



A critical assessment of potential measurement biases in the technology acceptance model: three experiments

FRED D. DAVIS

University of Maryland, College Park, U.S.A.

VISWANATH VENKATESH

Carlson School of Management, University of Minnesota, MN 55455, U.S.A.

(Received 13 February 1995 and accepted in revised form 7 December 1995)

The Technology Acceptance Model (TAM) is widely used by researchers and practitioners to predict and explain user acceptance of information technologies. TAM models system usage intentions and behavior as a function of perceived usefulness and perceived ease of use. The original scales for measuring the TAM constructs have been confirmed to be reliable and valid in several replications and applications spanning a range of technologies and user populations. However, a measurement bias may be present because the TAM instrument physically *groups* together the multiple items measuring each individual construct. Many scholars of instrument design would object to such item grouping, instead advocating that items from different constructs be *intermixed* in order to reduce "carryover" effects among the responses to multiple items targeting a specific construct, which might artificially inflate the observed reliability and validity. Three experiments involving two systems and a total of 708 subjects are reported which address whether such carryover biases are present in the TAM measures. All three studies found that item *grouping* vs. item *intermixing* had no significant effect (positive or negative) either on the high levels of reliability and validity of the TAM scales, or on the path coefficients linking them together. Ironically, open-ended verbal evaluations indicated that subjects were more confused and annoyed when items were *intermixed*, suggesting a tendency toward "output interference" effects, which themselves could have a biasing effect. Our findings suggest that those who employ the TAM measures should continue using the original (*grouped*) format in order to best predict and explain user acceptance of information technology.

© 1996 Academic Press Limited

1. Introduction

Information technology (IT) frequently undergoes dramatic changes within short periods of time (e.g. Buell & Brandt, 1990; Burgess, 1991). Organizations that supply or consume information technologies face risky investment decisions that can have major effects on their competitive position. Many examples of emerging information technologies have been touted with multi-billion dollar market projections, but remain shrouded in uncertainty as to their eventual effectiveness. These include multimedia (Milburn, 1988; Burgess, 1991; Lewis, 1991; Markoff, 1991a), pen-based computing (Buell & Brandt, 1990; Keefe, 1991; Markoff, 1991b), wireless networking (Hogberg, 1993; McCarroll, 1994) and personal digital assistants (Daly, 1993; Halfhill, 1993). Which of these technologies will succeed? In which IT should

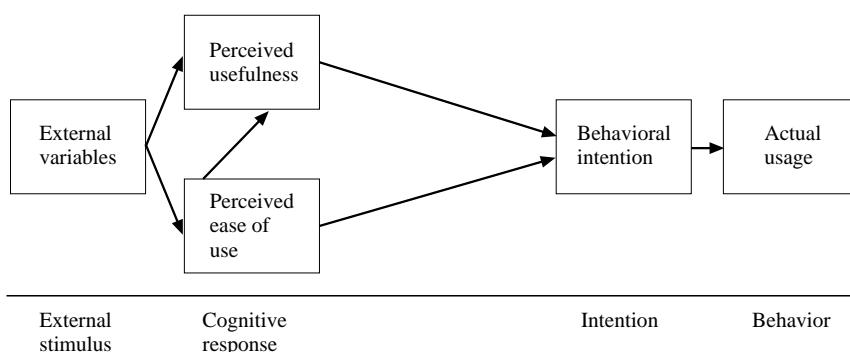


FIGURE 1. Technology acceptance model (TAM).

the organization invest? What are the effective applications going to be? Emerging information technologies cannot deliver improved organizational effectiveness if they are not accepted and used by potential users. Clearly, organizations need tools to direct resource allocation and investment decisions concerning the development and adoption of new and emerging information technologies.

1.1. TECHNOLOGY ACCEPTANCE MODEL (TAM)

One tool for assessing and predicting user acceptance of emerging IT which has gained popularity in recent years is the technology acceptance model (TAM) (Davis, 1986; Davis, 1989; Davis, Bagozzi & Warshaw, 1989; Davis, 1993). TAM was developed under contract with IBM Canada, Ltd. in the mid-1980s where it was used to evaluate the market potential for a variety of then-emerging PC-based applications in the area of multi-media, image processing, and pen-based computing in order to guide investments in new product development. TAM (Figure 1), which is an adaptation of the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975; Ajzen & Fishbein, 1980), was designed to understand the causal chain linking external variables to its user acceptance and actual use in a workplace. External variables such as objective system design characteristics, training, computer self-efficacy, user involvement in design, and the nature of the implementation process are theorized to influence behavioral intention to use, and ultimately usage, indirectly via their influence on perceived usefulness and perceived ease of use.

Research in psychology and TAM itself suggest that users' intention to use is the single best predictor of actual system usage. The intention to use (BI) is determined by one's attitude towards using. This attitude in turn is determined by two specific beliefs perceived usefulness (U), the user's perception of the degree to which using a particular system will improve her/his performance, and perceived ease of use (EOU), the user's perception of the extent to which using a particular system will be free of effort (Davis, 1989; Davis *et al.*, 1989). While TRA would suggest that the impact of beliefs on intention should be completely mediated by attitude towards the behavior (i.e. in this the case the behavior is system use), the original conceptualization of TAM and subsequent research have shown that effect of perceived usefulness on intention is only partially mediated by attitude towards using. This is explained by Davis *et al.* (1989) as being attributable to the fact that in work settings, people may use a technology even if they do not have a positive

attitude (affect) towards using the same because it may provide productivity enhancement (i.e. be useful). In keeping with this, the original theoretical conceptualization of TAM included the attitude construct. However, based on the empirical evidence, the final TAM model (see Davis *et al.*, 1989: pp. 995–996) excluded the attitude construct because attitude did not fully mediate the effect of perceived usefulness on intention.

TAM has proven to be among the most effective models in the information systems literature for predicting user acceptance and usage behavior. The original instrument for measuring these beliefs was developed and validated by Davis (1986; 1989; 1993), and Davis *et al.* (1989), and replicated by Adams, Nelson and Todd (1992), Mathieson (1991), Hendrickson, Massey, and Cronan (1993), and Segars and Grover (1993). The instrument has also been used extensively by researchers investigating a range of issues in the area of user acceptance (e.g. Moore & Benbasat, 1991; Olfman & Bostrom, 1991; Trevino & Webster, 1992; Chin & Gopal, 1993; Sjazna, 1994; Venkatesh & Davis, 1994; Davis & Venkatesh, 1995; Taylor & Todd, 1995; Venkatesh & Davis, forthcoming).

TAM uses multiple-items scales to operationalize BI, U, and EOU in order to measure these constructs more reliably than would be possible with single-item scales. The Cronbach alpha reliability of the TAM scales has generally been found to exceed 0.9 across numerous studies. The scales have also exhibited a high degree of convergent, discriminant, and nomological validity. These strong psychometric properties, combined with the high proportion of variance in intention to use explained by perceived usefulness and ease of use, have led to confidence in the instrument for studying IS acceptance.

1.2. POTENTIAL BIASES IN TAM

A major potential concern is that the high reliability and validity of the TAM scales and the large proportion of variance in intention explained by perceived usefulness and ease of use could simply be an artifact of the measurement approach, which groups together multiple items measuring a single construct. In the original TAM instrument, and in most replications and applications cited above, the multiple questions measuring intention to use were grouped together, as were the items measuring perceived usefulness and ease of use. Many empirical studies have demonstrated that the psychometric properties of measurement scales can be affected by the ordering of items within the questionnaire (e.g. Schuman & Presser, 1981; Bradburn, 1982; Budd, 1987; Harrison & McLaughlin, 1991). Classical psychometric theory (e.g. Nunnally, 1978) does not explicitly account for these “context” effects. There have been some recent attempts to gain a better understanding of the cognitive processes people use to respond to survey questions regarding beliefs, attitudes, intentions, and behavior, and how these may be affected by item context (for recent reviews and syntheses, see Feldman & Lynch, 1988; Tourangeau & Rasinski, 1988). However, the status of theory regarding context effects has generally been characterized as “rudimentary” (Feldman & Lynch, 1988: p. 421), encompassing multiple interrelated processes that are not well understood and are “difficult to predict and sometimes difficult to replicate” (Tourangeau & Rasinski, 1988: p. 299).

One central mechanism by which the response to one survey question is theorized to be influenced by responses to previous questions derives from the notion of “accessibility” (e.g. Fazio, Chen, McDonel & Sherman, 1981). Within this view, constructs to be measured by survey questions are understood as “structures that reside in long-term memory” (Tourangeau & Rasinski, 1988: p. 299) and that “momentarily activated cognitions have disproportionate influence over judgments made about an object or on related behaviors performed shortly after their activation” (Feldman & Lynch, 1988: p. 421). That is, responding to a question may make the associated construct more accessible than usual in long term memory. If immediately subsequent questions ask about the same construct, their responses may be unduly influenced by this heightened accessibility, which in turn may artificially increase the consistency across the multiple responses to the same construct compared to what would have been obtained under normal levels of accessibility.

In order to minimize such artificially inflated consistency, many researchers have advocated intermixing items targeting one construct with those of other constructs (e.g. Heider, 1958; Abelson, Aronson, McGuire, Newcomb, Rosenberg & Tannenbaum 1968; Tedeschi, 1981; Harrison & McLaughlin, 1988). However, there has been little research directly evaluating whether item intermixing helps. In one particularly relevant study, Budd (1987) compared two alternative formats of a questionnaire that measured the components of the theory of reasoned action (intention, attitude, beliefs, evaluations, subjective norm, and normative beliefs) for three health-related behaviors (smoking cigarettes, brushing teeth and exercising). Version A of the questionnaire grouped items separately by construct and behavior, whereas version B randomly intermixed items across both constructs and behaviors. Attitude was apparently the only construct measured with multiple items. Cronbach alpha reliability obtained for the attitude scale for the three behaviors of 0.84, 0.79, and 0.78 using the grouped questionnaire fell to 0.47, 0.52, and 0.50, respectively, for the intermixed questionnaire. Similarly, most key structural relations between model components for the three behaviors were significantly reduced, going from being significant empirical links to being non-significant in all cases except one. Budd’s study is a highly worrisome indication that item grouping may artificially inflate the psychometric properties and structural relations among measured constructs.

Recognizing the speculative status of theory regarding context effects in general and carryover effects in particular, theorists have advocated conducting experiments that systematically assess the degree to which question placement influences obtained reliability and validity (both trait and nomological validity) (Feldman & Lynch, 1988: p. 432). The present research examines the question of whether the item grouping inherent in the design of the TAM instrument has any discernible effect on the psychometric properties of TAM’s measures for BI, U, and EOU, including reliability, convergent validity, discriminant validity, and nomological validity (i.e. the structural path coefficients linking together BI, U, and EOU). This research thus goes beyond our earlier work investigating measurement biases in the TAM instrument (see Davis & Venkatesh, 1995). The answer to this question has far-reaching implications for the many researchers and practitioners who employ TAM under the assumption that its measures are reliable and valid.

We conducted three experiments to directly assess the effect of item grouping

vs. intermixing on the psychometric properties of TAM's scales for BI, U and EOU, while counterbalancing order of construct, order of system, and titling, which may otherwise confound the results. Experiment 1 compares six different versions of a questionnaire, two grouped and four intermixed, measuring U and EOU regarding WordPerfect, counterbalancing order of construct. Experiment 2 compares seven versions of a questionnaire, four grouped and three intermixed, measuring BI, U, and EOU regarding WordPerfect, counterbalancing order of construct and titles. Experiment 3 compares 10 versions of a questionnaire, four grouped and six intermixed, regarding BI, U, and EOU of WordPerfect and Lotus, counterbalancing order of system, and order of construct.

2. Method—Experiment 1

This experiment was aimed at assessing any changes in the reliability and convergent and discriminant validity of the perceived usefulness and ease of use scales due to item intermixing. Two hundred and eighty students at the University of Minnesota participated in the study by filling out one of the six versions of the eight-item questionnaire assigned at random. All participants were currently enrolled in an information systems course, and were requested to fill out the questionnaire in a classroom setting. Responses that were considered usable were those from subjects who reported having a working knowledge of the target system, WordPerfect. One hundred and eighty-two usable responses were obtained.

All versions of the questionnaire had four items each that measured the users' perceptions about the U and EOU of WordPerfect. To determine the usability of the responses, there was one screening question on all versions of the questionnaire that determined whether the respondent possessed working knowledge of WordPerfect. Appendix 1 lists the four U and four EOU items that were used (also included are the two BI items, pertinent to Experiments 2 and 3). The different treatment groups and numbers of subjects in each group are given in Table 1. The first two versions of the questionnaire had grouped items: the first version had four U items grouped together, followed by four EOU items; the second version had four EOU items, followed by four U items. The next two versions had items in an alternating order: the third version of the questionnaire had items in an alternating order with the first question being on U, and the fourth version had items in an alternating order with the first question being on EOU. These two versions with

TABLE 1
Experimental design—Experiment 1

		Item ordering		
		Grouped	Intermixed	
			Alternating	Random
First item on questionnaire	Perceived usefulness first	$n = 30$	$n = 31$	$n = 31$
	Perceived ease of use first	$n = 30$	$n = 30$	$n = 30$
	Total	$n = 60$	$n = 61$	$n = 61$

TABLE 2
Reliability of perceived usefulness and ease of use scales—Experiment 1

	Grouped	Intermixed	
		Alternating	Random
Perceived usefulness	0.98	0.96	0.97
Perceived ease of use	0.97	0.96	0.97

Note: $n = 60$ (grouped items). $n = 61$ (alternating items). $n = 61$ (random items).

intermixed (alternating) items were designed specifically to ensure that there were no two successive items related to the same construct. The last two versions of the questionnaire had items in a random order (generated using a random number table): the fifth version had items in a random order with the first item measuring U; the sixth version had items in a random order with the first item measuring EOU. The two versions (U first and EOU first) each of the grouped, alternating, and random item questionnaires were required to control for any possible priming effects (Posner, 1978) caused by responses to the first question (or construct).

3. Results—Experiment 1

Reliability (Cronbach alpha) was assessed for the U and EOU scales for all three treatment groups (grouped, alternating, and random). The reliability coefficient exceeded 0.95 for all three groups for both scales (Table 2). To assess convergent and discriminant validity, we performed a factor analysis (with direct oblimin rotation which allows the factors to be correlated) on all three groups (Table 3). All

TABLE 3
Factor analyses of scales—Experiment 1

	Grouped		Intermixed			
			Alternating		Random	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
U ₁	0.99	-0.03	0.91	0.00	0.94	-0.05
U ₂	0.97	-0.02	0.92	0.06	0.86	0.10
U ₃	0.99	-0.03	0.98	-0.02	0.92	0.06
U ₄	0.92	0.07	0.96	-0.04	0.96	-0.03
EOU ₁	-0.08	0.98	0.05	0.88	0.22	0.78
EOU ₂	-0.03	0.96	0.02	0.94	-0.18	0.91
EOU ₃	0.06	0.94	-0.01	0.97	0.07	0.88
EOU ₄	0.11	0.89	-0.05	0.97	0.18	0.83
Eigenvalue	5.93	1.50	5.11	2.00	5.30	1.42
% Variance	74.2	18.8	63.9	24.9	66.2	17.8

Note: U₁ to U₄: perceived usefulness items. EOU₁ to EOU₄: perceived ease of use items. $n = 60$ (grouped items). $n = 61$ (alternating items). $n = 61$ (random items). Principal components with oblimin rotation.

cross-loadings in both groups were under 0.25 for both scales. The clear discrimination of the two constructs in the grouped, alternating, and random conditions indicates that the grouping of items did not cause any artifactual discriminant validity.

To further examine convergent and discriminant validity, interitem correlations were computed. Tables 4(a), 4(b), and 4(c) list the interitem correlations for the grouped, alternating, and random conditions, respectively. The interitem correlations of items within U and within EOU were significantly greater than correlations between U and EOU items for all three groups.

4. Method—Experiment 2

In addition to assessing the reliability and validity of the scales, Experiment 2 examined any possible effect of item intermixing on the estimated path coefficients linking BI, U, and EOU (i.e. nomological validity). Two hundred and twenty students at the University of Minnesota participated in this study by filling out one of seven versions of a 10-item questionnaire assigned at random. The participants were enrolled in an information systems course and filled out the questionnaire in a classroom setting. Participants in this study were different from the participants in the first experiment (they were students enrolled in different sections of the same course as the participants of the first experiment). All participants possessed a working knowledge of the target system, WordPerfect. Two hundred and fourteen usable responses were obtained.

All versions of the questionnaire had 10 items (two measuring BI, four measuring U, and four measuring EOU). The items used are listed in Appendix 1. The experimental design is given in Table 5. The first four versions served the purpose of control. The first version of the questionnaire had two items measuring BI, followed by four items measuring U, followed by four items measuring EOU; the second version of the questionnaire had two items measuring BI, followed by four items measuring EOU, followed by four items measuring U. These two versions were presented with titles for each of the constructs. The third and fourth versions were similar to the first and second versions respectively, the only difference being that the third and fourth versions did not carry construct titles. These are the typical four formats of the TAM instrument used by researchers. Following Ajzen and Fishbein (1975: appendix A), questions about BI appear first in the original TAM instrument in order to reduce the possibility that respondents base their BI responses on implicit theories (e.g. DeNisi & Pritchard, 1978) linking BI to previous responses to U and EOU. For experimental purposes, three versions of the questionnaire—the fifth, sixth, and seventh versions—had items in an intermixed order with the first item on the questionnaire measuring BI, U, and EOU, respectively. This allowed us to control and test for any possible priming effects. Also, in these three intermixed versions of the questionnaire, no two items related to a particular construct were placed together.

Small sub-samples of subjects filling out different versions of the questionnaire were asked to provide a retrospective verbal evaluation of the questionnaire. These

TABLE 4
 (a) *Interitem correlations across constructs (grouped items)—Experiment 1*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	3.75	1.73								
U ₂	3.55	1.85	0.94							
U ₃	3.62	1.80	0.92	0.96						
U ₄	3.50	1.86	0.92	0.93	0.91					
EOU ₁	3.77	1.71	0.48	0.50	0.44	0.54				
EOU ₂	3.98	1.74	0.50	0.54	0.51	0.54	0.82			
EOU ₃	3.67	1.77	0.55	0.61	0.58	0.63	0.88	0.90		
EOU ₄	3.78	1.72	0.59	0.63	0.60	0.64	0.85	0.87	0.94	

Note: $n = 60$.

(b) *Interitem correlations across constructs (alternating items)—Experiment 1*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	4.44	1.77								
U ₂	4.23	1.73	0.78							
U ₃	4.31	1.53	0.87	0.92						
U ₄	4.15	1.60	0.79	0.93	0.89					
EOU ₁	3.85	1.77	0.37	0.37	0.41	0.39				
EOU ₂	4.10	1.79	0.45	0.45	0.41	0.32	0.81			
EOU ₃	4.02	1.87	0.38	0.41	0.39	0.38	0.81	0.88		
EOU ₄	4.05	1.79	0.31	0.42	0.35	0.36	0.77	0.88	0.92	

Note: $n = 61$.

(c) *Interitem correlations across constructs (random items)—Experiment 1*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	4.22	1.65								
U ₂	4.18	1.67	0.81							
U ₃	4.37	1.56	0.82	0.93						
U ₄	3.89	1.85	0.84	0.88	0.91					
EOU ₁	3.72	1.75	0.38	0.38	0.48	0.32				
EOU ₂	3.92	1.78	0.42	0.42	0.38	0.34	0.84			
EOU ₃	3.96	1.98	0.31	0.36	0.37	0.39	0.88	0.87		
EOU ₄	4.04	1.72	0.36	0.37	0.37	0.38	0.86	0.93	0.93	

Note: $n = 61$.

randomly selected subjects were asked to “please tell me your reaction to the questionnaire” immediately after filling it out, and their comments were tape recorded and transcribed.

TABLE 5
Experimental design—Experiment 2

	Item ordering		
	Grouped		
	Grouped/titled	Grouped/untitled	Intermixed
2 BI, 4 U, 4 EOU	<i>n</i> = 30	<i>n</i> = 30	
2 BI, 4 EOU, 4 U	<i>n</i> = 30	<i>n</i> = 30	
Total	<i>n</i> = 60	<i>n</i> = 60	
BI first			<i>n</i> = 31
U first			<i>n</i> = 32
EOU first			<i>n</i> = 31
Total			<i>n</i> = 94

Note: *n* = 214.

5. Results—Experiment 2

Reliability (Cronbach alpha) was assessed for all three scales (BI, U, and EOU) for all three treatment groups (grouped/titled, grouped/untitled, and intermixed). The reliability coefficient exceeded 0.90 for all three scales within each of the three treatments, and exceeded 0.95 in seven out of nine cases (Table 6). Convergent and discriminant validity of the U and EOU scales were assessed using factor analyses with direct oblimin rotation on all three groups (Table 7). All cross-loadings in both groups were under 0.30 for both scales. The clear discrimination of the two constructs in the grouped/titled, grouped/untitled and intermixed groups indicates that the grouping of items did not cause any artifactual discriminant validity.

To further examine convergent and discriminant validity, interitem correlations were computed. Tables 8(a), 8(b), and 8(c) list the interitem correlations for the grouped/titled, grouped/untitled, and intermixed groups, respectively. The

TABLE 6
Reliability of perceived usefulness and ease of use scales—Experiment 2

	Item ordering		
	Grouped		
	Grouped/titled	Grouped/untitled	Intermixed
BI	0.99	0.98	0.98
U	0.97	0.97	0.96
EOU	0.95	0.91	0.94

Note: *n* = 60 (grouped/titled items). *n* = 60 (grouped/untitled items). *n* = 94 (intermixed items).

TABLE 7
Factor analyses of scales—Experiment 2

	Grouped					
	Grouped/titled		Grouped/untitled		Intermixed	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
U ₁	0.93	0.06	0.91	0.06	0.93	0.04
U ₂	0.95	0.02	1.03	−0.11	1.00	−0.05
U ₃	1.00	−0.08	0.96	0.01	0.99	−0.03
U ₄	0.92	0.03	0.86	0.14	0.88	0.08
EOU ₁	0.04	0.93	0.00	0.86	0.00	0.91
EOU ₂	0.04	0.90	−0.11	0.93	−0.06	0.93
EOU ₃	0.00	0.95	0.06	0.88	0.02	0.92
EOU ₄	−0.06	0.99	0.27	0.73	0.08	0.88
Eigenvalue	5.59	1.64	5.51	1.36	5.51	1.52
% Variance	69.9	20.5	69.0	17.0	68.8	19.0

Note: U₁ to U₄: perceived usefulness items. EOU₁ to EOU₄: perceived ease of use items. $n = 60$ (grouped/titled items). $n = 60$ (grouped/untitled items). $n = 94$ (intermixed items). Principal components with oblimin rotation.

interitem correlations of items within U and within EOU were significantly greater than correlations between U and EOU items for all three groups.

To study the structural relationships among BI, U, and EOU, we performed regressions on the data pooled across all versions of the questionnaire (Table 9). Since the versions were counterbalanced for titles and order of construct, this allowed us to examine the effect of intermixing of items. A dummy variable, INTERMIX (0 when the items were not intermixed; 1 when the items were intermixed), was included. INTERMIX as a main effect and as an interaction with U and EOU had no significant effect on BI. This indicates that grouping vs. intermixing had no effect on the U-BI or EOU-BI structural relationships, nor did it have any direct effect on BI itself. Similarly, INTERMIX and its interaction with EOU did not have a significant impact on the EOU-U relationship.

The non-significance of these t -tests does raise the issue of accepting the null hypothesis that the regression terms are not significant. A power analysis was therefore conducted to assess the probability of type II error occurring (i.e. a significant difference existing in the proportion of variance explained that went undetected). We posited that a change in R^2 as small as 0.05 needs to be detected. A power analysis (see Cohen, 1987) revealed that our tests had a power of 0.95 in detecting a 0.05 change in the proportion of variance explained (i.e. only 5% of the time would we have failed to detect a change of 0.05 that actually exists). Such a low type II error emphasizes the robustness of the findings.

While statistical tests revealed no significant effect of item intermixing vs. grouping on the psychometric properties of the TAM scales, the retrospective verbal evaluations from subjects (Table 10) reveal some interesting qualitative effects.

TABLE 8
 (a) *Interitem correlations across constructs (grouped/titled items)—Experiment 2*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	3.62	1.75								
U ₂	3.58	1.64	0.92							
U ₃	3.91	1.56	0.87	0.91						
U ₄	3.55	1.67	0.88	0.84	0.89					
EOU ₁	3.79	1.79	0.37	0.33	0.36	0.46				
EOU ₂	4.02	1.66	0.39	0.33	0.48	0.41	0.88			
EOU ₃	4.17	1.57	0.42	0.34	0.55	0.39	0.87	0.93		
EOU ₄	3.99	1.84	0.38	0.32	0.45	0.34	0.81	0.91	0.88	

Note: $n = 60$.

(b) *Interitem correlations across constructs (grouped/untitled items)—Experiment 2*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	4.11	1.79								
U ₂	4.01	1.75	0.81							
U ₃	4.02	1.98	0.82	0.91						
U ₄	3.76	1.76	0.89	0.95	0.83					
EOU ₁	3.92	1.93	0.41	0.31	0.37	0.36				
EOU ₂	4.01	1.72	0.43	0.42	0.38	0.33	0.83			
EOU ₃	4.01	1.71	0.51	0.43	0.32	0.39	0.88	0.84		
EOU ₄	4.04	1.71	0.38	0.52	0.33	0.39	0.91	0.92	0.93	

Note: $n = 60$.

(c) *Interitem correlations across constructs (intermixed items)—Experiment 2*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	4.02	1.81								
U ₂	4.03	1.75	0.77							
U ₃	4.33	1.64	0.81	0.88						
U ₄	4.06	1.61	0.91	0.91	0.86					
EOU ₁	4.01	1.79	0.49	0.32	0.42	0.39				
EOU ₂	3.94	1.99	0.47	0.45	0.44	0.35	0.88			
EOU ₃	3.91	1.68	0.35	0.45	0.53	0.35	0.86	0.85		
EOU ₄	4.13	1.66	0.38	0.44	0.36	0.36	0.91	0.81	0.80	

Note: $n = 94$.

While subjects appeared to react very favorably to the grouped/titled version of questionnaire, and somewhat favorably to the grouped/untitled version of the questionnaire, intermixing of items tended to confuse and irritate them.

TABLE 9
Regressions—Experiment 2

Dependent variable	R^2	Independent variable	b	S.E. (b)	Path weight	t -statistic
BI	0.44	U	0.59	0.11		
		EOU	0.41	0.11	0.53	7.65***
		Constant	0.80	0.84	0.24	6.82***
BI	0.48	U	0.57	0.07		4.51***
		EOU	0.32	0.08	0.51	8.00***
		INTERMIX	-0.19	0.19	0.26	4.05***
		U X INTERM	0.19	1.23	-0.05	-1.05
		EOU X INTER	0.22	1.24	0.04	0.43
		Constant	0.58	0.16	0.05	0.51
U	0.23	EOU	0.86	0.33		3.60***
		Constant	0.94	1.32	0.22	2.61**
U	0.26	EOU	0.76	0.35		0.41
		INTERMIX	0.23	1.41	0.19	2.99**
		EOU X INTER	0.27	1.11	0.02	0.33
		Constant	0.87	1.22	0.08	0.58

Note: Although titles and order of construct were counterbalanced in the design, the data was pooled only after using dummy variables (for titles and order of construct) and determining that the main effects and interactions of these dummy variables were not significant.

BI: Behavioral intention; U: perceived usefulness; EOU: perceived ease of use; INTERMIX: dummy variable to indicate grouped vs. intermixed. $n = 214$.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

6. Method—Experiment 3

Experiment 3 replicated the above two experiments and addressed two additional key issues in order to more conclusively rule out possible measurement biases in TAM. First, it is possible that the length of questionnaire in the first two experiments (eight or 10 items) did not allow measurement biases to surface. Second, the validity of the findings in the first two experiments need to be assessed in the context of multiple behaviors. Such an assessment would allow a more direct comparison of this adaptation of the Theory of Reasoned Action to Budd's (1987) application of the Theory of Reasoned Action to health-related behaviors. Moreover, examining multiple behaviors affords the opportunity to perform multitrait-multimethod analyses of convergent and discriminant validity.

Three hundred and twenty-two students at Temple University participated in this study by filling out one of the 10 versions of a 20-item questionnaire assigned at random. By virtue of the participants being from a different university from the participants in experiments 1 and 2, all participants were different from those who

TABLE 10
Verbal evaluations by respondents—Experiment 2

Grouped—titled
Neat stuff. I don't like WordPerfect but the questionnaire was pretty well-structured. OK, I guess. It was like a job application form, dude. Pretty cool with the titles. They just wanted to know what I thought of WordPerfect and whether I planned on using it. It was pretty easy to read. No problems. At least not a pain in the butt. Boy, this was one of the better questionnaires I filled out this year, man. At least this one was so that I could do it quick.
Grouped—untitled
It was OK, I guess. They wanted to know what I thought about WordPerfect. No big deal. It's good to know that someone out here can write a decent bunch of questions to be asked. I don't know. It was OK. Cool stuff, I like filling out questionnaires that flow well. It's fun.
Intermixed
It was a terrible survey. It seemed like they were trying to trick me into something. You need to give them feedback on how to construct a better questionnaire. I can't understand why they were trying to confuse me like that. I don't know. They just kept jumping back and forth.

participated in the first two experiments. The participants were enrolled in an information systems course. In order to promote the formation of intentions and beliefs in long-term memory (see Feldman & Lynch, 1988), participants were allowed to use the target systems, WordPerfect and Lotus, for a period of 3 months before the administration of the questionnaires. Allowing subjects to use the target systems for longer periods of time follows the suggestion of DeNisi and Pritchard (1978) that subjects should work together for longer periods of time in order to form more accurate judgments. Three hundred and twelve usable responses were obtained.

The experimental design is given in Table 11. The first four versions of the questionnaire employed grouped items for control purposes. The first version of the questionnaire had two items measuring BI (WordPerfect), followed by four items measuring U (WordPerfect), followed by four items measuring EOU (WordPerfect), followed by the items measuring perceptions towards Lotus in the same order; the second version had two items measuring BI (WordPerfect), followed by four items measuring EOU (WordPerfect), followed by four items measuring U (WordPerfect), followed by items measuring perceptions towards Lotus in the same order. The third and fourth versions were similar to the first and second respectively, the only difference being that the Lotus items came before WordPerfect items. This helped

TABLE 11
Experimental design—Experiment 3

	Item ordering	
	Grouped	Unmixed
WordPerfect first (2 BI, 4 U, 4 EOU)	$n = 30$	
WordPerfect first (2 BI, 4 EOU, 4 U)	$n = 31$	
Lotus first (2 BI, 4 U, 4 EOU)	$n = 31$	
Lotus first (2 BI, 4 EOU, 4 U)	$n = 31$	
Total	$n = 123$	
WordPerfect BI first		$n = 31$
WordPerfect U first		$n = 31$
WordPerfect EOU first		$n = 31$
Lotus BI first		$n = 32$
Lotus U first		$n = 32$
Lotus EOU first		$n = 32$
Total		$n = 189$

$n = 312$.

control for order of systems. There were six experimental versions of the questionnaire. All these versions had items related to the two different systems in an intermixed order, and no two items related to the same construct appeared together. The fifth, sixth and seventh versions had items related to WordPerfect appearing as the first item on the questionnaire. The construct which the first item measured was BI, U, and EOU respectively. The eighth, ninth and tenth versions were similar to the fifth, sixth and seventh versions respectively; the only difference was that the first item pertained to Lotus.

In this experiment also, verbal evaluations were collected from sub-samples of respondents. As in the previous experiment, the objective of collecting these verbal protocols was to understand the reactions of subjects towards the different formats of questionnaires.

7. Results—Experiment 3

Reliability (Cronbach alpha) was assessed for all three scales (BI, U, and EOU) for both treatment conditions (grouped and intermixed) for both systems (WordPerfect and Lotus). The reliability coefficient was over 0.90 for both conditions for all three scales (Table 12). Convergent and discriminant validity of the U and EOU scales were assessed using factor analyses with direct oblimin rotation for both conditions and both systems [Tables 13(a) and 13(b)]. All cross-loadings in both groups were under 0.20 for both scales. The clear discrimination of the two constructs in the grouped and intermixed conditions indicates that the grouping of items did not cause any artifactual discriminant validity.

To further examine the convergent and discriminant validity, interitem correlations were computed [Tables 14(a) and 14(b) for WordPerfect, Tables 14(c) and 14(d)]. The interitem correlation of items within each construct (i.e. within U,

TABLE 12
Reliability of scales—Experiment 3

	Item ordering	
	Grouped	Intermixed
WordPerfect		
BI	0.91	0.96
U	0.98	0.97
EOU	0.91	0.90
Lotus		
BI	0.97	0.96
U	0.99	0.98
EOU	0.96	0.94

Note: $n = 123$ (grouped). $n = 189$ (intermixed).

within EOU) was significantly greater than correlations between U and EOU items for both conditions (grouped and intermixed) for WordPerfect and Lotus.

One key advantage of having two systems was to be able to perform a multitrait–multimethod (MTMM) analysis of convergent and discriminant validity (Campbell & Fiske, 1959). An MTMM analysis in this case has two correlation matrices for each of the two treatment conditions (grouped and intermixed). For the data from the grouped questionnaire formats, the first is a correlation matrix among the U items across the two systems [Table 15(a)], and the second is a correlation matrix among the EOU items across the two systems [Table 15(b)]. Similarly, Tables 15(c) and 15(d) present correlation matrices for the data from the intermixed questionnaire formats. The correlation matrices revealed that the U items correlated higher with other U items of the same system (monotrait correlations) than with the U items of the other system (heterotrait correlations). Likewise, the EOU items correlated higher within system than across systems. The same pattern held equally true for grouped and intermixed conditions. This demonstrates the robustness of the scales in terms of discriminating across systems.

For each of the two systems, the structural relationships among BI, U, and EOU were studied using a series of regressions on the data pooled across all versions of the questionnaire, counterbalanced for order of construct and order of system (Table 16—WordPerfect; Table 17—Lotus). A dummy variable, INTERMIX (0 when the items were not intermixed; 1 when the items were intermixed), was included. INTERMIX was neither significant as a main effect nor as an interaction with U and EOU (for WordPerfect and Lotus). This indicates that grouping vs. intermixing had no effect on the U-BI or EOU-BI structural relationships, nor did it have any direct effect on BI itself. Similarly, neither INTERMIX nor its interaction with EOU was significantly linked to U. A power analysis revealed that the power of our tests exceeded 0.995 for detecting a 0.05 change in the proportion of variance explained (i.e. in fewer than five cases in 1000 would we have failed to detect a change of 0.05 that actually exists).

TAM hypothesizes that system characteristics influence BI entirely indirectly via their effects on EOU and U. As part of the nomological validation, we sought to

TABLE 13
(a) *Factor analyses of scales (WordPerfect)—Experiment 3*

	Grouped		Intermixed	
	Factor 1	Factor 2	Factor 1	Factor 2
U ₁	0.97	0.01	0.90	0.10
U ₂	0.92	0.10	0.94	0.02
U ₃	0.99	-0.02	0.98	0.00
U ₄	0.99	-0.05	0.97	-0.05
EOU ₁	0.13	0.74	0.12	0.81
EOU ₂	-0.12	0.92	-0.16	0.96
EOU ₃	0.05	0.92	0.12	0.86
EOU ₄	0.00	0.92	0.18	0.84
Eigenvalue	4.95	2.00	5.36	1.49
% Variance	61.8	25.0	67.0	18.7

Note: U₁ to U₄: perceived usefulness items. EOU₁ to EOU₄: perceived ease of use items. $n = 123$ (grouped). $n = 189$ (intermixed). Principal components with oblimin rotation.

(b) *Factor analyses of scales (Lotus)—Experiment 3*

	Grouped		Intermixed	
	Factor 1	Factor 2	Factor 1	Factor 2
U ₁	0.98	0.02	0.94	0.07
U ₂	1.00	-0.02	1.00	-0.05
U ₃	0.98	0.00	0.97	0.01
U ₄	0.98	0.01	0.96	-0.01
EOU ₁	0.00	0.93	0.03	0.90
EOU ₂	-0.06	0.96	-0.08	0.95
EOU ₃	0.03	0.96	-0.03	0.98
EOU ₄	0.04	0.94	0.12	0.87
Eigenvalue	5.12	2.37	4.99	2.21
% Variance	5.12	2.37	4.99	2.22
	64.0	29.6	62.3	27.8

Note: U₁ to U₄: perceived usefulness items. EOU₁ to EOU₄: perceived ease of use items. $n = 123$ (grouped). $n = 189$ (intermixed). Principal components with oblimin rotation.

test whether the system-EOU and system-U relationships were affected by item intermixing. As part of this assessment, we performed the regressions needed to verify that U and EOU indeed mediate the system-BI relationship (Table 18). The data was pooled across systems and the dummy variable SYSTEM (0 for WordPerfect, 1 for Lotus) was introduced to represent system characteristics. Regressions 1 through 4 confirm that item intermixing, represented by the dummy variable INTERMIX, had no effect on the SYSTEM-U, SYSTEM-EOU, or EOU-U relationships. Regressions 5–7 were performed to verify the TAM hypothesis that EOU and U together mediate the SYSTEM-BI relationship. Regression 8 demonstrates that INTERMIX does not moderate the EOU-BI and

TABLE 14
 (a) *Interitem correlations across constructs (grouped WordPerfect items)—
 Experiment 3*

	Mean	S.D.	Usefulness				Ease of use				
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄	
U ₁	4.01	1.77									
U ₂	3.85	1.49	0.89								
U ₃	3.91	1.65	0.81	0.89							
U ₄	3.96	1.96	0.88	0.92	0.92						
EOU ₁	3.91	1.50	0.33	0.52	0.36	0.31					
EOU ₂	3.64	1.66	0.39	0.44	0.34	0.48	0.87				
EOU ₃	3.87	1.76	0.51	0.41	0.39	0.44	0.81	0.91			
EOU ₄	4.18	1.61	0.37	0.46	0.30	0.46	0.83	0.88	0.89		

n = 123.

(b) *Interitem correlations across constructs (intermixed WordPerfect items)—
 Experiment 3*

	Mean	S.D.	Usefulness				Ease of use				
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄	
U ₁	4.12	1.71									
U ₂	4.11	1.93	0.81								
U ₃	4.01	1.73	0.82	0.91							
U ₄	3.91	1.63	0.91	0.89	0.84						
EOU ₁	3.85	1.55	0.33	0.57	0.39	0.42					
EOU ₂	4.01	1.67	0.35	0.35	0.39	0.44	0.87				
EOU ₃	4.11	1.81	0.37	0.42	0.35	0.41	0.86	0.92			
EOU ₄	3.95	1.78	0.51	0.43	0.41	0.51	0.91	0.93	0.81		

n = 189.

U-BI relationships in the pooled data, consistent with the analyses performed by individual systems (Tables 16 & 17). As before, the power was found to exceed 0.995 in this case also, almost certainly ruling out type II errors (i.e. fewer than five chances in 1000).

The verbal evaluations (Table 19) added support to the verbal reports in the second experiment: the intermixed versions of the questionnaire seemed to annoy respondents.

8. Discussion

8.1 MEASUREMENT ISSUES

Results across all three experiments consistently indicate that the strong psychometric properties historically observed using the TAM instrument are *not* an artifact of grouping together the multiple items measuring a single construct. We found that intermixing (vs. grouping) items produced neither a significant degradation in the

TABLE 14
(c) *Interitem correlations across constructs (grouped Lotus Items)—Experiment 3*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	3.91	1.67								
U ₂	3.95	1.66	0.91							
U ₃	3.92	1.55	0.91	0.91						
U ₄	3.61	1.88	0.88	0.87	0.79					
EOU ₁	4.11	1.77	0.41	0.44	0.41	0.45				
EOU ₂	3.96	1.63	0.42	0.46	0.50	0.43	0.89			
EOU ₃	3.97	1.61	0.44	0.48	0.51	0.33	0.78	0.88		
EOU ₄	4.18	1.62	0.49	0.43	0.44	0.54	0.84	0.84	0.82	

n = 123.

(d) *Interitem correlations across constructs (intermixed Lotus items)—Experiment 3*

	Mean	S.D.	Usefulness				Ease of use			
			U ₁	U ₂	U ₃	U ₄	EOU ₁	EOU ₂	EOU ₃	EOU ₄
U ₁	3.80	1.68								
U ₂	3.81	1.60	0.81							
U ₃	3.50	1.68	0.83	0.91						
U ₄	3.92	1.70	0.91	0.88	0.86					
EOU ₁	3.45	1.71	0.33	0.44	0.44	0.44				
EOU ₂	3.63	1.54	0.34	0.38	0.46	0.51	0.88			
EOU ₃	3.82	1.64	0.33	0.36	0.46	0.58	0.83	0.87		
EOU ₄	3.81	1.67	0.35	0.44	0.41	0.44	0.81	0.82	0.81	

n = 189.

high reliability and validity of the TAM scales, nor a change in the path coefficients linking perceived usefulness and ease of use to usage intentions. Although the tests that were performed to demonstrate that item intermixing had no effect on the path coefficients entailed accepting the null (no effect) hypothesis, power analyses indicated that the tests were adequately powerful to detect small effect sizes (power was 0.95 in Experiment 2 and exceeded 0.995 in Experiment 3). To put this high statistical power in perspective, Baroudi and Orlikowski (1989) found no studies possessing statistical power in excess of 0.9 for detecting small effect sizes in their review of 149 studies published in leading MIS journals. The only observed effects of questionnaire format were the open-ended verbal evaluations, which indicated that subjects filling out the intermixed questionnaires were confused and annoyed, in contrast to those filling out the grouped questionnaires, who tended to express positive evaluations of the instrument.

The present findings disagree strongly with those of Budd (1987), who found that item intermixing had substantial deleterious effects on the reliability, validity, and estimated path coefficients for an application of the Theory of Reasoned Action to health-related behaviors. What could account for this discrepancy? Budd (1987) argues that the high reliability and strong empirical relationships observed using

TABLE 15
 (a) *Multitrait-multimethod analysis (grouped perceived usefulness items—
 Experiment 3)*

	WordPerfect				Lotus			
	WPU ₁	WPU ₂	WPU ₃	WPU ₄	LU ₁	LU ₂	LU ₃	LU ₄
WPU ₁								
WPU ₂	0.91							
WPU ₃	0.86	0.82						
WPU ₄	0.93	0.83	0.91					
LU ₁	0.42	0.39	0.32	0.40				
LU ₂	0.43	0.38	0.36	0.42	0.88			
LU ₃	0.38	0.44	0.38	0.39	0.86	0.91		
LU ₄	0.41	0.41	0.43	0.44	0.81	0.82	0.88	

Note: $n = 123$.

(b) *Multitrait-multimethod analysis (grouped perceived ease of use items—
 Experiment 3)*

	WordPerfect				Lotus			
	WPEOU ₁	WPEOU ₂	WPEOU ₃	WPEOU ₄	LEOU ₁	LEOU ₂	LEOU ₃	LEOU ₄
WPEOU ₁								
WPEOU ₂	0.91							
WPEOU ₃	0.86	0.88						
WPEOU ₄	0.81	0.83	0.82					
LEOU ₁	0.41	0.36	0.45	0.44				
LEOU ₂	0.44	0.42	0.41	0.43	0.88			
LEOU ₃	0.43	0.39	0.39	0.39	0.82	0.80		
LEOU ₄	0.38	0.38	0.37	0.52	0.91	0.87	0.80	

Note: $n = 123$.

grouped items is evidence that people were constructing attitudes in reaction to the measurement process where none existed before. Feldman and Lynch (1988) similarly theorize that carryover effects would worsen when the target constructs do not already exist in long-term memory. However, Budd's (1987) investigation dealt with what for most people would be expected to be familiar behaviors (smoking, brushing teeth, exercising). Thus, most people should have somewhat well-formed attitudes in long-term memory regarding these behaviors. Our open-ended verbal evaluations indicating that subjects found the intermixed format confusing and annoying suggest an alternative explanation for Budd's (1987) findings. By continuously shifting from construct to construct and behavior to behavior, the intermixed format may engender a process referred to as "output interference" (Hoch, 1984; Alba & Chattopadhyay, 1985) in which "cognitions activated in the process of making the first judgment may suppress the retrieval of cognitions that would have influenced the second judgment or behavior in the absence of prior measurement" (Feldman & Lynch, 1988: p. 422). Assuming that Budd (1987) included as few as two modal beliefs and salient referents, his questionnaire asked

TABLE 15
 (c) *Multitrait–method analysis (intermixed perceived usefulness items)—*
Experiment 3

	WordPerfect				Lotus			
	WPU ₁	WPU ₂	WPU ₃	WPU ₄	LU ₁	LU ₂	LU ₃	LU ₄
WPU ₁								
WPU ₂	0.87							
WPU ₃	0.89	0.87						
WPU ₄	0.91	0.86	0.92					
LU ₁	0.47	0.31	0.38	0.51				
LU ₂	0.48	0.47	0.40	0.47	0.92			
LU ₃	0.47	0.57	0.44	0.41	0.88	0.83		
LU ₄	0.37	0.42	0.45	0.52	0.91	0.84	0.85	

Note: $n = 189$.

(d) *Multitrait–multimethod analysis (intermixed perceived ease of use items)—*
Experiment 3

	WordPerfect				Lotus			
	WPEOU ₁	WPEOU ₂	WPEOU ₃	WPEOU ₄	LEOU ₁	LEOU ₂	LEOU ₃	LEOU ₄
WPEOU ₁								
WPEOU ₂	0.83							
WPEOU ₃	0.81	0.87						
WPEOU ₄	0.79	0.93	0.89					
LEOU ₁	0.35	0.47	0.36	0.37				
LEOU ₂	0.40	0.40	0.34	0.55	0.91			
LEOU ₃	0.42	0.37	0.31	0.46	0.86	0.87		
LEOU ₄	0.45	0.49	0.38	0.44	0.79	0.85	0.89	

Note: $n = 189$.

about each of nine constructs for each of three behaviors, plus two measures of motivation to comply which were behavior-invariant, for a total of 29 distinct constructs. Asking subjects to respond to questions that randomly alternate among 29 constructs may be sufficient to bring about a degree of output interference that would substantially degrade the psychometric properties of the scales. Budd (1987: p. 104) acknowledges that the random ordering may have increased measurement error. In Experiment 3, above, we examined three constructs (BI, U and EOU) across two systems (WordPerfect and Lotus) for a total of six distinct constructs.

The verbal evaluations of questionnaire formats by respondents highlights the extent of subjective differences that could be detected by looking into the “black box” (e.g. Todd & Benbasat, 1987). While the intermixed version of the questionnaire did not degrade psychometric properties in this case, the heightened confusion and annoyance reported verbally by subjects suggests a potential tendency toward output interference. Ironically, a procedure designed to improve measurement quality, item intermixing, appears to have the potential for actually *degrading*

TABLE 16
Regression (WordPerfect)—Experiment 3

Dependent variable	R^2	Independent variable	b	S.E. (b)	Path weight	t -statistic
BI	0.47	U	0.60	0.10	0.43	9.61***
		EOU	0.41	0.12	0.25	6.83***
		Constant	1.01	0.16		4.07***
BI	0.49	U	0.55	0.09	0.41	8.00***
		EOU	0.36	0.09	0.28	4.05***
		INTERMIX	-0.89	1.29	-0.09	-1.05
		U × INTERM	0.78	1.21	0.04	1.24
		EOU × INTER	0.22	1.64	0.08	1.56
		Constant	0.61	0.14		3.99***
U	0.20	EOU	0.41	0.09	0.28	5.23***
		Constant	1.04	1.06		0.95
U	0.23	EOU	0.51	0.11	0.25	4.19***
		INTERMIX	0.98	1.89	0.00	0.21
		EOU × INTER	0.55	0.87	0.11	1.03
		Constant	0.84	1.36		0.55

Note: Although order of construct and order of system were counterbalanced in the design, the data was pooled only after using dummy variables (for order of construct and order of system) and determining that the main effects and interactions of these dummy variables were not significant.

BI: Behavioral intention; U: perceived usefulness; EOU: perceived ease of use; INTERMIX: dummy variable to indicate grouped vs. intermixed. $n = 312$.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$.

measurement quality. Moreover, all else being equal, researchers should attempt to employ procedures that involve the fairest treatment of human subjects, and therefore should not annoy them, particularly when there is no methodological leverage to be gained. Although more research is clearly needed to better understand possible output interference and carryover effects, given the available evidence, we recommend that those who use TAM continue to use the original grouped format.

IT PRODUCT ASSESSMENT

TAM is one of the most powerful models explaining IT acceptance and use. It is a fairly rational perspective, which is evident from the key constructs in the model: perceived usefulness and perceived ease of use, and also from the basic psychological theory it was adapted from (i.e. Theory of Reasoned Action). While TAM does not cover all possible factors affecting product assessment, empirical research has revealed that the constructs of perceived usefulness and perceived ease of use compare favorably with alternative approaches such as: realism of expectations (Ginzberg, 1981), motivational force (DeSanctis, 1983), user satisfaction (Srinivasan, 1985), user information satisfaction and user involvement (Baroudi, Olson, & Ives, 1986), value (Swanson, 1987), theory of planned behavior (Mathieson, 1991), and theory of trying (Bagozzi, Davis, & Warshaw, 1992). The focus of this research has

TABLE 17
Regressions (Lotus)—Experiment 3

Dependent variable	R^2	Independent variable	b	S.E. (b)	Path weight	t -statistic
BI	0.49	U	0.97	0.20	0.41	6.22***
		EOU	0.89	0.16	0.24	4.76***
		Constant	1.01	0.06		10.00***
BI	0.51	U	1.02	0.19	0.42	6.07***
		EOU	1.06	0.19	0.24	7.02***
		INTERMIX	0.09	1.39	0.04	0.65
		U × INTERM	0.32	1.32	0.05	0.99
		EOU × INTER	0.21	0.87	0.11	1.21
		Constant	0.55	0.09		6.32***
U	0.27	EOU	0.55	0.19	0.20	6.13***
		Constant	0.05	2.07		0.00
U	0.30	EOU	0.48	0.21	0.18	5.25***
		INTERMIX	0.32	0.94	0.11	1.06
		EOU × INTER	0.10	1.52	0.02	0.56
		Constant	0.12	1.65		1.40

Note: Although order of construct and order of system were counterbalanced in the design, the data was pooled only after using dummy variables for order of construct and order of system) and determining that the main effects and interactions of these dummy variables were not significant.

BI: Behavioral intention; U: perceived usefulness; EOU: perceived ease of use; INTERMIX: dummy variable to indicate grouped vs. intermixed.

$n = 312$.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$.

been to examine biases in TAM measures rather than to compare (with other approaches) the power of TAM as a predictor of IT acceptance and use. However, the findings from several pieces of past research (has established TAM as one of the best models explaining IT acceptance and use) combined with the findings of our research (that concludes that TAM measures appear to be free of measurement biases) suggest that TAM is a robust model both from a theoretical framework and measurement standpoint that can be relied upon to study acceptance and use of information technologies. Further, TAM has been very successfully applied to the study of a wide range of issues related to acceptance including our most recent work (Venkatesh & Davis, 1995), which employed the model to examine the feasibility of early user acceptance of unfamiliar technologies.

These studies have important implications for research and practice. The findings offer reassurance to researchers who use the TAM instrument to study acceptance of emerging information technologies. TAM has been widely used by many researchers with minor modifications. The use of TAM to study issues related to IS acceptance is further recommended given the high reliability and validity of these scales. The present research employed subjects who were predetermined to be familiar with the target systems. Further research is needed to determine whether the findings would

TABLE 18
Regressions—TAM validation (data pooled across systems)

Dependent variable	R^2	Independent variable	b	S.E. (b)	Path weight	t -statistic
1. U	0.28	EOU	0.98	0.18	0.46	8.00***
		S	0.55	0.12	0.25	9.12***
		Constant	0.23	0.51		6.25***
2. U	0.31	EOU	0.88	0.21	0.44	10.61***
		S	0.67	0.21	0.24	5.11***
		INTERMIX	0.20	0.86	0.11	1.41
		EOU \times INTER	0.19	1.22	0.09	0.96
		S \times INTERMI	0.23	0.88	0.12	0.99
3. EOU	0.48	Constant	1.44	0.44		4.55***
		S	1.01	0.25	0.21	2.54*
4. EOU	0.52	Constant	1.44	0.05		9.33***
		S	0.88	0.16	0.27	4.24***
		INTERMIX	0.18	1.23	0.08	1.23
		S \times INTERMI	0.21	1.88	0.02	1.01
5. BI	0.53	Constant	1.23	0.08		10.22***
		S	0.81	0.06	0.55	11.55***
		Constant	1.11	1.45		0.43
6. BI	0.50	U	1.22	0.19	0.36	5.52***
		EOU	0.38	0.07	0.21	3.21***
		Constant	0.32	1.21		1.03
7. BI	0.52	U	0.55	0.14	0.39	3.45***
		EOU	0.55	0.21	0.24	4.61***
		S	0.16	0.98	0.05	0.56
		Constant	0.81	0.09		4.55***
8. BI	0.56	U	0.48	0.16	0.32	4.28***
		EOU	0.41	0.20	0.18	5.02***
		S	0.38	0.51	0.07	1.08
		INTERMIX	0.12	1.33	0.04	0.84
		U \times INTERMI	0.19	0.98	0.09	1.11
		EOU \times INTER	0.41	1.99	0.10	1.22
		S \times INTERMI	0.67	1.12	0.04	0.31
		Constant	1.11	0.12		6.08***

Note: Although order of construct and order of system were counterbalanced in the design, the data was pooled only after using dummy variables (for order of construct and order of system) and determining that the main effects and interactions of these dummy variables were not significant. BI: Behavioral intention; U: perceived usefulness; EOU: perceived ease of use; INTERMIX: dummy variable to indicate grouped vs. intermixed.

$n = 624$.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$.

generalize to situations involving unfamiliar systems. Also, in these three experiments, we used technologies that were already being widely used. Future research should examine the issue of whether these findings would generalize to situations involving “emerging technologies” as they emerge.

TABLE 19
Verbal evaluations by respondents—Experiment 3

Grouped
I have no complaints. I figure they wanna know what I think of those two softwares.
What can I say? No sweat. Questions about my feeling towards WordPerfect and Lotus. I wish I could have told them I dislike these DOS packages. Microsoft Works is the way to go.
Easy to fill out.
Made sense to me. Survey about WordPerfect and Lotus.
I did what they told me to do. Fill out the questionnaire. Went through quickly though it was a couple of pages. It had a sense of coherence and order. I liked it.
Intermixed
Why do you keep repeating yourself in the different questions? It's so annoying that you keep jumping back forth between WordPerfect and Lotus.
Can't you just finish with one system before going on to another?
It would make my life easier if you just put them in some damn coherent order.
I was just plain irritated as they kept jumping around.
It was OK. I just thought they could have done it better though. Maybe some titles, I don't know. Seemed like they were trying to test me.
Boring. Confused the heck out of me.

More broadly, the generalizability of our findings beyond TAM warrant further investigation. The findings do, however, reinforce the need for clear theoretical definition and careful selection of items that measure research construct(s) (see Bagozzi, 1980). Practitioners can continue to rely on these scales which have been significantly linked to actual system use. Since investment in emerging information technologies is in the order of billions of dollars, reliable and valid scales to predict and explain user acceptance are of critical importance. By ruling out carryover biases in these scales, this study is encouraging to practitioners who use these scales to evaluate risky investments in information technologies.

We wish to thank Bob VanCleave and Gerald Smith of the University of Minnesota and Khemdut Purang and Laurie Shteir of Temple University for their help in conducting the studies, and Barbara Klein and Dale Goodhue University of Minnesota for their comments on drafts of this paper. We also wish to thank Tracy Sykes for her tremendous efforts in editing the paper and improving its readability. Last but not least, we wish to express our sincerest thanks to the Associate Editor and two anonymous reviewers for their many useful comments and suggestions which helped improve the paper.

References

- ABELSON, R., ARONSON, E., MCGUIRE, W., NEWCOMB, T., ROSENBERG, M. & TANNENBAUM, P. (1968). *Theories of Cognitive Consistency: A Sourcebook*. Chicago, IL: Rand McNally.
- ADAMS, D. A., NELSON, R. R. & TODD, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: a replication. *MIS Quarterly*, **16**, 227–250.

- AJZEN, I. & FISHBEIN, M. (1980). *Understanding Attitudes and Predicting Social Behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
- ALBA, J. W. & CHATTOPADHYAY, A. (1985). The effects of context and part-category cues on the recall of competing brands. *Journal of Marketing Research*, **22**, 340–349.
- BAGOZZI, R. P. (1980). *Causal Models in Marketing*. New York, NY: John Wiley and Sons.
- BAGOZZI, R. P., DAVIS, F. D. & WARSHAW, R. P. (1992). Development and test of a theory of technological learning and usage. *Human Relations*, **45**, 659–686.
- BAROUDI, J. J. & ORLIKOWSKI, W. J. (1989). The problem of statistical power in MIS research. *MIS Quarterly*, **13**, 87–106.
- BAROUDI, J. J., OLSON, M. H. & IVES, B. (1986). An empirical study of the impact of user involvement on system usage and information satisfaction. *Communications of the ACM*, **29**, 232–238.
- BRADBURN, N. (1982). Question-wording effects in surveys. In R. HOGARTH, Ed. *Question Framing and Response Consistency*, pp. 65–76. San Francisco, CA: Jossey-Bass.
- BUDD, R. J. (1987). Response bias and the theory of reasoned action. *Social Cognition*, **5**, 95–107.
- BUELL, B. & BRANDT, R. (1990). The pen: computing's next big leap. *Business Week*, May 14, 128–129.
- BURGESS, J. (1991). Telecomputphoneos: the next step in the information rage. *The Washington Post Weekly Edition*, August 5–11, 20.
- CAMPBELL, D. T. & FISKE, D. W. (1989). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**, 81–105.
- CHIN, W. W. & GOPAL, A. (1993). An examination of the relative importance of four belief constructs on the GSS adoption decision: a comparison of four methods. *Proceedings of the 26th Hawaii International Conference on System Sciences*, pp. 548–557. Los Alamitos: IEEE Computer Society Press.
- COHEN, J. (1987). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum, 1987.
- DALY, J. (1993). Move over, Dick Tracy. *Forbes ASAP*, **151**, 48–51.
- DAVIS, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: theory and results*. Ph.D. dissertation, MIT Sloan School of Management, Cambridge, MA.
- DAVIS, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, **13**, 319–340.
- DAVIS, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, **38**, 475–487.
- DAVIS, F. D. & VENKATESH, V. (1995). Measuring user acceptance of emerging information technologies: an assessment of possible method biases. *Proceedings of the 28th Annual Hawaii International Conference on System Sciences*. pp. 729–736. Los Alamitos: IEEE Computer Society Press.
- DAVIS, F. D., BAGOZZI, R. P. & WARSHAW, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, **35**, 982–1003.
- DE NISI, A. S. & PRITCHARD, R. D. (1978) Implicit theories of performance as artifacts in survey research: a replication and extension. *Organizational Behavior and Human Performance*, **21**, 358–366.
- DESANCTIS, G. (1983). Expectancy theory as an explanation of voluntary use of a decision support system. *Psychological Reports*, **52**, 247–260.
- FAZIO, R. H., CHEN, J., MCDONEL, E. C. & SHERMAN, S. J. (1981). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, **18**, 339–357.
- FELDMAN, J. M. & LYNCH, J. G. JR. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, **73**, 421–435.
- FISHBEIN, M. & AJZEN, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

- GINZBERG, M. J. (1981). Early diagnosis of MIS implementation failure: promising results and unanswered questions. *Management Science*, **27**, 459–478.
- HALFHILL, T. R. (1993). PDAs arrive but aren't quite here yet. *Byte*, **18**, 66–94.
- HARRISON, D. A. & McLAUGHLIN, M. E. (1991). Exploring the cognitive processes underlying responses to self-report instruments: effects of item context on work attitude measures. *Academy of Management Best Paper Proceedings*, pp. 310–314. Miami Beach, FL.
- HEIDER, F. (1988). *The Psychology of Interpersonal Relations*. New York, NY: Wiley.
- HENDRICKSON, A. R., MASSEY, P. D. & CRONAN, T. P. (1993). On the test–retest reliability of perceived usefulness and perceived ease of use scales. *MIS Quarterly*, **17**, 227–230.
- HOCH, S. J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 649–662.
- HOGBERG, H. (1993). Wide-area wireless data: currently available systems and applications. *Telecommunications*, **December**, 33–35.
- KEEFE, P. (1991). Pen-based PCs poised for breakthrough. *Computerworld*, **25**, 111, 117.
- LEWIS, P. H. (1991). Multimedia's new world. *New York Times*, November 5.
- MARKOFF, J. (1991a). Mouse! movie! sound! action! *The New York Times*, October 27, 11.
- MARKOFF, J. (1991b). Getting computers to accept poor penmanship. *The New York Times*, October 16, C9.
- MATHIESON, K. (1991). Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. *Information Systems Research*, **2**, 173–191.
- MCCARROLL, T. (1994). War of the wireless. *Time*, March 14, 89.
- MILBURN, K. (1988). It's show time. *Infoworld*, June 6, **10**, 49–51; 54–61, 64, 65.
- MOORE, G. C. & BENBASAT, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, **2**, 192–222.
- NUNALLY, J. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.
- OLFMAN, L. & BOSTROM, R. P. (1991). End-user software training: an experimental comparison of methods to enhance motivation. *Journal of Information Systems*, **1**, 249–266.
- POSNER, M. (1978). *Chronometric Explorations of the Human Mind*. Hillsdale, NJ: Erlbaum.
- SCHUMAN, J. & PRESSER, S. (1981). Questions and answers in attitude surveys: experiments in question form, wording, and context. New York, NY: Academic Press.
- SEGARS, A. H. & GROVER, V. (1993). Re-examining perceived ease of use and usefulness: a confirmatory factor analysis. *MIS Quarterly*, **17**, 517–525.
- SJAZNA, B. (1994). Software evaluation and choice predictive validation of the technology acceptance instrument. *MIS Quarterly*, **18**, 319–324.
- SRINIVASAN, A. (1985). Alternative measures of system effectiveness: associations and implications. *MIS Quarterly*, **9**, 243–253.
- SWANSON, E. B. (1987). Information channel disposition and use. *Decision Sciences*, **18**, 131–145.
- TEDESCHI, J. (1981). *Impression Management Theory and Social Psychological Research*. New York, NY: Academic Press.
- TAYLOR, S. A. & TODD, P. A. (1995). Understanding information technology usage: a test of competing models. *Information Systems Research*, **6**, 144–176.
- TODD, P. A. & BENBASAT, I. (1987). Process tracing methods in decision support systems research: exploring the black box. *MIS Quarterly*, **11**, 493–512.
- TOURANGEAU, R. & RASINSKI, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, **103**, 299–314.
- TREVINO, L. K. & WEBSTER, J. W. (1992). Flow in computer-mediated communication: electronic mail and voice mail evaluation and impacts. *Communication Research*, **19**, 539–573.
- VENKATESH, V. & DAVIS, F. D. (1994). Modeling the determinants of perceived ease of use. *Proceedings of International Conference on Information Systems*, pp. 213–227. Vancouver, Canada, 227.

VENKATESH, V. & DAVIS, F. D. (1995). *Pre-prototype user acceptance testing: three validation experiments*. Working Paper, University of Minnesota, Minneapolis.

VENKATESH, V. & DAVIS, F. D. (forthcoming). A model of the antecedents of perceived ease of use: development and test. *Decision Sciences*.

Paper accepted for publication by Associate Editor, Professor A. Sutcliffe.

Appendix. Items used

INTENTION TO USE

Assuming I had access to WordPerfect, I intend to use it

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

Given that I had access to WordPerfect, I predict that I would use it.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

PERCEIVED USEFULNESS

Using Wordperfect improves my performance in my degree program.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

Using WordPerfect in my degree program increases my productivity.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

Using WordPerfect enhances my effectiveness in my degree program.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

I find WordPerfect useful in my degree program.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

PERCEIVED EASE OF USE

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

Interacting with WordPerfect does not require a lot of my mental effort.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

I find WordPerfect easy to use.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------

I find it easy to get WordPerfect to do what I want it to do.

Strongly Agree	+3	+2	+1	0	-1	-2	-3	Strongly Disagree
----------------	----	----	----	---	----	----	----	-------------------