the technology available at no risk to the individual adopter, thus minimizing its discriminating power. Thus, researchers may be able to control for this factor, rather than directly measure it and thus drop this scale from their study. On the other hand, **trialability** should be a significant concern for those who would adopt an innovation *at their own risk,* and thus studies of consumer behaviour, for example, should consider this factor.

Deletion of items from scales is another means of reducing the instrument length, but this should be done with caution. Reduction may threaten both the validity and reliability of the scale. Nevertheless, thirteen items were identified, which if deleted *for this sample,* would not have had a significant negative effect on ALPHA, and should not affect the content validity of the scales. The result would be a 25-item (identified in Appendix 3 by an asterisk) instrument, measuring eight constructs. Table 5 indicates the ALPHA scores for the shortened scales. Again, it must be cautioned that these scales are untested in a new sample.

5. Summary and Conclusions

The instrument development research described here offers several contributions. The most obvious is the creation of an overall instrument to measure various perceptions of using an information technology innovation. The creation process included surveying known existing instruments, choosing appropriate items, creating new items as necessary, and then undertaking an extensive scale development process. It is believed that the method of developing the scales provides a high degree of confidence in their content and construct validity. The result is a parsimonious, 34-item instrument, comprising seven scales, all with acceptable levels of reliability. This instrument can now be used to investigate how perceptions affect individuals' actual use of information technology as well as other innovations.

The development process also helped to clarify and refine some of the definitions of the various PCI. The development steps showed that **Observability** as originally defined by Rogers seemed to be tapping two distinctly different constructs, **Result Demonstrability** and **Visibility.** As a result, two different scales were developed. The development process also further demonstrated the validity and reliability of the two scales imported from Davis (1986). Finally, the issue of the effects of corporate mandates on adoption is one which has been raised as a further factor which needs

TABLE 5
ALPHA Coefficients of Short Scales

Construct	Items	ALPHA
Relative Advantage	5	0.90
Compatibility	3	0.86
Ease of Use	4	0.84
Result Demonstrability	4	0.79
Image	3	0.79
Visibility	2	0.83
Trialability	2	0.71
Voluntariness	2	0.82
Total Number of Items	25	

addressing in diffusion research (Bayer and Melone 1989). In this study, **Perceived Voluntariness** was shown to be more than a binary variable. Many researchers assume, based on respondents' level in the hierarchy or on job descriptions, that certain behaviours are either voluntary or compulsory. Based on this assumption, conclusions are drawn about the effects of other variables. Use of the **Voluntariness** scale should help to clarify assumptions about the freedom of choice in adopting innovations.

An eighth scale to measure **Compatibility** was also developed. It was initially refined during the sorting rounds to drop references to the **needs** of the potential adopter, as this tended to be confounded with **Relative Advantage**. While the card sorting indicated that it appeared to be conceptually distinct from **Relative Advantage,** and while its scale has good reliability, principal components analysis grouped its items with those of **Relative Advantage.** This indicates that still more work is needed in this area to investigate the relationship between these two constructs.

Finally, it is believed that this project further extends our instrument development methodology. The process involved several deliberate steps to develop the scales, and the results show that these steps had the desired results. The sorting routines seem to have been especially valuable, given the relatively distinct factors which emerged from the unconstrained factor analysis. In addition, having judges provide their own labels for the groupings helps in verifying the construct validity of the scales.

In recommending this instrument to researchers investigating perceptions of innovations, a few points of caution are appropriate. While the various items were developed to be as general as possible, they were worded and tested with respect to a particular innovation, the Personal Work Station, in a particular context, organizational work. Nevertheless, it is believed that they could be easily reworded by substituting the names of different IT innovations, though additional checks for validity and reliability would be prudent after rewording. Second, other contexts, such as

consumer buying behaviour, may introduce other factors which need examination. Therefore, further investigation is needed with other innovations in other contexts to investigate the generalizability of the scales.

One scale's reliability is relatively lower than the commonly desired standard of 0.80, although there is potential for significant improvement in the scale by including previously deleted items. As was discussed earlier, however, the desired reliability levels may differ depending on the purposes of the research. Researchers using these scales will have to consider such factors. In conclusion, it is believed that the final instrument, which was developed based on a model of general factors that have predicted the adoption of innovations quite successfully, offers a useful tool for the study of the initial adoption and diffusion of innovations.*

Acknowledgements. The authors would like to thank all those individuals who assisted in this study, especially Professor Craig Pinder in helping to create the sorting procedures, and Professor Peter Todd in the development of the inter-judge agreement measures. Furthermore, this study would not have been possible without the help of the 21 judges who actually carried out the sorting steps, and the eight companies who participated in the field testing of the instrument. Finally, the comments of Professors Burt Swanson, David Goldstein, Sue Conger, Wynne Chin and anonymous reviewers of an earlier draft of this paper were also most helpful. The cooperation of all these individuals is most appreciated.

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada.

* Robert W. Zmud, Associate Editor. This paper was received on January 28, 1991 and has been with the authors 2 months for 1 revision.

Appendix 1

Item Placement Scores

The **item placement ratio** is an indicator of how many items were placed in the intended, or target, category by the judges. As an example of how this measure could be used, consider the simple case of **three** theoretical constructs with **ten items** developed for each construct. With a panel of four judges, a theoretical total of 40 placements (4 judges times 10 items) could be made within each construct. Thus a **THEORETICAL** versus **ACTUAL** matrix of item placements could be created as shown in the figure below (including an ACTUAL "Not Applicable" column where judges could place items which they felt fit none of the categories).

CONSTRUCTS		ACTUAL					
		A	B	C	N/A	Total	% Hits
THEORETICAL	A	**36**	2	1	1	40	90
	B	12	**24**	4	0	40	60
	C	0	0	**40**	0	40	100

Item Placements: 120 Hits: 100 Overall "Hit Ratio": 83%

Examination of the *diagonal* of the matrix shows that with a theoretical maximum of 120 target placements (three constructs at 40 placements per construct), a total of 100 "hits" were achieved, for an overall "hit ratio" of 80%. Secondly, and more importantly, examination of each *row* shows how the items created to tap the particular constructs are actually being classified. For example, Row C shows that all 40-item

placements were within the target construct, but that in Row B, only 60% (24/40) were within the target. In the latter case, 12 of the placements were made in Construct A, which might indicate that the items underlying these placements are not differentiated enough from the items created for Construct A. This would lead one to have confidence in a scale based on Row C, but be hesitant about accepting any scale based on Row B. Finally, examination of the *off-diagonal* entries indicates how factorially complex any construct might be. Actual constructs based on *columns* with a high number of entries in the off-diagonal might be considered to be too ambiguous. Likewise, the items themselves in the off-diagonals might be ambiguous, so any consistent pattern of item misclassification should be examined.

In the tables that follow, the actual scores for the four sorting rounds are provided.

Appendix 1.
Item Placement Ratios

FIRST SORTING ROUND

	ACTUAL CATEGORIES									
TARGET CATEGORY	VOL	IMAGE	REL ADV	COMPAT	EOU	TRIAL	OBS	N/A	TOT	TGT %
VOL	20		1	3					24	83
IMAGE		29	3						32	91
REL ADV		1	71						72	99
COMPAT	5	1	3	42	4		1	8	64	66
EOU			2		69			1	72	96
TRIAL	17			1		38			56	68
OBSERV	1	1	23		3	2	24	2	56	43
Total Item Placements: 376			Hits: 293				Overall Hit Ratio: 78%			

SECOND SORTING ROUND

	ACTUAL CATEGORIES									
TARGET CATEGORY	VOL	IMAGE	REL ADV	COMPAT	EOU	TRIAL	OBS	N/A	TOT	TGT %
VOL	23		1						24	96
IMAGE		32							32	100
REL ADV		3	65	2				2	72	90
COMPAT	1		1	40				2	44	91
EOU				1	65		1		68	96
TRIAL					1	50			52	96
OBSERV		1	8	1		2	32	1	52	96
Total Item Placements: 336			Hits: 308				Overall Hit Ratio: 92%			

Appendix 1. (*continued*).
Item Placement Ratios

THIRD SORTING ROUND

	ACTUAL CATEGORIES										
TARGET CATEGORY	VOL	IMAGE	REL ADV	COMPAT	EOU	TRIAL	RES DEM	VISI-BILITY	N/A	TOT	TGT %
VOL	28		2							30	93
IMAGE	7	29	4							40	73
REL ADV			83		2					85	98
COMPAT	1		5	29	13	3		3	1	55	53
EOU					85					85	100
TRIAL	3				2	54		2	4	65	83
RES DEM			1			1	47		1	50	94
VISIBILITY						12		33		45	73
Total Item Placements: 445			Hits: 388			Overall Hit Ratio: 85%					

FOURTH SORTING ROUND

	ACTUAL CATEGORIES										
TARGET CATEGORY	VOL	IMAGE	REL ADV	COMPAT	EOU	TRIAL	RES DEM	VISI-BILITY	N/A	TOT	TGT %
VOL	20									230	100
IMAGE		26	1						1	28	93
REL ADV			59	3		2	3		1	68	87
COMPAT				43		1				44	98
EOU				1	63	4				68	93
TRIAL	7					37				44	84
RES DEM			1		1		29		1	32	91
VISIBILITY							2	34		36	94
Total Item Placements: 340			Hits: 311			Overall Hit Ratio: 92%					

Appendix 2. Judges' Labels for Categories

First Sorting Round Individual Judge's Construct Labels

	JUDGES			
CONSTRUCTS	A	B	C	D
VOLUNTARINESS	Access/ Barriers	Voluntariness	Personal Fit	Voluntariness
IMAGE	Image/ Visibility	Image	Image	Status
RELATIVE ADVANTAGE	Advantages	Benefits	General Usefulness	Value
COMPATABILITY	Fit with Personal Style	Experience	Fit with Job	Change in Work Patterns
EASE OF USE	Required Skills	Understanding	Ease of Use	Ease
TRIALABILITY		Trialability	Trialability	Availability
RESULT DEMONSTRABILITY	Explaining Results	Perception of Results	Measureability	

First Sorting Round Panel's Labels and Definitions

Voluntariness: the degree to which the use of the PWS is perceived as being voluntary.
Image: the degree to which the use of the PWS enhances one's image or status within the organization.
Benefit: the degree to which the use of the PWS is a benefit in one's job.
Compatibility: the degree to which use of the PWS is compatible with, or requires change, in one's job.
Ease of Use: the degree to which the **PWS** is easy to learn and use.
Trialability: the degree to which it is possible to try using the **PWS.**
Measurability: the degree to which the benefits of using the **PWS** are measurable

Third Sorting Round Individual Judge's Construct Labels

	JUDGES				
CONSTRUCTS	V	W	X	Y	Z
VOLUNTARINESS	Voluntary Usage	Voluntariness & Risk of Failure	PWS Required for Job?	External Pressure to Use PWS	Mgmt Pressure to use PWS
IMAGE	Status	Status	Status		Image
RELATIVE ADVANTAGE	Efficiency	Advantages of PWS	Job Enhancement	Costs/Benefits of Using PWS	Effects of PWS on Job Performance
COMPATIBILITY	Compatibility	Effect on Work Habits	Initial Perceptions of Using PWS		Change required in work habits
EASE OF USE	Ease of Use	User Friendliness		Ease of Use	User Friendliness
RESULT DEMONSTRABILITY	Tangibility	Measureability	Ability to communicate knowledge	Understanding consequences of using PWS	Communicating pros and cons of PWS usage
VISIBILITY	Exposure	Familiarity with PWS	Awareness of PWS	Exposure to PWS	Exposure to PWS usage
TRIALABILITY		Accessibility	Testing of PWS	Ease of experimentation	Climate for risk taking

Appendix 3. List of Items by Construct

Voluntariness

1. My superiors expect me to use a PWS.
2. My use of a PWS is voluntary (as opposed to required by my superiors or job description).
3.* My boss does not require me to use a PWS.
4.* Although it might be helpful, using a PWS is certainly not compulsory in my job.

Relative Advantage

1.* Using a PWS enables me to accomplish tasks more quickly.
2.* Using a PWS improves the quality of work I do.
3.* Using a PWS makes it easier to do my job.
4. The *disadvantages* of my using a PWS far outweigh the advantages. (See Note a.)
5. Using a PWS improves my job performance.
6. Overall, I find using a PWS to be advantageous in my job.
7.* Using a PWS enhances my effectiveness on the job.
8.* Using a PWS gives me greater control over my work.
9. Using a PWS increases my productivity.

Compatibility

1.* Using a PWS is compatible with all aspects of my work.
2. Using a PWS is completely compatible with my current situation.
3.* I think that using a PWS fits well with the way I like to work.
4.* Using a PWS fits into my work style.

Image

1. Using a PWS improves my image within the organization.
2. Because of my use of a PWS, others in my organization see me as a more valuable employee. (See Note a.)
3.* People in my organization who use a PWS have more prestige than those who do not.
4.* People in my organization who use a PWS have a high profile.
5.* Having a PWS is a status symbol in my organization.

Ease of Use

1. I believe that a PWS is cumbersome to use.
2. It is easy for me to remember how to perform tasks using a PWS. (See Note a.)
3. My using a PWS requires a lot of mental effort.
4. Using a PWS is often frustrating.
5.* My interaction with a PWS is clear and understandable. (See Note a.)
6.* I believe that it is easy to get a PWS to do what I want it to do.
7.* Overall, I believe that a PWS is easy to use.
8.* Learning to operate a PWS is easy for me.

Result Demonstrability

1.* I would have no difficulty telling others about the results of using a PWS.
2.* I believe I could communicate to others the consequences of using a PWS.
3.* The results of using a PWS are apparent to me.
4.* I would have difficulty explaining why using a PWS may or may not be beneficial.

Visibility

1. I have seen what others do using their PWS.
2.* In my organization, one sees PWS on many desks.
3. I have seen a PWS in use outside my firm. (See Note a.)
4.* PWS are *not* very visible in my organization.
5. It is easy for me to observe others using PWS in my firm.

I have had plenty of opportunity to see the PWS being used. (See Note b.)
I have not seen many others using a PWS in my department. (See Note b.)

Trialability

1. I've had a great deal of opportunity to try various PWS applications.
2. I know where I can go to satisfactorily try out various uses of a PWS.

Appendix 3. (*continued*).
List of Items by Construct

3. A PWS was available to me to adequately test run various applications.
4.* Before deciding whether to use any PWS applications, I was able to properly try them out.
5.* I was permitted to use a PWS on a trial basis long enough to see what it could do.
I am able to experiment with the PWS as necessary. (See Note b.)
I can have PWS applications for long enough periods to try them out. (See Note b.)
I did not have to expend very much effort to try out the PWS. (See Note c.)
I don't really have adequate opportunities to try out different things on the PWS. (See Note c.)
A proper on-the-job tryout of the various uses of the PWS is not possible. (See Note c.)
There are enough people in my organization to help me try the various uses of the PWS. (See Note c.)

Notes

a. The indicated items were all deleted as the result of the first factor analysis and hence were not in the final scales.

b. The indicated items, which were deleted after the initial test, are suggested as candidates for inclusion in any expanded scale

c The indicated items, which were not in the final instrument, had item-scale correlations less than 0.40 in the initial test and are suggested as secondary candidates for lengthening the scale.

d. *—indicates items suggested for inclusion in any "short" scales.

Appendix 4. Rotated Factor Matrix

Numbers in brackets refer to respective item numbers for each scale in Appendix 3.

ITEM	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
RELATIVE ADV (3)	0.83014	−0.20739	0.09293	−0.05519
RELATIVE ADV (2)	0.80094	−0.17114	0.04037	0 08332
RELATIVE ADV (7)	0.78119	−0.01433	−0.05072	0.12049
RELATIVE ADV (9)	0.76953	−0.10946	0.10502	0.10199
RELATIVE ADV (6)	0.76644	−0.09817	0.12219	0.02485
RELATIVE ADV (5)	0.74996	−0.12779	0 11349	0 15115
COMPATIBILITY (4)	0.73611	−0.11529	0.23797	0.12184
COMPATIBILITY (3)	0.72881	−0.00200	0.18394	0.15350
RELATIVE ADV (1)	0.72514	−0.20658	0.12878	−0.03200
RELATIVE ADV (8)	0.72305	0.02852	0.08674	0 16745
COMPATIBILITY (2)	0.66501	−0.33841	0 15294	−0.00917
COMPATIBILITY (1)	0.58393	−0.36829	0 11385	0.10820
VOLUNTARINESS (3)	−0.24846	0.83002	0.01897	−0.11224
VOLUNTARINESS (4)	−0.22549	0.82352	−0.05387	−0.04325
VOLUNTARINESS (2)	−0.13496	0.75335	0 04197	0 01211
VOLUNTARINESS (1)	−0.39014	0.66053	−0.00389	−0.12728
EASE OF USE (4)	0.15799	0.06633	0.75956	−0.03231
EASE OF USE (3)	0.01615	0.06851	0 71169	0.08009
EASE OF USE (7)	0.31585	−0.13606	0.67534	0.03490
EASE OF USE (8)	0.30347	−0.09989	0 59111	−0.02837
EASE OF USE (6)	0.42866	−0.12978	0 53742	0.11739
EASE OF USE (1)	0.36079	0.01858	0.44854	−0 19142
IMAGE (5)	0.00098	−0.01664	0.08473	0.81987
IMAGE (3)	0.12683	−0.07179	−0.11402	0.81140
IMAGE (4)	0.15317	−0.03459	−0.10110	0.80571
IMAGE (1)	0.36678	−0.13300	−0.01785	0.60337
RESULT DEMON (1)	0.22381	−0.05311	0.15225	0.00869
RESULT DEMON (2)	0.24694	−0.05343	0 13015	0.14440
RESULT DEMON (4)	0.29714	0.20501	0 02087	−0.04363
VISIBILITY (4)	0.08991	−0.11546	0.08289	−0.00983
VISIBILITY (2)	0.09697	−0.19103	0.15220	0.07149
VISIBILITY (5)	0.02766	−0.01957	0.08790	0.09944

Appendix 4. (*continued*).
Rotated Factor Matrix

VISIBILITY (1)	0.18064	0.17062	−0.16816	−0.13335
TRIALABILITY (4)	−0.05055	0.22710	0.07451	0.02418
TRIALABILITY (5)	−0.08048	0.21537	0.00930	0.09812
TRIALABILITY (3)	0.04267	−0.13613	−0.10745	0.01328
TRIALABILITY (2)	0.03066	−0.09130	−0.00903	0.10412
TRIALABILITY (1)	0.29325	−0.13219	0.06494	0.11698
Percentage of Variance Explained by Each Factor (Cumulative in Brackets)				
First Analysis	32.2 (32.2)	7.9 (40.1)	6.2 (46.3)	5.0 (51.3)
Second Analysis	28.0 (28.0)	9.1 (37.1)	7.6 (44.7)	6.4 (51.1)

ITEM	FACTOR 5	FACTOR 6	FACTOR 7
RELATIVE ADV (3)	0.06912	0.08608	−0.00374
RELATIVE ADV (2)	0.03793	0.05807	0.00817
RELATIVE ADV (7)	0.17573	0.01759	−0.00384
RELATIVE ADV (9)	0.18030	−0.00078	0.03019
RELATIVE ADV (6)	0.12181	0.02884	−0.05624
RELATIVE ADV (5)	0.16653	0.04896	−0.00639
COMPATIBILITY (4)	0.19560	0.07945	−0.01528
COMPATIBILITY (3)	0.13890	0.08462	−0.02136
RELATIVE ADV (1)	0.04817	0.06957	0.01356
RELATIVE ADV (8)	0.14521	0.05715	0.10302
COMPATIBILITY (2)	0.03006	0.10007	0.02579
COMPATIBILITY (1)	0.07125	−0.01437	0.01708
VOLUNTARINESS (3)	−0.03411	−0.08338	0.03974
VOLUNTARINESS (4)	−0.01256	−0.03048	−0.01179
VOLUNTARINESS (2)	0.06035	−0.00232	0.25423
VOLUNTARINESS (1)	−0.02652	−0.24926	0.00463
EASE OF USE (4)	−0.03083	−0.02595	0.02579
EASE OF USE (3)	−0.05601	0.10558	−0.14609
EASE OF USE (7)	0.31625	0.06347	0.02563
EASE OF USE (8)	0.34284	0.02146	0.09205
EASE OF USE (6)	0.26751	0.08114	0.20712
EASE OF USE (1)	0.00950	0.12652	0.08718
IMAGE (5)	−0.01055	0.00627	0.07916
IMAGE (3)	0.15986	0.04397	0.01247
IMAGE (4)	0.07527	0.00302	0.02588
IMAGE (1)	−0.13407	0.05213	0.13006
RESULT DEMON (1)	0.77657	−0.02705	0.04038
RESULT DEMON (2)	0.77462	0.08508	0.14002
RESULT DEMON (4)	0.61593	0.11056	0.02345
VISIBILITY (4)	−0.01435	0.85762	−0.05833
VISIBILITY (2)	−0.02133	0.76693	−0.12640
VISIBILITY (5)	0.15549	0 70787	0.29863
VISIBILITY (1)	0.20036	0.49983	0.25815
TRIALABILITY (4)	0.03866	−0.07538	0.78058
TRIALABILITY (5)	−0.14145	−0.05224	0.72220
TRIALABILITY (3)	0.29085	0.26712	0.61574
TRIALABILITY (2)	0.34848	0.36574	0.48406
TRIALABILITY (1)	0.32407	0.15852	0.48206
Percentage of Variance Explained by Each Factor (Cumulative in Brackets)			
First Analysis	4.5 (55.8)	3.8 (59.6)	3.1 (62.7)
Second Analysis	4.6 (55.7)	4.0 (59.6)	3.6 (63.2)

Appendix 5. Inter-Item Correlations by Scale

In the matrices that follow, items are identified by their respective number in Appendix 3

Relative Advantage and Compatibility

	RA1	RA2	RA3	RA5	RA6	RA7	RA8	RA9	COMP1	COMP2	COMP3	COMP4
RA1	1.00											
RA2	0.65	1.00										
RA3	0.75	0.72	1.00									
RA5	0.50	0.61	0.60	1.00								
RA6	0.53	0.58	0.68	0.69	1.00							
RA7	0.45	0.58	0.58	0.72	0.66	1.00						
RA8	0.46	0.55	0.53	0.56	0.50	0.60	1.00					
RA9	0.55	0.60	0.63	0.67	0.57	0.77	0.68	1.00				
COMP1	0.61	0.53	0.68	0.50	0.55	0.41	0.44	0.48	1.00			
COMP2	0.53	0.51	0.55	0.44	0.42	0.41	0.46	0.47	0.63	1.00		
COMP3	0.53	0.64	0.61	0.53	0.57	0.52	0.55	0.50	0.56	0.46	1.00	
COMP4	0.51	0.62	0.61	0.63	0.61	0.57	0.59	0.60	0.58	0.51	0.72	1.00

Voluntariness

	VOL1	VOL2	VOL3	VOL4
VOL1	1.00			
VOL2	0.46	1.00		
VOL3	0.66	0.61	1.00	
VOL4	0.55	0.60	0.75	1.00

Image

	IMAGE1	IMAGE2	IMAGE4	IMAGE5
IMAGE1	1.00			
IMAGE2	0.42	1.00		
IMAGE4	0.40	0.66	1.00	
IMAGE5	0.45	0.57	0.53	1.005

Ease of Use

	EOU1	EOU3	EOU4	EOU6	EOU7	EOU8
EOU1	1.00					
EOU3	0.31	1.00				
EOU4	0.37	0.25	1.00			
EOU6	0.35	0.46	0.37	1.00		
EOU7	0.42	0.43	0.41	0.68	1.00	
EOU8	0.37	0.44	0.38	0.53	0.60	1.00

Result Demonstrability

	RD1	RD2	RD4
RD1	1.00		
RD2	0.64	1.00	
RD4	0.43	0.49	1.00

Trialability

	TRIAL1	TRIAL2	TRIAL3	TRIAL4	TRIAL5
TRIAL1	1.00				
TRIAL2	0.53	1.00			
TRIAL3	0.38	0.47	1.00		
TRIAL4	0.20	0.26	0.35	1.00	
TRIAL5	0.11	0.15	0.27	0.56	1.00

Appendix 5. (*continued*).
Inter-Item Correlations by Scale

Visibility

	VIS1	VIS2	VIS4	VIS4
VIS1	1.00			
VIS2	0.14*	1.00		
VIS4	0.30	0.66	1.00	
VIS5	0.40	0.40	0.50	1.00

Note All correlations in the above matrices were significant at the 0.001 level, except for one (in the Visibility scale) indicated by an asterisk (*).

Bibliography

Ajzen, Icek, and Martin Fishbein, *Understanding Attitudes and Predicting Behavior,* Prentice-Hall Inc., Englewood Cliffs, NJ, 1980.

Alexander, Mary, "The Adoption and Implementation of Computer Technology in Organizations: The Example of Database Machines," unpublished Doctoral Dissertation, Indiana University, 1989.

Bailey, J. E. and S. W. Pearson, "Development of a Tool for Measuring and Analyzing Computer Satisfaction," *Communications of the ACM,* 26, 10 (October 1983), 785–793.

Bayer, Judy and Nancy Melone, "A Critique of Diffusion Theory as a Managerial Framework for Understanding Adoption of Software Engineering Innovations," *The Journal of Systems and Software,* 9 (1989), 161–166.

Benbasat, Izak, "Laboratory Experiments in Information Systems Studies with a Focus on Individuals: A Critical Appraisal," *The Information Systems Research Challenge. Experimental Research Methods.* Vol. 2, Izak Benbasat (Ed.), Harvard Business School, Boston, MA, 1990, 33–47.

Bohrnstedt, George W., "Measurement," in Peter H. Rossi, James D. Wright, and Andy B. Anderson, (Eds.), *Handbook of Survey Research,* Academic Press Inc., San Diego, 1983. 69–121.

Bollen, Kenneth A., and Rick H. Hoyle, "Perceived Cohesion: A Conceptual and Empirical Examination," *Social Forces,* 69, 2 (December 1990), 470–504.

Bolton, William T., "The Perception and Potential Adoption of Channel 2000: Implications for Diffusion Theory and Videotext Technology," Unpublished Doctoral Dissertation, Ohio State University, 1981.

Brancheau, James C. and James C. Wetherbe, "The Adoption of Spreadsheet Software: Testing Innovation Diffusion Theory in the Context of End-User Computing," *Information Systems Research,* 1, 2 (June 1990), 115–143.

Burt, R. S., "Interpretational Confounding of Unobserved Variables in Structural Equation Models," *Sociological Methods & Research,* 5 (August 1976), 3–52.

———, "Social Contagion and Innovation: Cohesion versus Structural Equivalence," *American Journal of Sociology,* 92, 6 (May 1987), 1287–1335.

Campbell, D. T. and D. W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin,* 56 (1959), 81–105.

Cohen, J. A., "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement,* 20 (1960), 37–46.

Comrey, A. L., *A First Course in Factor Analysis,* Academic Press, New York, 1973.

Cronbach, Lee J., *Essentials of Psychological Testing,* (Third Ed.), Harper and Row, New York, 1970.

Davis, Fred D., "A Technology Acceptance Model for Empirically Testing New End User Information Systems: Theory and Results," Unpublished Doctoral Dissertation, Massachusetts Institute of Technology, 1986.

———, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly,* 13, 3 (September 1989), 319–340.

———, Richard P. Bagozzi and Paul R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science,* 35, 8 (August 1989), 982–1003.

Dickson, Gary W., Izak Benbasat and William R. King, "The Management Information Systems Area: Problems, Challenges and Opportunities," *Proceedings of the First International Conference on Information Systems,* E. R. McLean (Ed.), Philadelphia, PA, December 1980, 1–7.

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

Downs, George W. Jr and Lawrence B. Mohr, "Conceptual Issues in the Study of Innovation," *Administrative Science Quarterly*, 21 (December 1976), 700–714

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

Fornell, Claes, "Issues in the Application of Covariance Structure Analysis: A Comment," *Journal of Consumer Research*, 9 (March 1983), 443–448.

Gerbing, David W. and John E. Hunter, *ITAN A Statistical Package for ITem ANalysis with Correlational Data Including Multiple Groups Confirmatory Factor Analysis*, Portland State University, Portland, OR, 1988.

Gorsuch, Richard L., "Exploratory Factor Analysis," in John R Nesselroade and Raymond B. Cattell, *Handbook of Multivariate Experimental Psychology*, (Second Ed.) Plenum Press, New York, 1988, 231–258.

Guttman, L., "A Basis for Analyzing Test-Retest Reliability," *Psychometrika*, 10, 4 (December 1945), 255–282.

Holloway, Robert E., "Perceptions of an Innovation. Syracuse University Project Advance," Unpublished Doctoral Dissertation, Syracuse University, 1977.

Huber, George P., "Cognitive Style as a Basis for MIS and DSS Design: Much Ado About Nothing?", *Management Science*, 29, 5 (May 1983), 567–582.

Hurt, H. Thomas and Rob Hubbard, "The Systematic Measurement of the Perceived Characteristics of Information Technologies: Microcomputers as Innovations," presented at the ICA Annual Conference, Montreal, Quebec, May 1987

Ives, Blake and Margrethe H Olson, "User Involvement and MIS Success: A Review of Research," *Management Science*, 30, 5 (May 1984), 586–603.

———, ——— and Jack J. Baroudi, "The Measurement of User Information Satisfaction " *Communications of the ACM*, 26, 10 (October 1983), 785–793.

Jarvenpaa, Sirkka, "The Effect of Task Demands and Graphical Format on Information Processing Strategies," *Management Science*, 35, 3 (March 1989), 285–303.

———, Gary W. Dickson and Gerardine DeSanctis, "Methodological Issues in Experimental IS Research: Experiences and Recommendations," *MIS Quarterly*, 9, 2 (June 1985), 141–156.

Johnson, Bonnie M. and Ronald E. Rice, *Managing Organizational Innovation*, Columbia University Press, New York, 1987.

———, G. W Dickson and G. DeSanctis, "Methodological Issues in Experimental IS Research," *Management Science*, 30, 5 (May 1984), 586–603.

Keen, Peter G. W. "MIS Research: Reference Disciplines and a Cumulative Tradition," *Proceedings of the First International Conference on Information Systems*, E. R. McLean (Ed.), Philadelphia, PA, December 1980, 9–18.

Kerlinger, Fred N. *Foundations of Behavioral Research*, (Third Ed.), Holt, Rinehart and Winston, New York, 1978.

Kwon, Tae H. and Robert W Zmud, "Unifying Fragmented Models of Information Systems Implementation," in *Critical Issues in Information Systems Implementation*, R. J. Boland and R. A. Hirschheim (Eds.), John Wiley and Sons, Chichester, 1987, 227–251.

Larcker, David F. and V. Parker Lessig, "Perceived Usefulness of Information: A Psychometric Examination," *Decision Sciences*, 11 (1980), 121–134.

Lastovicka, John L. and Kanchana Thamodaran, "Common Factor Score Estimates in Multiple Regression Problems," *Journal of Marketing Research*, 28 (February 1991), 105–112.

Leonard-Barton, Dorothy, "Implementation Characteristics of Organizational Innovations," *Journal of Communication Research*, 15, 5 (October 1988), 603–631.

——— and Isabelle Deschamps, "Managerial Influence in the Implementation of New Technology," *Management Science*, 34, 10 (October 1988), 1252–1265

Moore, Gary C., "End User Computing and Office Automation: A Diffusion of Innovations Perspective," *INFOR*, 25, 3 (1987), 214–235.

Nunnally, Jum C., *Psychometric Theory*, McGraw Hill, New York, 1967.

Ostlund, Lyman E., "Product Perceptions and Predispositional Factors as Determinants of Innovative Behavior," Unpublished Doctoral Dissertation, Harvard University, 1969.

Rice, Ronald, Personal Communication, January 1987.

Rogers, Everett M., *Diffusion of Innovations,* (Third Ed.), The Free Press, New York, 1983.

Swanson, E. Burton, *Information Systems Implementation Bridging the Gap Between Design and Utilization,* Richard Irwin, Inc., Homewood, IL, 1988.

Tabachnik, Barbara and Linda S. Fidell, *Using Multivariate Statistics,* (Second Ed.), Harper & Row, New York, 1989.

Taylor, Ronald and Izak Benbasat, "A Critique of Cognitive Styles Theory and Research," *Proceedings of the First International Conference on Information Systems,* E. R. McLean (Ed.), Philadelphia, PA, December 1980, 82–90.

Todd, Peter A. and Izak Benbasat, "An Experimental Investigation of the Impact of Computer Based Decision Aids on the Process of Preferential Choice," Working Paper, School of Business, Queen's University, Kingston, Ontario, October 1989.

Tornatzky, Louis G. and Katherine J. Klein, "Innovation Characteristics and Innovation Adoption-Implementation: A Meta-Analysis of Findings," *IEEE Transactions on Engineering Management,* EM-29, 1 (February 1982), 28–45.

Treacy, Michael E., "Towards a Cumulative Tradition of Research on Information Technology as a Strategic Factor," Working Paper, Center for Information Systems Research, Sloan School of Management, MIT, 1986.

Vessey, Iris, "An Investigation of the Psychological Processes Underlying the Debugging of Computer Programs," Unpublished PhD Dissertation, Department of Commerce, University of Queensland, April 1984.

Zajonc, Robert B. and Hazel Markus, "Affective and Cognitive Factors in Preferences," *Journal of Consumer Research,* 9 (September 1982), 123–131.

Zaltman, Gerald, Robert Duncan, and Jonny Holbek, *Innovations and Organizations,* Wiley and Sons, New York, 1973.