Communications of the Association for Information Systems

Volume 4, Article 7
October 2000

STRUCTURAL EQUATION MODELING AND REGRESSION: GUIDELINES FOR RESEARCH PRACTICE

David Gefen
Management Department
LeBow College of Business
Drexel University
gefend@drexel.edu

Detmar W. Straub
Department of Computer Information Systems
Robinson College of Business
Georgia State University

Marie-Claude Boudreau
Management Information System Department
Terry College of Business
University of Georgia

TUTORIAL

STRUCTURAL EQUATION MODELING AND REGRESSION: GUIDELINES FOR RESEARCH PRACTICE

David Gefen
Management Department
LeBow College of Business
Drexel University

Detmar W. Straub
Department of Computer Information Systems
Robinson College of Business
Georgia State University

Marie-Claude Boudreau
Management Information System Department
Terry College of Business
University of Georgia

ABSTRACT

The growing interest in Structured Equation Modeling (SEM) techniques and recognition of their importance in IS research suggests the need to compare and contrast different types of SEM techniques so that research designs can be selected appropriately. After assessing the extent to which these techniques are currently being used in IS research, the article presents a running example which analyzes the same dataset via three very different statistical techniques. It then compares two classes of SEM: covariance-based SEM and partial-least-squares-based SEM. Finally, the article discusses linear regression models and offers guidelines as to when SEM techniques and when regression techniques should be used. The article concludes with heuristics and rule of thumb thresholds to guide practice, and a discussion of the extent to which practice is in accord with these guidelines.

Keywords: IS research methods; measurement; metrics; guidelines; heuristics; structural equation modeling (SEM); LISREL; PLS; regression; research techniques; theory development; construct validity; research rules of thumb and heuristics; formative constructs; reflective constructs.

Note: The paper is written in such a way that readers with basic knowledge of multivariate statistics can follow the logic and examples. It does not assume the reader is already conversant with LISREL, PLS, or other SEM tools. This tutorial contains:

- straightforward examples to illuminate more complex topics,
- a glossary whose entries are linked to the text, and
- a rudimentary structural model applying the Technology Acceptance Model (TAM) to e-Commerce. This model is analyzed in three ways: (1) PLS, (2) LISREL, and (3) linear regression.

Because of the large number of notes associated with this paper, they are presented as end notes at the end of this paper rather than as footnotes.

I. INTRODUCTION

Structural Equation Modeling (SEM) techniques such as LISREL[1] and Partial Least Squares (PLS) are second generation data analysis techniques [Bagozzi and Fornell, 1982] that can be used to test the extent to which IS research meets recognized standards for high quality statistical analysis. That is to say, they test for statistical conclusion validity [Cook and Campbell, 1979]. Contrary to first generation statistical tools such as regression, SEM enables researchers to answer a set of interrelated research questions in a

- single,
- systematic, and
- comprehensive analysis

by modeling the relationships among multiple independent and dependent constructs simultaneously [Gerbing and Anderson, 1988]. This capability for simultaneous analysis differs greatly from most first generation regression models such as linear regression, LOGIT, ANOVA, and MANOVA, which can analyze only one layer of linkages between independent and dependent variables at a time. This ability is demonstrated by the running example in this paper (Section II) that applies the Technology Acceptance Model (TAM) [Davis, 1989] to the problem of e-commerce acceptance.

FIRST GENERATION vs. SECOND GENERATION MODELS

SEM permits complicated variable relationships to be expressed through hierarchical or non-hierarchical, recursive or non-recursive structural equations, to present a more complete picture of the entire model [Bullock et al., 1994, Hanushek and Jackson, 1977]. In TAM [Davis, 1989], for example, the intention to use a new information technology is the product of two beliefs:

1. the perceived usefulness (PU) of using the IT and
2. the perceived ease of use of using it (EOU).

But TAM also posits that perceived usefulness depends upon ease of use. Using SEM, these three paths can be modeled in one analysis (Figure 1).

Using first generation regression models two unrelated analyses are required (H_1 and H_2 in one analysis and H_3 in a second analysis):

1. examining how items load on the constructs via factor analysis, and then,
2. a separate examination of the hypothesized paths, run independently of these factor loadings.

The intricate causal networks enabled by SEM characterize real-world processes better than simple correlation-based models. Therefore, SEM is more suited for the mathematical modeling of complex processes to serve both theory [Bollen, 1989] and practice [Dubin, 1976].

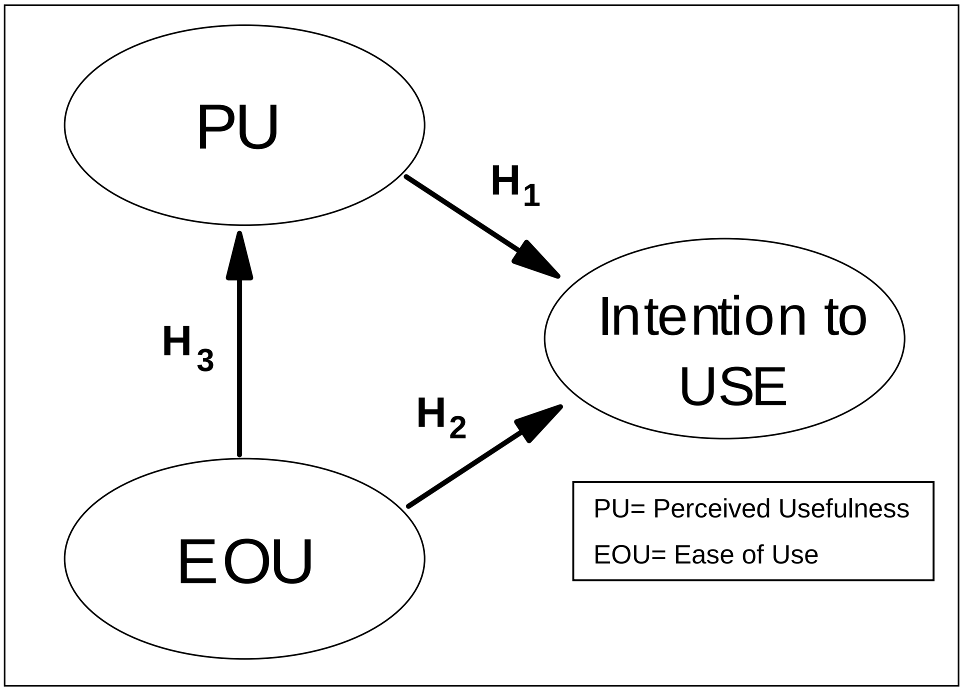


Figure 1. The TAM Model

Unlike first generation regression tools, SEM not only assesses

- the *structural model* – the assumed causation among a set of dependent and independent constructs – but, in the same analysis, also evaluates the
- *measurement model* – loadings of observed items (measurements) on their expected latent variables (constructs).

The combined analysis of the measurement and the structural model enables:

- measurement errors of the observed variables to be analyzed as an integral part of the model, and
- factor analysis to be combined in one operation with the hypotheses testing.

The result is a more rigorous analysis of the proposed research model and, very often, a better methodological assessment tool [Bollen, 1989, Bullock et al., 1994, Jöreskog and Sörbom, 1989].

Thus, in SEM, factor analysis and hypotheses are tested in the same analysis. SEM techniques also provide fuller information about the extent to which the research model is supported by the data than in regression techniques.

THE EXTENT TO WHICH SEM IS BEING USED

Not surprisingly, SEM tools are increasingly being used in behavioral science research for the causal modeling of complex, multivariate data sets in which the researcher gathers multiple measures of proposed constructs [Hair et al., 1998].[2] Indeed, even a casual glance at the IT literature suggests that SEM has become *de rigueur* in validating instruments and testing linkages between constructs.

Before describing in greater depth the methods and approaches adopted in SEM vis-à-vis regression, it is useful to know the extent to which SEM is currently being used in IS research. The results of analyzing techniques used in empirical articles in three major IS journals (*MIS Quarterly, Information & Management* and *Information Systems Research)* during the four year period between January 1994 and December 1997 are shown in Table 1. Consistent with Straub [1989], the qualifying criteria for the sample were that the article employed either:

- correlation or statistical manipulation of variables or
- some form of data analysis, even if the data analysis was simply descriptive statistics.

Studies using archival data (e.g., citation analysis) or unobtrusive measures (e.g., computer system accounting measures) were omitted from the sample unless it was clear from the methodological description that key variable relationships being studied could have been submitted to validation procedures. The number of articles published by each journal (n) and the percentage using SEM techniques are shown in the table. Most of the 171 articles selected were field studies (74%); the remainder were field experiments (6%), laboratory experiments (15%) and case studies (5%) that used quantitative data.

Table 1. Use of Structural Equation Modeling Tools 1994-1997

SEM Approaches	I&M (n=106)	*ISR* (n=27)	*MISQ* (n=38)	All Three Journals
PLS	2%	19%	11%	7%
LISREL	3%	15%	11%	7%
Other *	3%	11%	3%	4%
Total %	8%	45%	25%	18%

* Other includes SEM techniques such as AMOS and EQS.

Table 1 clearly shows that SEM has been used with some frequency for validating instruments and testing linkages between constructs in two of three widely known IS journals. In *ISR*, 45% of the positivist, empirically-based articles used SEM; in *MISQ*, it was 25%. From the first appearance of SEM in 1990 in the major IS journals [Straub, 1990], usage grew steadily. By the mid-1990's SEM was being used in about 18% of empirical articles across the three journals, with PLS and LISREL being the two most common techniques. Other SEM tools, such as EQS and AMOS, were used less often, but this is most likely because of the slowness of diffusion of innovation and is not a statement about the power or capability of these particular packages.

WHAT IS IN THIS PAPER

To help the reader understand the differences among LISREL, PLS, and linear regression, this article presents a running example of the analysis of a Technology Acceptance Model (TAM) dataset that uses these three statistical techniques. The running example begins in Section II. It can be skimmed or skipped by readers familiar with the three techniques.

Despite increased interest and the growing literature of individual SEM models, there is no comprehensive guide for researchers on when a specific form of SEM should be employed. To inform research practice and to explore the dimensions of the problem, Section III compares the two most widely used

SEM models in the IT literature: LISREL and PLS. PLS and LISREL represent the two distinct SEM techniques, respectively:

- partial-least-squares-based and
- covariance-based SEM,

In Section IV, the paper summarizes the major assumptions of the two SEM models. Based on this analysis, guidelines are presented in Section V for when to choose one of the two SEM models or one of the first generation regression models.

A summary of the major guidelines in Sections III, IV, and V, is presented below in Tables 2 and 3. Table 2 summarizes the objective behind each technique and limitations relating to sample size and distribution. A detailed discussion with citations on these issues can be found in Overview of Analytical Techniques in Section III. Table 3 summarizes guidelines based on the capabilities of each technique. These guidelines are discussed in detail and with citations in The SEM Model, also in Section III.

II. RUNNING EXAMPLE OF USE OF SEM VERSUS FIRST GENERATION STATISTICAL TECHNIQUES

For those IS researchers who are not familiar with SEM, this section presents a sample analysis of a typical dataset that uses the three techniques discussed in this article: [3]

1. linear regression
2. LISREL
3. PLS

Table 2. Comparative Analysis between Techniques

Issue	LISREL	PLS	Linear Regression
Objective of Overall Analysis	Show that the null hypothesis of the entire proposed model is plausible, while rejecting path-specific null hypotheses of no effect.	Reject a set of path-specific null hypotheses of no effect.	Reject a set of path-specific null hypotheses of no effect.
Objective of Variance Analysis	Overall model fit, such as insignificant χ^2 or high AGFI.	Variance explanation (high R-square)	Variance explanation (high R-square)
Required Theory Base	Requires sound theory base. Supports confirmatory research.	Does not necessarily require sound theory base. Supports both exploratory and confirmatory research.	Does not necessarily require sound theory base. Supports both exploratory and confirmatory research.
Assumed Distribution	Multivariate normal, if estimation is through ML. Deviations from multivariate normal are supported with other estimation techniques.	Relatively robust to deviations from a multivariate distribution.	Relatively robust to deviations from a multivariate distribution, with established methods of handling non-multivariate distributions.
Required Minimal Sample Size	At least 100-150 cases.	At least 10 times the number of items in the most complex construct.	Supports smaller sample sizes, although a sample of at least 30 is required.

TAM AS DOMAIN FOR RUNNING EXAMPLE

The domain of the running example is the Technology Acceptance Model (TAM), a widely researched theoretical model that attempts to explain the adoption of new information technologies. A partial listing of previous TAM studies, presented in Appendix A, shows the extent to which this model has been examined in IS research. TAM, based on the *Theory of Reasoned Action* [Ajzen and Fishbein, 1980, Fishbein and Ajzen, 1975], is a straightforward model of IT adoption that contends that beliefs such as system perceived usefulness (PU) and perceived ease-of-use (EOU) impact:

1. attitudes toward use,
2. intentions to use (IUSE), and ultimately
3. IT acceptance (most often measured as utilization).

Table 3. Capabilities by Research Approach

Capabilities	LISREL	PLS	Regression
Maps paths to many dependent (latent or observed) variables in the same research model and analyze all the paths simultaneously rather than one at a time.	Supported	Supported	Not supported
Maps specific and error variance of the observed variables into the research model.	Supported	Not supported	Not supported
Maps reflective observed variables	Supported	Supported	Supported
Maps formative observed variables	Not supported	Supported	Not supported
Permits rigorous analysis of all the variance components of each observed variable (common, specific, and error) as an integral part of assessing the structural model.	Supported	Not supported	Not supported
Allows setting of non-common variance of an observed variable to a given value in the research model.	Supported	Not supported	Supported by adjusting the correlation matrix.
Analyzes all the paths, both measurement and structural, in one analysis.	Supported	Supported	Not supported
Can perform a confirmatory factor analysis	Supported	Supported	Not supported
Provides a statistic to compare alternative confirmatory factor analyses models	Supported	Not supported	Not supported

Figure 1, shown in Section I and repeated below, illustrates the basic research model used throughout this tutorial. The causal linkages in TAM are thoroughly explained in the literature and need not be repeated here. Suffice it to say, TAM studies typically involve up to three hypotheses associated with these fundamental constructs (Table 4). First, PU is expected to influence outcome variables such as intention to use the system (see H_1). Researchers in this research stream choose outcomes depending on the questions they are investigating and the research methods they have selected. Attitudes toward use are also chosen as DVs (dependent variables) as are several standard IT use variables. The latter relationship is, perhaps, the most consistent finding in TAM studies with self-reported usage variables (see Straub, Limayem, and Karahanna [1995], however; this relationship raises a serious question about the possibility of common methods variance in most TAM studies). Moreover, it has come to represent the most interesting derivative work trying to explain the conditions and antecedents to PU and EOU.

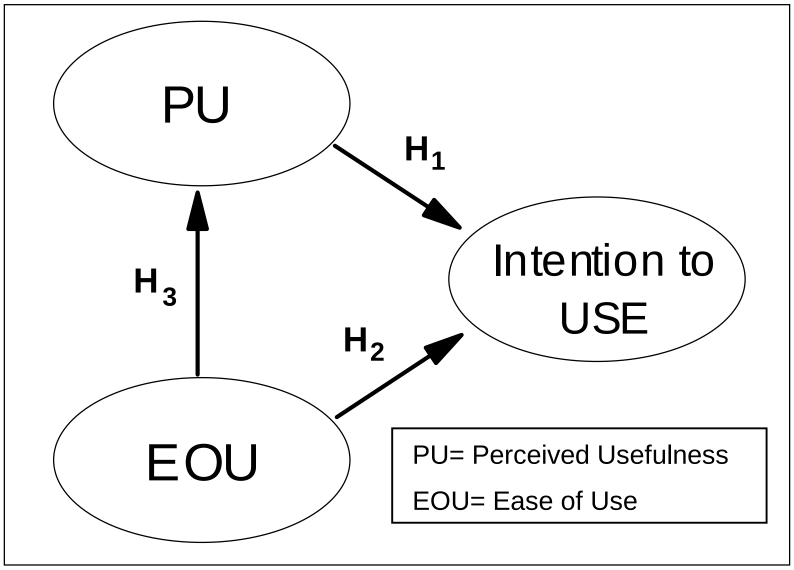


Figure 1. Basic TAM Model Used as Running Example

Table 4. Typical TAM Hypotheses

	Hypothesis
H_1	PU will impact the system outcome construct, Intention to Use the System.
H_2	EOU will impact the system outcome construct, Intention to Use the System.
H_3	EOU will impact PU.

.

In the original TAM studies by Davis [1989] and Davis et al. [1989], EOU was also thought to influence User Acceptance (a surrogate for IT Usage). With respect to H_2 in Table 4, these studies and subsequent studies did not find consistent results.[4] One empirically-derived explanation for why EOU did not produce invariant effects on system outcomes was offered by Davis [1989]. He argued that EOU may affect system outcomes only through the intermediate or intervening variable PU (i.e., H_3). His experiment confirmed this statistical explanation, which has also been posited and confirmed by later research (e.g., Adams et al. [1992], Gefen [2000], Gefen and Straub [2000], Keil et al. [1995], Venkatesh and Davis [1994]).

While a literature review and in-depth discussion of the TAM research model are not necessary here, elaboration of the measurement and data gathering are relevant. The instrument used to collect the data is shown in Appendix B. While the measures are based on previously validated instruments in the literature, the current study re-validates these measures, as recommended by Straub [1989].

METHODOLOGY

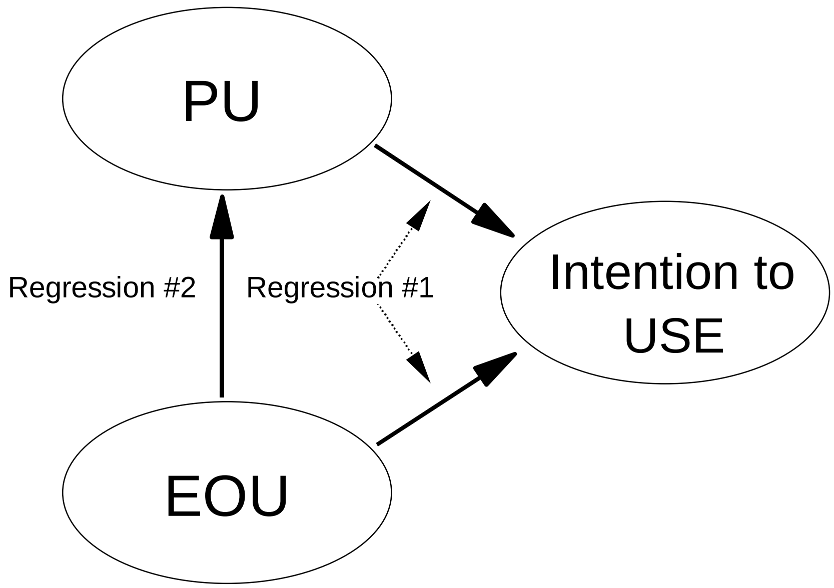
To test TAM via the three statistical techniques, we conducted a free simulation experiment [Fromkin and Streufert, 1976] with student subjects. As indicated in Appendix B, subjects were asked to use the Internet during the laboratory experiment to access Travelocity.com, thoroughly review the site, and then answer questions about it. In free simulation experiments, subjects are placed in a real-world situation and then asked to make decisions and choices as part of the experiment. Since there are no preprogrammed treatments, the experiment allows the values of the IVs (independent variables) to range over the natural range of the subject's experience. In effect, the experimental tasks induce subject responses, which are then measured via the research instrument.

Subjects were students taking MBA courses at the Lebow College of Business at Drexel University, a large accredited urban research university in Philadelphia. Most of the subjects were well acquainted with commercial Web sites where products and services are offered for sale, so the technology itself was not a novelty to them. Many were also familiar with the specific Web site selected for study, Travelocity.com. To permit controlling for possible effects from prior experience, we also measured the extent of this activity for each subject. One hundred and sixty subjects took part in the experiment. The exercise was optional for the course, which can be interpreted to mean that there should be no confounding effects from coercing subjects into participation. Participation in the experiment was voluntary and the students were not rewarded for taking part in it. Even so, 93% of the students volunteered to take part in the study.

DATA ANALYSIS USING LINEAR REGRESSION

Because linear regression cannot test all three relationships in a single statistical test, it is necessary to use two separate regressions to test the model fully. In regression #1, IUSE is the dependent variable and PU and EOU are independent variables. In regression #2, PU is regressed on EOU as the only independent variable. To perform linear regression analysis on the data, the researcher must first create an index for each of the constructs or variables. As shown in Appendix B, the index represents the value of the construct by averaging the subject responses to items PU1-PU6 for PU, items EOU1-EOU6 for EOU, and items IUSE1-IUSE3 for IUSE.

The findings from the statistical tests are shown in Figure 2. As is common in the literature [Gefen and Straub, 2000], H_1 and H_3 are significant and in the posited directions while H_2 is not. Using an index (average) for the constructs in the TAM testing is acceptable because the items making up the instruments scales were tested to ensure that they formed strong unities and demonstrate good measurement properties (construct validity and reliability). The tests most frequently used are factor and reliability analyses [Straub, 1989]. In this case, a Principal Components Analysis (PCA) of the primary research constructs showed extremely clean loadings in the factor structure, as depicted in Table 5. The only loading that was marginal was PU1, which was still above the commonly cited .40 minimum loading level [Hair et al., 1998]. The reliabilities reported are Cronbach's αs, and all are well above the cited minimums of .60 [Nunnally, 1967] or .70 [Nunnally, 1978, Nunnally and Bernstein, 1994]. Note that in the example all six PU items are included. Had PU1 been dropped, the factor analyses in PCA, LISREL, and PLS, would have shown a cleaner factor-loading pattern. (The same item also cross-loaded on the EOU factor in other e-commerce studies [Gefen and Straub, 2000].) The item was included because dropping it does not change the regression patterns and the objective is to use established scales "as is" in this demonstration.



	DV	F (R^2)	IV	Coefficient (T-value)
Regression #1	Intention to Use	23.80** (.24)	PU	.41 (4.45**)
			EOU	.10 (1.07)
Regression #2	PU	124.01** (.44)	EOU	.66 (11.14**)

** = Significant at the .01 level

Figure 2. TAM Causal Path Findings via Linear Regression Analysis

DATA ANALYSIS USING LISREL

To estimate coefficients, researchers employing LISREL typically use a different algorithm than the algorithm used for linear regression. Instead of minimizing variance as in regression, the most common LISREL estimation method maximizes likelihood.[5] The differences between the typical LISREL approach and that of regression will be examined in greater detail later in the paper. For the moment, it is sufficient to say that the preliminary factor and reliability analyses that are required to legitimate indices in linear regression

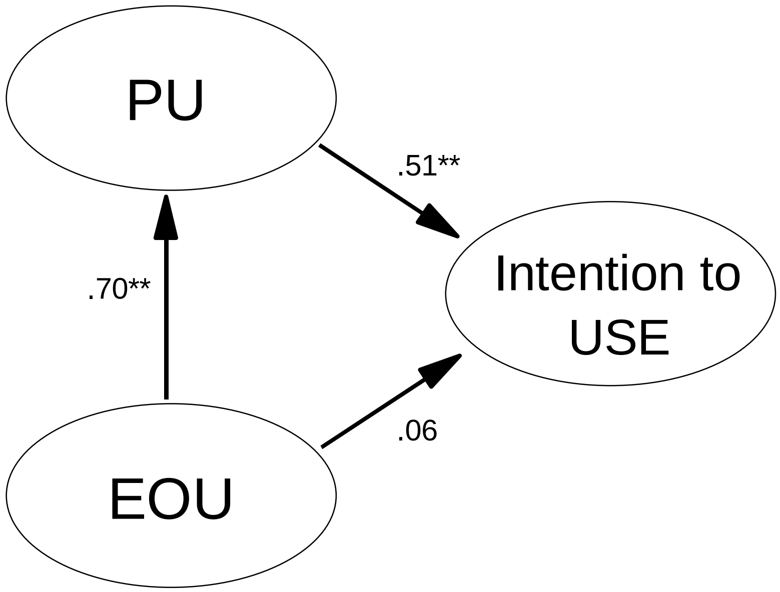
Table 5. Factor Analysis and Reliabilities for Example Dataset

Construct	Item	Factors 1	2	3	Cronbach's α
Perceived Usefulness (PU)	PU1	**.543**	.277	.185	.91
	PU2	**.771**	.178	.053	
	PU3	**.827**	.315	.185	
	PU4	**.800**	.268	.234	
	PU5	**.762**	.352	.236	
	PU6	**.844**	.437	.290	
Perceived Ease-of-Use (EOU)	EOU1	.265	**.751**	.109	.93
	EOU2	.217	**.774**	.150	
	EOU3	.270	**.853**	.103	
	EOU4	.303	**.787**	.105	
	EOU5	.248	**.831**	.179	
	EOU6	.242	**.859**	.152	
Intention To Use (IUSE)	IUSE1	.183	.147	**.849**	.80
	IUSE2	.224	.062	**.835**	
	IUSE3	.139	.226	**.754**	

Rotation Method: Varimax with Kaiser Normalization (Rotation converged in 6 iterations)

are not necessary in SEM techniques like LISREL and PLS because the testing of measurement properties of the instruments is simultaneous with the testing of hypotheses. The coefficients in LISREL can be read in a manner very similar to regression, that is, the standardized coefficients, known as betas and gammas, indicate the relative strength of the statistical relationships. And the loadings from the instrument items to the constructs (termed "latent" variables in SEM) can, once one recalibrates the scaling and examines the t-values, be interpreted in a similar manner to factor analysis.

We will discuss the LISREL findings in the same order in which the findings were discussed in the regression analysis. Unlike regression, however, it is only necessary to conduct a single LISREL run, in that the technique can consider the underlying structural relationships of all the latent variables at once. Moreover, it can also estimate the strength of the measurement items in loading on their posited latent variable or construct. Using the same dataset as in the regression runs (plus factor analysis and reliability tests), a single LISREL run produced the results shown in Figure 3 and Table 6. The SMC in Figure 3 is the LISREL equivalent of an R^2 in linear regression. It shows the percent of



LISREL Fit Indices
X^2 = 160.17
df = 87
AGFI = .84
RMR = .047

Link	**Coefficient (T-value)**	**SMC**
PU -> Intended Use	.51 (3.94**)	.30
EOU -> Intended Use	.06 (.48)	
EOU -> PU	.70 (7.05**)	.48

** = Significant at the .01 level

Figure 3. TAM Standardized Causal Path Findings via LISREL Analysis

explained variance in the latent variable [Bollen, 1989]

As in the regression analysis, H_1 and H_3 are significant and in the posited directions. H_2, likewise, is not significant. Moreover, LISREL provides several indications of the extent to which the sampled data fits the researcher-specified model. In this case, both the ratio of the χ^2 to the degrees of freedom (160.17/87=1.84) and the adjusted goodness of fit (AGFI) index (.84) tell the researcher that the model is a reasonably good-fitting model.[6] Finally, due to the low standardized root mean square residual (RMR), it is not unreasonable to conclude that the data fits the model. Dropping PU1 significantly improves the fit indexes (almost all the published LISREL analyses of TAM have dropped items).[7] So that readers can make straightforward comparisons, we will use the same tabular format as Table 5 to present the LISREL-generated factor loadings and reliabilities. Table 6 shows that the measurement properties for the instrument items using the confirmatory factor analysis (CFA) capability of LISREL are remarkably similar to those of the PCA performed earlier. All meet a standard for significance at the .01 level. The reliabilities are likewise respectable.

Table 6. Standardized Loadings and Reliabilities in LISREL Analysis

		Latent Construct Loading (and Error)			Reliability
Construct	**Item**	***PU***	***EOU***	***IUSE***	**Coefficient**
Perceived Usefulness (PU)	PU1	0.99 (.50)			.95
	PU2	1.10 (.39)**			
	PU3	0.93 (.45)**			
	PU4	1.07 (.26)**			
	PU5	1.10 (.29)**			
	PU6	1.11 (.24)**			
Perceived Ease-of-Use (EOU)	EOU1		0.78 (.45)		.94
	EOU2		0.95 (.38)**		
	EOU3		0.92 (.25)**		
	EOU4		0.99 (.31)**		
	EOU5		1.00 (.27)**		
	EOU6		0.94 (.21)**		
Intention To Use (IUSE)	IUSE1			1.36 (.34)	.95
	IUSE2			2.17 (.38)**	
	IUSE3			1.15 (.53)**	

The first item loading in each latent variable is fixed at 1.00 and does not have a t- value.
** Significant at the .01 level

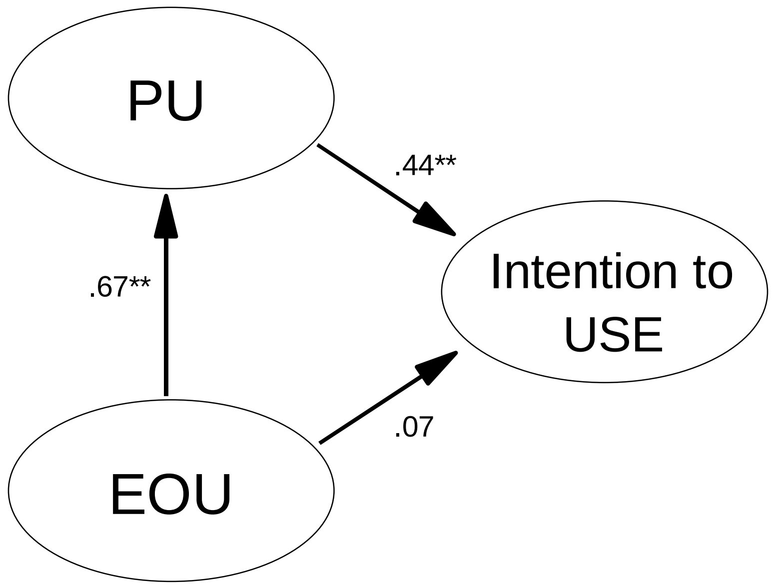
More details about each of these statistics are given below, but it is sufficient to point out at this time that the results of the LISREL analysis are in complete accord with those of the regression analysis. The primary differences that the reader may wish to take note of is that when all of the causal paths are tested in the same model, there is not a statistical issue with the lack of connection between runs, which characterizes all regression analyses. It is possible in regression, for example, to misinterpret the underlying causality in that no single run can partial out all the variance in complex research models.

DATA ANALYSIS USING PLS

In estimating its coefficients, PLS uses algorithms that have elements in common with both linear regression and LISREL. Like regression, it works with the variance of the individual data item from the means. In partialing out variance for the entire research model via iterative analysis, PLS resembles LISREL. In fact, it is this latter characteristic, that it works with the entire structure of the research model, that allows it to be categorized as a SEM technique.

Coefficients in PLS, shown in Figure 4, can be read in a manner very similar to regression and LISREL, that is, the standardized coefficients indicate the relative strength of the statistical relationships. Moreover, loadings from the instrument items to the constructs can also be interpreted in a similar manner to the PCA that precede regression runs[8] and the CFA that is utilized in LISREL. Using the same dataset as in the two previous analyses, a single PLS run produced the results shown in Figure 4 and Tables 7 and 8.

As before, H_1 and H_3 are significant while H_2 is not. While there are no overall model fit statistics produced by PLS, it can estimate t-values for the loadings utilizing either a jackknife or bootstrap technique. The loadings and the significance level of their t-values are shown in Table 7. Note that item loadings on their respective construct are presented by PLS, but that cross-loadings need to be calculated as the correlation of each standardized item with its factor scores on the constructs. Assessing the confirmatory factor analysis in PLS is then done by verifying that the AVE (discussed later) of each construct is larger than its correlations with the other constructs and that each item loading in the factor analysis is much higher on its assigned construct (factor) than on the other constructs. Table 8 shows the correlation and AVE table. The AVE is presented in the diagonal with a gray background.



Link	**Coefficient (T-value)**	R^2
PU -> Intended Use	.44 (3.69**)	.24
EOU -> Intended Use	.07 (.12)	
EOU -> PU	.67 (10.20**)	.44

** = Significant at the .01 level

Figure 4. TAM Causal Path Findings via PLS Analysis

Table 7. Loadings in PLS Analysis

		Latent Construct		
Construct	**Item**	***PU***	***EOU***	***IUSE***
Perceived Usefulness (PU)	PU1	**.776****	.613	.405
	PU2	**.828****	.498	.407
	PU3	**.789****	.448	.302
	PU4	**.886****	.558	.353
	PU5	**.862****	.591	.451
	PU6	**.879****	.562	.406
Perceived Ease-of-Use (EOU)	EOU1	.534	**.802****	.323
	EOU2	.557	**.839****	.338
	EOU3	.467	**.886****	.260
	EOU4	.562	**.843****	.289
	EOU5	.542	**.865****	.304
	EOU6	.508	**.889****	.288
Intention To Use (IUSE)	IUSE1	.350	.270	**.868****
	IUSE2	.380	.234	**.858****
	IUSE3	.336	.280	**.814****

N.B. A reliability statistic not automatically produced in PLS.
** Significant at the .01 level

Table 8. AVE and Correlation Among Constructs in PLS Analysis

AVE/ Correlation	IUSE	PU	EOU
IUSE	.721		
PU	.468	.742	
EOU	.359	.632	.738

SUMMARY AND CAVEAT

What do these three analyses of this sample dataset show? It is clear that in this particular circumstance, the analyses produced remarkably similar results. The reader should not generalize that this will always be the case, however. When certain endogenous constructs are added to this basic model, for example, the SEM analytical techniques — LISREL and PLS — come to different conclusions than linear regression. As developed by Straub [1994], Gefen and Straub [1997], and Karahanna and Straub [1999], the construct social presence-information richness (SPIR) has been found to predict PU. But in the dataset used for the running example, SPIR is statistically significant in two separate SEM analyses, but not in a regression analysis. Whether this difference is obtained because regression cannot partial out variance for the entire model whereas SEM can, or for some other reason, is not easy to determine. In spite of the fact that the measurement properties of the instrument seem to be acceptable, no instrument perfectly captures the phenomenon and the interaction between the measurement characteristics and the statistical technique may spell the difference. Then, again, as we shall shortly see, the assumptions and algorithms used in each of the techniques vary quite a bit and this could be the explanation.

The point is not to resolve this particular issue here. What is critical to note is that there may be subtle or even gross differences between analytical inferences about statistical conclusion validity depending on the researchers' choices — in sample, in instrument, in method, and in analytical technique.

III. SEM RESEARCH MODELS

Given the heavy increase in the use of SEM in well known IS journals, how does one know when the SEM statistics confirm or disconfirm hypotheses? Before addressing this key question, it is important to understand the central characteristics of the SEM techniques and what distinguishes them from ordinary least squares regression (linear regression models).

DIAGRAMMATIC SYNTAX

One of the most notable differences between SEM and its first generation predecessors, a difference that also indicates the nature of the analysis being performed, is the special diagrammatic syntax used in SEM. A sample of this syntax is presented in the theoretical model presented in Figure 5.

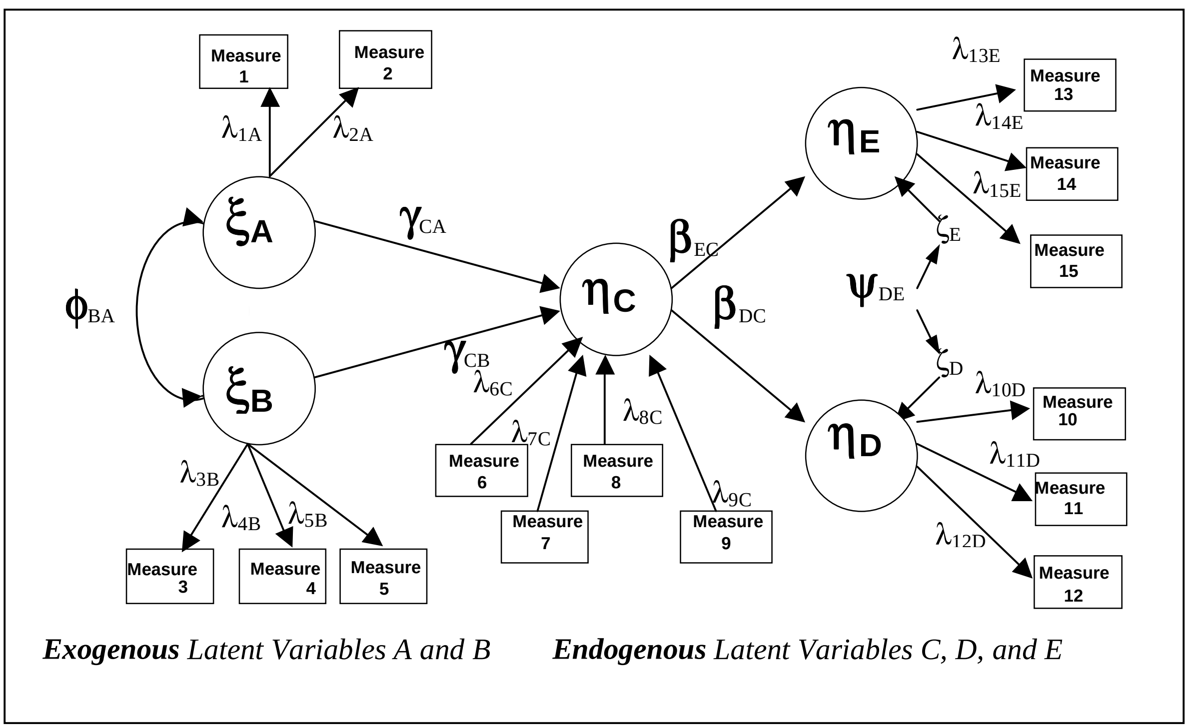


Figure 5. Generic Theoretical Network with Constructs and Measures

In LISREL terminology, the structural model contains the following:

- exogenous latent constructs called Xi or Ksi (ξ), depending on the dictionary used.
- endogenous latent constructs called Eta (η).
- paths connecting ξ to η represented statistically as Gamma (γ) coefficients.
- paths connecting one η to another are designated Beta (β).
- shared correlation matrix among ξ ; called Phi (ϕ).
- shared correlation matrix among the error terms of the η called Psi (ψ).
- the error terms themselves are known as ζ (Zeta).

To illustrate, IUSE and PU would be considered to be endogenous constructs in the TAM running example used earlier. Both are predicted by one or more other variables, or latent constructs. EOU, however, would be considered to be an exogenous latent construct in that no other variable in this particular model predicts it. The causal path PU (ξ_1) $\Rightarrow$ IUSE (ξ_2) was estimated as a β coefficient. The causal path EOU (η_1) $\Rightarrow$ PU (ξ_1) was estimated as a γ coefficient.[9]

In addition, the measurement model consists of:

- X and Y variables, which are observations or the actual data collected. X and Y are the measures of the exogenous and endogenous constructs, respectively. Each X should load onto one ξ, and each Y should load onto one η.
- Lambda X (λ_X) representing the path between an observed variable X and its ξ, i.e., the item loading on its latent variable.
- Theta Delta (Θ_δ) representing the error variance associated with this X item, i.e., the variance not reflecting its latent variable ξ.
- Lambda Y (λ_Y) representing the path between an observed variable Y and its η, i.e., the item loading on its latent variable.

- Theta Epsilon (Θ_ε) representing the error variance associated with this Y item, i.e., the variance not reflecting its latent variable η.

The Θ_δ and Θ_ε matrixes are diagonal by default, meaning that an error term is supposed to load only on its corresponding item. The λ_X and λ_Y matrixes are full and fixed, requiring the researcher to connect each item to its latent construct.

In the running example, the X observed variables were items EOU1-EOU6, since these measures are thought to reflect the latent construct EOU. For PU, the Y observed variables were PU1-PU6; for IUSE, the Y items were IUSE1-IUSE3.

Figure 5 shows the standard representation of these elements. Boxes represent X and Y items, observations, or empirical data that the researchers collected. These data are assumed to contain measurement error, not typically drawn in the diagram but always considered as part of the complete statistical model. With respect to the latent variables (constructs) of the model, these observations either reflect or form the latent constructs, and, thus, are said to be either reflective or formative. These latent variables – named A, B, C, D, and E in Figure 5 – are displayed as circles or ellipses.

Latent variables or research constructs cannot be measured directly. Note that the arrows connecting latent variables A, B, D and E to the measurement (also known as "indicator" or "observed") variables point *away from* the latent variables. The direction of the arrows indicates that LISREL assumes that the measurement variables reflect the construct represented by the latent variable. In PLS, however, arrows may also point *to* (rather than *from*) a latent variable if they are formative (see explanation below), as shown with latent construct C. As mentioned immediately above, the latent variables also have an error element that is typically not drawn in the diagram but is always part of the complete statistical model.

Arrows in the diagram between the latent constructs represent the researcher's hypothesized causation paths, estimating the extent to which the latent variables vary *linearly* with other latent variables in the model. Coefficients estimating the strength of the relationships are either βs or γs, depending on whether they represent early stage (effects of exogenous latent variables on endogenous latent variables) or late stage relationships (effects of endogenous latent variables on other endogenous latent variables) in the model. Latent variables may be correlated not only through hypothesized cause-effect relationships but also through correlated error variance. In this case, the correlation is shown with a double headed curved arrow, as between latent variables D and E, where the arrow connects the two error components, ζ, of the two constructs.

THE TWO PRIMARY METHODS OF SEM ANALYSIS

The holistic analysis that SEM is capable of performing is carried out via one of two distinct statistical techniques:

1. covariance analysis – employed in LISREL, EQS and AMOS – and
2. partial least squares – employed in PLS and PLS-Graph [Chin, 1998b, Thompson et al., 1995].

These two distinct types of SEM differ in the objectives of their analyses, the statistical assumptions they are based on, and the nature of the fit statistics they produce.

The statistical objective of PLS is, overall, the same as that of linear regression, i.e., to show high R^2 and significant t-values, thus rejecting the null hypothesis of no-effect [Thompson et al., 1995]. The objective of covariance-based SEM, on the other hand, is to show that the null hypotheses — the assumed research model with all its paths — is *insignificant,* meaning that the complete set of paths as specified in the model that is being analyzed is plausible, given the sample data. Moreover, its goodness of fit tests, such as χ^2

test the restrictions implied by a model. In other words, the objective of covariance-based SEM is to show that the operationalization of the theory being examined is corroborated and not disconfirmed by the data [Bollen, 1989, Hair et al., 1998, Jöreskog and Sörbom, 1989].

Another important difference between the two SEM techniques is that covariance-based SEM techniques, unlike PLS, enable an assessment of *unidimensionality*. Unidimensionality is the degree to which items load only on their respective constructs without having "parallel correlational pattern(s)" [Segars, 1997]. In factor analysis terms, unidimensionality means that the items reflecting a single factor have only that one shared underlying factor among them. Accordingly, there should be no significant correlational patterns among measures within a set of measures (presumed to be making up the same construct) except for the correlation associated with the construct itself (see also Anderson et al. [1987]). Unidimensionality cannot be assessed using factor analysis or Cronbach's α [Gerbing and Anderson, 1988, Segars, 1997].

An example of unidimensionality and parallel correlational patterns can clarify these terms. A student's GPA is the average of his or her course grades. Assume there are only 10 courses in a narrow subject area and all students take all 10 courses. All things being equal other than instructor, course grades in a factor analysis should all load onto one factor — the GPA for this set of courses. This can be verified using a factor analysis. It is possible, however, that some of the grades are related to each other *beyond* their loading onto the GPA factor. Such a circumstance could occur, for example, when two course sections are taught by a very lenient professor who tries to help his students by giving them higher grades than other professors in this same course. As a result, his two course sections would show a parallel correlational pattern. They would share variance with the overall course grades (the GPA factor), but would *also* have a significant shared variance between them. Likewise, if several of the courses were graded based on a take-home exam rather than on a traditional in-class examinations, it is unlikely that the 10 courses would show unidimensionality because the courses with the take-home exam would probably share a factor among themselves beyond the factor that is associated with all the grades of all the courses. In this hypothetical circumstance, it is likely that the take-home exam courses would share the "GPA" factor with the other courses, but would, in addition, have another shared factor among themselves reflecting the unique variance relating to take-home grades.

Unidimensionality testing can uncover such cases. When there is unidimensionality, there is no significant shared variance among the items beyond the construct which they reflect. In addition, while both methods of SEM provide for factor analysis**,** covariance-based SEM also provide the ability to compare alternative pre-specified measurement models and examine, through statistical significances, which is better supported by the data [Jöreskog and Sörbom, 1989]. Assuming that the models are nested, this type of CFA enables the comparison of two separate measurement models for the same data and a significance statistic for which model is superior [Segars, 1997].[10] Finally, covariance-based SEM provides a set of overall model-fit indices that include a wide set of types of fit (unlike the single F statistic in linear regression and the R^2 that is derived from this F-value). Covariance-based SEM is thought to provide better coefficient estimates and more accurate model analyses [Bollen, 1989].

OVERVIEW OF ANALYTICAL TECHNIQUES

Differences between SEM methods are the result of the varying algorithms for the analytical technique. Covariance-based SEM uses model fitting to compare the covariance structure fit of the researcher's model to a best possible fit covariance structure. Indices and residuals provided tell how closely the proposed model fits the data as opposed to a best-fitting covariance structure. Covariance-based SEM tests the *a priori* specified model against population estimates derived from the sample.[11,12] When the research model has a sound theoretical base, its overall objective is theory testing. Thus, these types of modeling examine whether the data is statistically congruous with an assumed multivariate distribution [Bollen, 1989, Hair et al., 1998, Jöreskog and Sörbom, 1989].[13] Covariance-based SEM techniques emphasize the overall fit of the entire observed covariance matrix with the hypothesized covariance model; for this reason, they are best suited for confirmatory research.

Our running example provides a straightforward translation of these terms. The TAM research model expresses certain causal paths that are specified in the theory or represent refinements or testable propositions by IS researchers. If this model is an accurate description of the system use/technology acceptance phenomenon, then the relationships between observed measures of these constructs in the theoretical model should be superior to a LISREL-generated model of no-fit. In other words, data gathered from the field or from experimental subjects should correspond well to patterns that are hypothesized by the research model. By comparing the sample data and its various path-, item loading-, and error variance-estimates to a null model, it is possible to see how good the researcher's TAM theoretical model really is.

PLS, the second major SEM technique, is designed to explain variance, i.e., to examine the significance of the relationships and their resulting R^2, as in linear regression. Consequently, PLS is more suited for predictive applications and theory building, in contrast to covariance-based SEM. Some researchers, thus, suggest that PLS should be regarded as a complimentary technique to covariance-based SEM techniques [Chin, 1998b, Thompson et al., 1995] — possibly even a forerunner to the more rigorous covariance-based SEM [Thompson et al., 1995]. Using OLS (Ordinary Least Squares) as its estimation technique, PLS performs an iterative set of factor analyses combined with path analyses until the difference in the average R^2 of the constructs becomes insignificant [Thompson et al., 1995]. Once the measurement and structural paths have been estimated in this way, PLS applies either a jackknife or a bootstrap approach to estimate the significance (t-values) of the paths.

Neither of these PLS significance estimation methods require parametric assumptions. PLS is thus especially suited for the analysis of small data samples and for data that does not necessarily exhibit the multivariate normal distribution required by covariance-based SEM [Chin, 1998b, Thompson et al., 1995]. This characteristic of PLS is in contrast to covariance-based SEM which requires a sample of at least 100 [Hair et al., 1998] or 150 [Bollen, 1989] because of the sensitivity of the χ^2 statistic to sample size [Bollen, 1989, Hair et al., 1998].[14] Nonetheless, even in PLS the sample size should be a large multiple of the number of constructs in the model since PLS is based on linear regression. One guideline for such a sample size in PLS is that the sample should have at least ten times more data-points than the number of items in the most complex construct in the model [Barclay et al., 1995].

Just as the objectives of the two types of SEM differ, so do their analysis algorithms. Covariance-based SEM applies second order derivatives, such as Maximum Likelihood (ML) functions, to maximize parameter estimates. Though LISREL uses ML estimates as a default, it can also be set to estimate these coefficients using other established estimation techniques, including Unweighted Least Squares (ULS), Generalized Least Squares (GLS), and Weighted Least Squares (WLS), among others. ULS can be used when the observed variables have the same units; GLS and ML are appropriate when the observed variables are known to be multivariate-normal, although they are applicable even when the observed variables deviate from this assumption [Jöreskog and Sörbom, 1989]. As to WLS, this estimation method should be used when polychoric correlations have been generated or when there are substantial deviations from a multivariate-normal distribution [Bollen, 1989, Jöreskog and Sörbom, 1983, Jöreskog and Sörbom, 1989].[15]

PLS, on the other hand, applies an iterative sequence of OLS and multiple linear regressions, analyzing one construct at a time [Thompson et al., 1995]. Rather than estimating the variance of all the observed variables, as in covariance-based SEM, PLS estimates the parameters in such a way that will minimize the residual variance of all the dependent variables in the model [Chin, 1998b]. Consequently, PLS is less affected by small sample sizes [Thompson et al., 1995], as in the case of linear regression models in general [Neter et al., 1990]. PLS, like linear regression models [Neter et al., 1990], is also less influenced by deviations from multivariate normal distribution [Chin, 1998b, Thompson et al., 1995], although sample size considerations influence the strength of the statistical test [Cohen, 1977, Cohen, 1988]. Comparisons based on all three aspects discussed were presented in Table 2 in Section I.

In the running example, it is clear that the data gathered from the free simulation experiment produces normalized/standardized path coefficients and R-squares that are similar across all three techniques. In minimizing the residual variance between the indicators of the latent variables PU and IUSE, EOU and IUSE, and EOU and PU, the statistical linkages in PLS between these constructs proves to be consistent with TAM theory. Moreover, despite the use of different estimation methods, the regression approaches reached comparable percent of explained variance (R^2 and SMC) and comparable standardized path coefficients.

THE SEM MODEL

The SEM model contains two inter-related models — the measurement model and the structural model. Both models are explicitly defined by the researcher. Pragmatically speaking, the researcher expresses which items load onto which latent variables and which latent constructs predict which other constructs through software packages specifically designed for these techniques, or, by one's expression of the equations via generalized packages like SAS. The measurement model defines the constructs (latent variables) that the model will use, and assigns observed variables to each. The structural model then defines the causal relationship among these latent variables (see Figure 5; the arrows between the latent variables represent these structural connections). The measurement model uses factor analysis to assess the degree that the observed variables load on their latent constructs (ξ and η, for exogenous and endogenous constructs, respectively). The manifest or observed variables are identified as Xs and Ys, for items reflecting the exogenous and endogenous constructs, respectively. SEM estimates item loading (λ) and measurement error for each observed item (Θ_δ and Θ_ε, respectively for X and Y items).

The item loadings provided by SEM are analogous to a factor analysis where each factor is, in effect, a latent variable. SEM techniques also explicitly assume that each of the observed variables has unique measurement error.[16] Measurement error represents both inaccuracy in participant responses and their measurement, as well as inaccuracies in the representation of the theoretical concept by the observed variables. Consequently, covariance-based techniques are well suited for the analysis of models containing variables with measurement error [Bullock et al., 1994, Hair et al., 1998, Jöreskog and Sörbom, 1989], facilitating a transition from exploratory to confirmatory analysis.[17]

Typically, a latent variable will be estimated based on multiple observed variables. Nonetheless, SEM does permit the use of constructs represented by single items. In such cases, in covariance-based SEM alone, the researcher explicitly sets parameters for the reliability and loading of the observed variable. Having a single item reflect a construct would be appropriate when the researcher uses an established scale with a known reliability and wishes to use an index of the scale as a whole, or when there is, indeed, only one item with little or no assumed measurement error, as with gender or age [Hair et al., 1998].

The structural model estimates the assumed causal and covariance linear relationships among the exogenous (ξ) and endogenous (η) latent constructs.[18] (As explained earlier, these paths are called γ when they link exogenous and endogenous latent constructs, and β when they link endogenous latent constructs.) SEM also estimates the shared measurement error for the constructs (ϕ and ψ, for exogenous and endogenous latent constructs respectively).[19] By allowing the researcher to specify these γ and ψ paths, SEM can support multi-layered causal models.

Covariance-based SEM and PLS differ, however, in the types of relationship they support between the observed variables and their associated latent constructs. PLS supports two types of relationship, *formative* and *reflective*.[20] Formative observed variables, as their name implies, "cause" the latent construct, i.e., represent different dimensions of it. Latent variables attached to formative measures are the summation of the formative observed variables associated with them [Campbell, 1960, Cohen et al., 1990, Thompson et al., 1995]. These observed variables are not assumed to be correlated with each other or to represent the same underlying dimension [Chin, 1998a].

The latent construct "Technological Environment," for example, might be measured by the extent of the IT infrastructure, but also by the level of technical support. These measures could be uncorrelated, but each viewed as "forming" the construct.

Reflective observed variables, on the other hand, reflect the latent variable and as a representation of the construct should be unidimensional and correlated [Gerbing and Anderson, 1988]. To emphasize this difference, formative items are drawn with an arrow leading to the latent construct, while reflective items are drawn with an arrow leading away from the latent construct. PLS supports both types of observed variables whereas covariance-based SEM has been interpreted to support only reflective observed variables [Chin, 1998b, Thompson et al., 1995].[21] According to one interpretation, reflective observed variables should be preferred to formative ones when there is a relevant theory and when the objective is theory testing rather than theory building [Chin, 1998b].

An example might better clarify the difference between reflective and formative observed variables. When a construct, such as intelligence, cannot be measured directly, researchers measure it indirectly using several indicator variables. In the case of intelligence these indicator variables might be scores obtained from a test. When the scores are assumed to measure the same underlying aspect of intelligence, they are reflective. This situation would occur, for example, when a researcher is measuring algebraic intelligence and the indicator variables chosen evaluate aptitudes for addition, division, subtraction, and multiplication. On the other hand, when more than one aspect of intelligence is being measured, such as when the exam tests both algebraic and linguistic intelligence using one indicator variable each, then the indicator variables would be formative of a construct for “intelligence.” It is conceivable and often the case that an individual’s algebraic and linguistic intelligence can be reasonably thought of as composite elements (or sub-constructs/meso-level constructs) of the molar-level construct "intelligence," but not necessarily highly correlated with each other. Therefore, they are formative rather than reflective of the molar construct "intelligence." Whereas both algebraic intelligence and linguistic intelligence are viable sub-constructs in this situation, the nature of constructs chosen by the researcher in other situations will determine whether the measures are better seen as formative or reflective.

The ability to analyze complex models (like that shown in Figure 5) in a single, unified process is a major advantage of both types of SEM over first generation regression models. In first generation regression models, item loadings on the latent variables must be analyzed in a separate step (as shown in the TAM running example in Section II) and the linkage to each dependent variable must be assessed independently (other than MANOVA, of course).[22] SEM analysis also generally results in a more rigorous variance analysis [Bollen, 1989], and enables the researcher to include not only common variance but also specific and error variance explicitly into the research model [Hair et al., 1998].[23]

Some SEM, such as LISREL, also permit the researcher to specify how the specific and error variance of each observed variable relates to those of other observed variables. Accordingly, LISREL allows the setting and fixing of the item loading and measurement error of the observed variables [Bollen, 1989]. Setting the items loading, however, should not be exercised unless there is a good reason for doing so, such as comparing samples or when it is known that there is little or no measurement error (e.g., when measuring gender or age).[24] Table 3 in Section II presented guidelines based on capabilities by research approach.

APPLYING CRITERIA TO THE RUNNING EXAMPLE

How would these criteria for analytical method choice apply in the case of the TAM running example? In the first case, as indicated earlier, TAM is a mature theoretical research stream in IS research. As such, the relationships between the basic constructs are relatively well understood. Based on Table 2, therefore, TAM testing should use confirmatory analytical techniques, which, in this case, means that any of the three methods would be appropriate although LISREL and regression are to be preferred as they are especially suited for testing theory. Given that the sample size exceeds the minimal requirements for LISREL, which is the most demanding in this regard, any of these techniques would also be appropriate with regard to this criterion.

There are, however, conditions where the use of linear regression and PLS would be the most appropriate choices for the TAM running example. If the sample size for the TAM researchers had been low, then the power of a LISREL analysis would have suffered badly and PLS, which can work with much smaller samples, would have been a better choice. The tradeoff in this situation would be that PLS is best used for exploratory research, but can, when necessary, serve for confirmatory work.

Regression might have been an appropriate choice if the researcher wished to make specific and direct comparisons to other studies that used this technique in the research tradition. By the same token, ANOVA or MANOVA might be employed for these same reasons. The statistics generated by regression and older statistical techniques seem to be more amenable to meta-analysis, which might also be a factor in its selection. Researchers who want to add to the research tradition and meta-analyze the cumulative effect of TAM studies would find it simpler to work with regression, ANOVA, t-tests, and simple or partial correlations.

Finally, if the LISREL TAM model had refused to converge, as it did in some of the runs with our sample data when the SPIR variables were included, PLS or regression may also be a better choice. One should never conclude that the refusal of LISREL to converge represents anything other than the inability of the matrices to be reduced, which is the mathematical method used for maximum likelihood estimation. Lack of convergence does not suggest anything definitive about the model itself (as is obvious in the TAM case presented here) or its hypothesized causal paths. If LISREL reports that the reason for non-convergence is that a matrix is not positive definite, then two rows (item measures) are likely so similar that matrix reduction cannot be carried out, but this would imply more about measurement than about the underlying theory being tested and relationships between constructs. Moving to another technique is a perfectly acceptable alternative in such a case.

STATISTICS IN SEM

Just as the two types of SEM techniques differ in their underlying statistical assumptions and estimation methods, so do the statistics they produce. First, it is important to note in this respect that covariance-based SEM, unlike linear regression models and PLS, does not always converge and produce interpretable results. A covariance-based SEM model that does not converge will have to be modified or the theory base reassessed when the model:

- does not converge,
- warns of a non-positive definite covariance matrix, or
- adds a ridge to the covariance matrix,

Lack of convergence notwithstanding, the next few paragraphs describe SEM statistics, starting with covariance-based SEM statistics.

Covariance-based SEM packages generate statistics at three levels:

1. at the individual path and construct level.
2. at the overall model fit level.
3. individual path modification indexes.

At the individual path level, SEM estimates item loadings and measurement error along with their respective t-values. Construct reliability, the analog of a Cronbach's α, can then be derived from these statistics.[25] As with Cronbach's α statistics, construct reliability should be above .70 [Hair et al., 1998, Segars, 1997]. SEM also estimates the coefficients and t-values representing the relationships among the latent constructs γs, βs, ϕs, and ψs. As in linear regression, a t-value is associated with each of these. The t-values of the γs and βs need to be significant to support the hypothesized paths (above

1.96 or 2.56, for alpha protection levels of .05 and .01, respectively).

The next important statistic in this group is the Squared Multiple Correlation (SMC) of each of the exogenous latent constructs. Equivalent to an R^2 in linear regression, the SMC is the explained variance of each latent construct [Bollen, 1989].

The second set of statistics deals with the entire model fit. The most important of these statistics is the likelihood-ratio chi-square (χ^2). Technically speaking, the χ^2 statistic should be *insignificant* with a p-value above .05, because an insignificant χ^2 shows good model fit [Jöreskog and Sörbom, 1989].[26] However, this criterion is satisfied only rarely because χ^2 is sensitive to larger sample sizes and the power of the test [Jöreskog and Sörbom, 1989]. Therefore the ratio of χ^2 to degrees of freedom is sometimes examined.[27] Some commentators recommend that the ratio of χ^2 to degrees of freedom be between 1 and 2 [Hair et al., 1995, Hair et al., 1998]. But the IS literature has been more forgiving in this regard, recommending just a χ^2 as small as possible [Segars and Grover, 1993] and showing a ratio of χ^2 to degrees of freedom smaller than 3:1 [Chin and Todd, 1995].

Finally, the most widely used overall model fit indices are the Goodness of Fit Index (GFI), the Adjusted Goodness of Fit Index (AGFI), and the Root Mean Residual (RMR). GFI measures the absolute fit (unadjusted for degrees of freedom) of the combined measurement and structural model to the data. AGFI adjusts this value to the degrees of freedom in the model. The standardized RMR (Root Mean Residuals), on the other hand, assesses the residual variance of the observed variables and how the residual variance of one variable correlates with the residual variance of the other items. It is important to note that large standardized RMR values mean high residual variance, and that such values reflect a poorly fitting model. Thresholds for these indices in IS research are above .90, above .80, and below .05, respectively [Chin and Todd, 1995, Segars and Grover, 1993]. A more restrictive .90 threshold for AGFI is sometimes cited (e.g., Chin and Todd [1995], Hair et al. [1998]).

Another important fit index is the Normed Fix Index (NFI), which measures the normed difference in χ^2 between a zero factor null model with no common variance across measures and a proposed multi-factor model [Bentler, 1990].[28] Typically, NFI should be above .90 [Chin and Todd, 1995, Hair et al., 1998].

The third set of statistics is the modification indexes. Some SEM, notably LISREL, provide modification indices that estimate the difference in model fit χ^2 for each possible individual additional path. A value in these so-called modification matrices [Jöreskog and Sörbom, 1989] above 3.84 suggests that adding that path may significantly improve model fit [Hair et al., 1998]. This criterion is analogous to the way stepwise linear regression chooses to add IVs to the regression model, except that stepwise linear regression analyzes the change in the F statistic. Researchers should be cautious, however, to add only paths justified by theory and not attempt to retrofit the model [Bullock et al., 1994, Hair et al., 1998].

Please note that the LISREL statistics in the TAM running example exceed all of the thresholds just cited. The fit indices are good, and the residual variance is low. The ratio of χ^2 to degrees of freedom is well within boundaries. The T-values indicate that the paths that are posited to be significant are significant and those that were not expected to be significant, are, indeed, not significant. A minimalist interpretation is that statistical conclusion validity is in favor of the TAM research model and that the data does not disconfirm the theory. In spite of this conclusion, measurement issues in TAM remain. Common methods variance could be a serious problem for nearly all TAM studies to date [Straub et al., 1995].

PLS has a less extensive set of statistics. At the measurement model level, PLS estimates item loadings and residual covariance. At the structural level, PLS estimates path coefficients and correlations among the latent variables, together with the individual R^2 and AVE (Average Variance Extracted)[29] of each of the latent constructs. T-values of both paths and loadings are then calculated using either a jackknife or a bootstrap method. Good model fit is established with significant path coefficients, acceptably high R^2 and internal consistency (construct reliability) being above .70 for each construct [Thompson et al., 1995]. Convergent and discriminant validity are assessed by checking that the AVE of each construct is larger than its correlation with the other constructs, and that each item has a higher loading (calculated as the correlation between the factor scores and the standardized measures) on its assigned construct than on the other constructs. The implications of these issues are presented in Table 9.

Table 9. Comparative Analysis Based on Statistics Provided by SEM

Statistics	LISREL	PLS	Regression
Analysis of overall model fit	Provided	Provided	Provided
Analysis of individual causation paths	Provided	Provided	Provided
Analysis of individual item loading paths	Provided	Provided	Not provided
Analysis of residual non-common error	Provided	Not Provided	Not provided
Type of variance examined	1. Common 2. Specific 3. Error	Common Combined specific and error	Common
Analysis of statistical power	Not available	Available through the f^2 statistic.	Available

Again, the PLS run in the TAM running example generates statistics that infer that the instrument has acceptable measurement properties and that the hypothesized relationships are supported by the data. T-values were all significant for every item loading onto the latent constructs and for every path except for the EOU ⇒ IUSE link (as predicted). Explained variance is in keeping with other studies in the tradition.

ADDITIONAL ANALYSES: NESTED MODELS AND INTERACTION EFFECTS

Good fit indices show that the data support the proposed model, but they do not indicate that the selected model is necessarily parsimonious or the best model among a set of theoretically feasible models. These issues can be examined in covariance-based SEM techniques in a manner analogous to the way nested linear regressions can examine the significance of the difference in the F and in the R^2 statistics between nested models via a stepwise liner regression. The application of nested models in SEM is discussed in Appendix C. The implications are presented in Table 10.

Table 10. Comparative Analysis Based on Capabilities

Capabilities	**LISREL**	**PLS**	**Regression**
Examines interaction effect on cause-effect paths	Supported	Supported	Supported
Examines interaction effect on item loadings	Supported	Not readily supported	Not supported
Examines interaction effect on non-common variance	Supported	Not readily supported	Not supported
Examines interaction effect on the entire model	Supported	Not readily supported	Not supported
Can cope with relatively small sample size	Problematic	Supported	Supported
Readily examines interaction effect with numerous variable levels	Problematic	Supported	Supported
Can constrain a path to a given value	Supported	Not supported	Not supported
Examines nested models	Supported	Supported	Supported

Another examination that is sometimes necessary is the analysis of interaction effects. In linear regression and analysis of variance models examining this is relatively simple. One adds a new variable to the regression model, calculated as the product of the assessed independent variables that are assumed to interact, and then rerun the regression [Neter et al., 1990]. However, this procedure does not work well in covariance-based SEM because, inevitably, such a calculated new variable will have high shared residual variance with the variables from which it is derived.[30] As with any other high residual variance, this deviation will then be reflected in the RMR statistic. Consequently, interaction effects are assessed in a different manner in covariance-based SEM. The recommended approach is to use multi-sample analysis [Jöreskog and Sörbom, 1989].

Multi-sample analysis is performed in covariance-based SEM by examining the parameter estimates of exactly the same model run with two distinct samples, and constraining the ϕ and/or the ψ elements of the second sample to be equal to those derived for the first sample. Alternatively, the two-sample analysis can constrain any or several of the paths γ, β, λ_X, Θ_δ, λ_Y, Θ_ε in the second model to equal those in the first model. Thus, LISREL can examine an interaction effect of the kind examined in linear regression by constraining the γ or the β paths in one sample to be equal to those estimated by LISREL in the other sample. If the χ^2 of the model with the constrained paths is significantly smaller than the χ^2 of the model with the unconstrained paths, given the difference in degrees of freedom between the two χ^2, then there is a significant interaction effect [Jöreskog and Sörbom, 1989].

For example, examining a gender effect on a given model would require running the theoretical model on the sub-sample of one gender first, and then running exactly the same model with the sub-sample of the other gender but constraining the paths to the path estimates obtained from the first gender. Constraining the other paths in this manner would permit the exploration of other types of interaction effects, some of which cannot be examined in linear regression, such as whether item loadings differ across sub-populations.

Examining interactions in this manner, however, requires a separate sample for each interaction value. For example, an interaction effect based on gender would require two samples and one analysis to compare the two genders, but an interaction effect based on a four-value category interaction would require 4 samples and 6 comparative analyses [Jöreskog and Sörbom, 1989]. Consequently, this type of analysis is not very practical once the number of interaction categories is large because of the need to collect separate samples for each category and the probability of getting a significant t-value in one of the tests purely by chance.[31] The implications of these issues are presented in Table 10.

IV. WHEN TO USE LINEAR REGRESSION IN PREFERENCE TO SEM

INTERPRETING CAUSAL RELATIONSHIPS IN SEM

Establishing causation is difficult in research. Typically, establishing causation requires showing [Cook and Campbell, 1979]:

1. association,
2. temporal precedence, and
3. isolation.

Association means that when the "cause" event happens, it is very likely that the "effect" event will happen too. For example, when fires break out firefighters are usually there. Thus, "fires" are associated with "firefighters". Association is typically measured through correlation. Correlation alone, however, is not enough to establish causation; it is also necessary to establish that the "cause" event occurred before the "effect" event. Thus, one may conclude that the fires cause the arrival of the firefighters, and not vice versa, because the fires occur first. One would be mistaken, however, to conclude that fires cause firefighters to come, because there are other events involved, specifically, somebody calling the fire-department. Without showing that no other event was involved, concluding that such causation occurred would be misleading. Establishing that no such other event occurred is called isolation[32] or ruling out rival hypotheses [Cook and Campbell, 1979].

Consequently, statistical analysis alone cannot prove causation, because it does not establish isolation or temporal ordering [Bollen, 1989, Bullock et al., 1994]. Nonetheless, correlation analysis, including linear regression and SEM, can be used to show that the correlations found in the data are in accordance with the causation predicted by an established theory-base [Bollen, 1989]. These principles apply equally well to SEM, except that corroborating causation in this manner is more difficult in SEM because of the complexity of the structural models it supports and the large number of alternative, but statistically equivalent, models that can be supported by the same data. These effects have been extensively studied with regard to covariance-based SEM, where it has been shown, for example, that reversing the direction of any causation path or replacing it with a correlation path will produce an equivalent model with the same fit indices [Stelzl, 1986]. This concern for equivalence of models and the concern for "over-fitting" the model to the data and consequently coming up with non-generalizable results is a major reason why covariance-based SEM should be used as a confirmatory and not as an exploratory method [Bullock et al., 1994, Hair et al., 1998].

Another concern in inferring a cause-effect related issue in SEM is specification errors, i.e., not specifying an important construct in the model and/or not specifying enough observed measurements for each construct [Bagozzi and Baumgartner, 1994].[33] Bias created by either of these problems can result in an incorrect interpretation of the results, as in other types of statistical analysis [Hair et al., 1998].

Because of over-fitting, the fact that the same data can support many equivalent models, and specification errors, the assumed causation in covariance-based SEM should be based on a theoretical rationale supported by data. In other words, the assertion of causation is applicable in SEM only when and because the data analysis corroborates theory-based causation hypotheses (as specified in the structural model) [Bollen, 1989, Bullock et al., 1994, Hair et al., 1998]. Consequently, covariance-base SEM should be used as a confirmatory analysis method only. It needs to show that the hypotheses are plausible given the data. PLS, on the other hand, does not require strong theory and can be used as a theory-building method [Chin, 1998b, Thompson et al., 1995]. The implications of these issues are presented in Table 11.

INHERENT ANALYTICAL ASSUMPTIONS

Another major concern when using SEM is inherent assumptions, such as data distribution assumptions. Apart from the assumed multi-normal distribution

Table 11. Comparative Analysis Based on Capabilities

Capabilities	LISREL	PLS	Regression
Establishment of causation	No	No	No
Possible over-fitting	Problematic	Less problematic	Less problematic
Testing of suspected non-linear effect	Problematic	Problematic	Mitigated by data transformation
Suspected influential outliers	Problematic	Problematic	Mitigated by data transformation
Suspected heteroscedasticity	Problematic	Problematic	Mitigated by data transformation
Suspected polynomial relation	Problematic	Problematic	Mitigated by data transformation

that is important when ML estimation is used (discussed above), a central assumption in SEM is that the relationship between the observed variables and their constructs and between one construct and another is *linear*. SEM has no established tools for handling variations from this assumption, unlike linear regression that has established methods of identifying and proven remedial data transformational methods for handling data that has nonlinear relationships. Linear regression can also deal with multicollinearity (violations of the assumed independence of predictor variables), outliers, heteroscedasticity (unequal variance among the measurement items), and polynomial relationships (such as: $Y = b_0 + b_1X + b_2X^2$) [Hair et al., 1998, Neter et al., 1990]. No such remedies are available yet in SEM. SEM has no tools to identify, let alone handle, these violations of the major distribution assumptions. Using linear regression is advisable in these cases, as shown in Table 11.

V. WIDELY USED VALIDATION HEURISTICS IN SEM

Validity rules of thumb are pragmatic measures indicating patterns of behavior that are acceptable within a scientific community. There is no recognized means of verifying the truth of such heuristics, other than through tradition or evaluation of best of breed practice. It is traditional, for example, to accept a p-value of .05 in SEM [Jöreskog and Sörbom, 1989], just as the .01 and .05 thresholds are the accepted heuristics in linear regression [Neter et al., 1990]. As with first generation regression models, there is no mathematical or other means for establishing these levels [Nunnally, 1967, Nunnally, 1978, Nunnally and Bernstein, 1994]. Nonetheless, rules of thumb are desirable because of their practicality, enabling researchers to utilize them as *de facto* standards. A summary of key heuristics is presented in Table 12.

Table 12. Heuristics for Statistical Conclusion Validity (Part 1)

Validity	**Technique**	**Heuristic**
Construct Validity		
Convergent Validity	CFA used in covariance-based SEM only.	GFI > .90, NFI > .90, AGFI > .80 (or >.90) and an insignificant χ^2, to show unidimensionality. In addition, item loadings should be above .707, to show that over half the variance is captured by the latent construct [Chin, 1998b, Hair et al., 1998, Segars, 1997, Thompson et al., 1995].
Discriminant Validity	CFA used in covariance-based SEM only.	Comparing the χ^2 of the original model with an alternative model where the constructs in question are united as one construct. If the χ^2 is significantly smaller in the original model, discriminant validity has been shown [Segars, 1997].
Convergent & Discriminant Validities	PCA used in PLS can assess factor analysis but not as rigorously as a CFA in LISREL does and without examining unidimensionality	Each construct AVE should be larger than its correlation with other constructs, and each item should load more highly on its assigned construct than on the other constructs.
Reliability		
Internal Consistency	Cronbach's α	Cronbach's αs should be above .60 for exploratory research and above .70 for confirmatory research [Nunnally, 1967, Nunnally, 1978, Nunnally and Bernstein, 1994, Peter, 1979].
	SEM	The internal consistency coefficient should be above .70 [Hair et al., 1998, Thompson et al., 1995].
Unidimensional Reliability	Covariance-based SEM only.	Model comparisons favor unidimensionality with a significantly smaller χ^2 in the proposed measurement model in comparison with alternative measurement models [Segars, 1997].

Table 12. Heuristics for Statistical Conclusion Validity (Part 2)

Model Validity		
AGFI	LISREL	AGFI > .80 [Segars and Grover, 1993]
Squared Multiple Correlations	LISREL, PLS	No official guidelines exist, but, clearly, the larger these values, the better
χ^2	LISREL	Insignificant and χ^2 to degrees of freedom ratio of less than 3:1 [Chin and Todd, 1995, Hair et al., 1998]
Residuals	LISREL	RMR <.05 [Hair et al., 1998]
NFI	LISREL	NFI > .90 [Hair et al., 1998]
Path Validity Coefficients	LISREL	The β and γ coefficients must be significant; standardized values should be reported for comparison purposes [Bollen, 1989, Hair et al., 1998, Jöreskog and Sörbom, 1989] .
	PLS	Significant t-values [Thompson et al., 1995].
	Linear Regression	Significant t-values [Thompson et al., 1995].
Nested Models		
	LISREL	A nested model is rejected based on insignificant βs and γs paths and an insignificant change in the χ^2 between the models given the change in degrees of freedom [Anderson and Gerbing, 1988] [Jöreskog and Sörbom, 1989] .
	PLS	A nested model is rejected if it does not yield significant a f^2 [Chin and Todd, 1995].
	Linear Regression	A nested model in a stepwise regression is rejected if it does not yield a significant change in the F statistic (reflected directly in the change in R^2) [Neter et al., 1990].

Given that these guidelines are what amount to *de facto* SEM standards for the IS field, we collected data (in the same research discussed in Section 1) on the extent to which IT research follows these guidelines. As can be seen from Table 13 and Table 14, there are areas of concern and areas where the field is doing remarkably well.

What should be said about the reporting of SEM covariance-based statistics in the IS literature? The grayed rows in Table 13 are, in our view, both a critical and minimal set of statistics for establishing construct validity and the truth of theoretical models, and so we will concentrate on these rows. The lack of reporting of AGFI across all three journals is, frankly, disturbing. As argued above, the adjusted goodness of fit reports whether the theory fits the data or not, given a statistical adjustment for degrees of freedom. Readers are left in serious doubt as to the merit of the case when this statistic is absent. As Table 13 notes, when this statistic is being reported, the values on the whole seem to meet our rule of thumb, which is a hopeful sign.

Table 13. Number Of Covariance-based SEM Articles Reporting SEM Statistics in IS Research

Statistics	*I&M* (n=6)	*ISR* (n=7)	*MISQ* (n=5)	All Journals (n=18)
GFI reported	3 (50%)	3 (43%)	1 (20%)	7 (39%)
Of GFI reported, number > 0.90	1 (33%)	2 (67%)	1 (100%)	4 (57%)
AGFI reported	2 (33%)	2 (29%)	1 (20%)	5 (28%)
Of AGFI reported, number > 0.80	1 (50%)	2 (100%)	1 (100%)	4 (80%)
RMR reported	2 (33%)	4 (57%)	2 (40%)	8 (44%)
Of RMR reported, number < 0.05	0 (0%)	1 (25%)	1 (50%)	2 (25%)
χ^2 insignificance reported	3 (50%)	2 (29%)	0 (0%)	5 (28%)
Of χ^2 insig. reported, number > .05	3 (100%)	1 (50%)	0 (0%)	4 (80%)
Ratio χ^2 / df reported	5 (83%)	6 (86%)	4 (80%)	15 (83%)
Of ratio χ^2 / df reported, number < 3	5 (100%)	5 (83%)	2 (50%)	12 (80%)
SMC	2 (33%)	3 (43%)	2 (40%)	7 (39%)
NFI reported	3 (50%)	3 (43%)	3 (60%)	9 (50%)
Of NFI reported, number > .90	2 (67%)	3 (100%)	3 (100%)	8 (89%)
CFI reported	3 (50%)	2 (29%)	1 (20%)	6 (33%)
T-values or significance of paths	4 (67%)	6 (86%)	4 (80%)	14 (78%)
Construct Reliability reported	5 (83%)	7 (100%)	4 (80%)	16 (89%)
Use of Nested Models	4 (67%)	6 (86%)	3 (60%)	13 (72%)

Notes: Rows in gray should receive special attention when reporting results
11 articles used LISREL, 6 EQS, and 1 AMOS

Table 14. Number of PLS Studies Reporting PLS Statistics in IS Research (Rows in gray should receive special attention when reporting results)

PLS Statistics	*I&M* (n=2)	*ISR* (n=5)	*MISQ* (n=4)	All Journals (n=11)
R^2 reported	2 (100%)	5 (100%)	4 (100%)	11 (100%)
AVE reported	2 (100%)	5 (100%)	3 (75%)	10 (91%)
T-values or significance of paths	2 (100%)	5 (100%)	4 (100%)	11 (100%)
Construct Reliability reported	2 (100%)	4 (80%)	3 (75%)	9 (82%)
Use of Nested Models	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Expressing the extent to which the model explained the variance in the dataset for each exogenous variable, the SMCs are likewise being reported at low levels, across all journals. Again, it is difficult to see how a researcher can hope to defend the explanatory power of his/her model without this statistic. Since there are no rules of thumb for explained variance, it only remains for researchers to convince reviewers/editors that the values reported are sufficiently high to indicate that the theory has reasonable explanatory power. It is purely a matter of good argumentation and not something that authors should, therefore, avoid.

Whereas reporting of RMRs is roughly as deficient as reporting of the AGFIs and SMCs, and also an area that calls for greater attention, the disclosure of χ^2 / df ratio, t-values, and construct reliability is generally good. It is curious that editors and reviewers are apparently stringent with regard to these statistics, but not so with AGFI, SMC and RMR. Another encouraging signal is that when these statistics are reported, they generally meet or exceed the rules of thumb articulated in Table 12.

Other than nested models, all of the PLS statistics shown in Table 14 should be reported, and usually are. Perhaps because there are fewer overall statistics offered to the researcher in PLS, these have most often been placed in the public forum for readers.

A final note about sample size may also be useful at this juncture. In spite of the fact that PLS can be run with relatively small sample sizes, these, on average, were larger than those in the LISREL articles. The mean for PLS articles was 295 (minimum 40, maximum 1020) whereas for LISREL, it was 249 (minimum 41, maximum 451). The low minimum among the LISREL articles raises a flag, in that the rules of thumb recommend at least 100.

VI. CONCLUSION

Covariance-based SEM, PLS-based SEM, and linear regression models overlap in many ways, including analysis objectives, distribution assumptions, and etiological and correlational linearity assumptions. Nonetheless, there are distinct differences among the three approaches that makes each more or less appropriate for certain types of analysis. Furthermore, even when all three techniques are appropriate, the resulting set of supported hypotheses in the model may be more or less credible because of underlying data distribution assumptions and the analysis methods employed.

Thus, choosing an analysis method based correctly on the research objectives and the limitations imposed by the sample size and distribution assumptions is crucial. The importance of establishing statistical conclusion validity using such tools in positivist research cannot be overemphasized. It is, in essence, the strength of evidence researchers have to report in order to prove that their models are supported by data collected. Indeed, studies lacking strong statistical conclusion validity are highly questionable [Cook and Campbell, 1979]. This paper has presented key criteria for effective practices in the use of new and old tools for this form of validation. These guidelines are summarized in the tables throughout the tutorial.

The meta-analysis shown in Tables 13 and 14 indicates that much still must be done in this regard. There is wide disparity among journals on utilization of SEMs. In *ISR*, for instance, 45% of empirical articles use SEM techniques, whereas in *MISQ*, this figure is closer to 25%. Assuming that SEM techniques represent state-of-the-art in many research settings, this discrepancy must be heeded. Editors and reviewers may want to encourage authors to use SEM tools, where appropriate. Nonetheless, as noted in this article, there are situations where SEM tools are not called for. In such cases, editors and reviewers will want to ensure that authors are not over-using the techniques, by, perhaps, choosing them for mimetic rather than for solid, technical reasons.

To internalize such statistical knowledge, editors, associate editors, and reviewers will want to immerse themselves in at least the three (or four, including factor analysis) techniques touched on in this article. There are many instances where an editor will be confronted with disagreements among the methodological experts asked to review and where merely adding another knowledgeable reviewer is not going to resolve the issue. The reviewing process should not be a vote. It should be a set of judgments, where more knowledgeable opinions are weighted more heavily than those of less understanding.

Hopefully, this article has resulted in a renewable and upskilling of some faculty in this area. Courses in LISREL are *de rigeur* for many doctoral graduates since 1990 and in doctoral-granting institutions where it is not, such courses need to be added. The history of our oldest academic journals, such as *MIS Quarterly*, is testimony to the requirement for post-millennium researchers to be careful methodologists as well as content specialists.

Guidelines as to when to use each SEM and what statistics need to be reported are clearly necessary. In this tutorial, we have summarized some of the most important aspects to be considered when choosing a SEM technique and we have reviewed the most widely used statistics reported together with their established thresholds. As can be seen from Tables 13 and 14, many studies report only a partial set of these statistics, and, even then, many of these statistics fall short of the common thresholds. As in any other statistical method, when the statistics are not within their respective thresholds, the conclusions drawn based on the analysis are potentially flawed. Applying the appropriate analysis technique, given the research objective and the data, reporting the appropriate statistics, and ensuring that their values are within the established thresholds, is crucial in LISREL [Chin, 1998a, Jöreskog and Sörbom, 1989], PLS [Chin, 1998a], and linear regression models [Cohen, 1988, Cook and Campbell, 1979, Hair et al., 1998, Neter et al., 1990, Nunnally and Bernstein, 1994]. Guidelines for such clear reporting are obviously necessary for good positivist science [Chin, 1998a].

We hope this tutorial provides researchers with a helpful and practical tool toward reaching these objectives.

Editor's note: This article was received on February 29, 2000. It was with the authors for 6 months for 2 revisions, and was published on October 24, 2000

ENDNOTES

[1] LISREL is a registered trademark of SSI: http://www.ssicentral.com/lisrel/mainlis.htm

[2] A February 2000 on-line search on ABI-Inform yielded 194 articles that utilized LISREL analytical techniques. In that many articles using LISREL may not even mention this fact in the abstract or headings, this undoubtedly represents only a portion of all uses of LISREL in business studies.

[3] Professors Dale Goodhue (Carlson School of Management, Minnesota), Fred Davis (University of Arkansas), and Ron Thompson (Wake Forest University) compared these techniques in a panel-tutorial in the 1990 ICIS Conference in Copenhagen. None of their findings are reproduced here in any way, although our results are strikingly similar.

[4] Gefen and Straub [2000] present a theoretical explanation for this lack of consistency and empirical findings which support this interpretation.

[5] LISREL can use one of several estimation techniques. The most commonly used method, and the default, is Maximum Likelihood. This is the method also used in this analysis.

[6] As we shall see later in the paper, some methodologists suggest a .90 threshold for this value while others use a .80 standard. Accordingly, .84 is somewhere in between and, because of the low RMR, was deemed to be acceptable in this case.

[7] See Gefen And Straub's [2000] synopsis of these studies.

[8] In fact, some methodologists interpret PLS as a PCA technique. We do not intend to enter into this debate in this paper, however.

[9] It is useful to note that these distinctions are artificial--there is no substantive difference between a gamma and a beta. Maintaining the distinction achieves some computational efficiency, but that is its only real function.

[10] This is achieved by comparing the χ^2 of the two models and choosing the model with a significantly smaller χ^2 [Segars, 1997].

[11] Mathematically, this is expressed as H(o): $\Sigma = \Sigma(\theta)$, where Σ is the population covariance matrix represented by the covariance matrix of the observed variables, and $\Sigma(\theta)$ is the null hypothesis covariance structure hypothesized by the researcher and written as a function of the research model's parameters, θ [Bollen, 1989].

[12] Multiple-item scales can be introduced into the analysis because correlations among common and unique error terms in LISREL do not have to be automatically assigned a zero value. As in confirmatory factor analysis, this allows overt modeling of the measurement error (in LISREL these matrices are called Θ_δ and Θ_ε, for X and Y measures, respectively). The communality of variance is reflected as loadings on the latent construct that are thought to underlie the multiple items [Bollen, 1989].

[13] See Jöreskog and Sörbom [1989] for a detailed discussion of how variations from the multi-normal distribution affect the fit indexes.

[14] Though some of the estimation techniques, such as ML and GLS, do not actually require a multivariate normal distribution to estimate the model parameters, the estimations they provide still need to be "interpreted with caution" [Jöreskog and Sörbom, 1989] (p. 21). Moreover, the χ^2 statistic may show an unjustified but acceptable fit in sample sizes smaller than 100 [Bollen, 1989, Hair et al., 1992].

[15] Intervals between ranked data points do not have to be equally distributed, as in interval-scaled data. If one assumes that the distances between these points are, on the whole, randomly distributed, statistical tests can be performed on the data. Polychoric distributions, therefore, are the distributions against which the differences between ranks can be checked [Jöreskog and Sörbom, 1989].

[16] LISREL examines the extent to which this measurement error is correlated with the measurement error of other observed variables. The larger these standardized residuals are, the worse the model fit.

[17] A confirmatory analysis attempts to *support* a predefined hypothesized relationship, rather than examine all the possible relationships and select the one that has the best statistical fit.

[18] These are also known as predictor and criterion variables, respectively.

[19] In addition, there are many package-specific assumptions. For example, LISREL assumes (unless explicitly specified otherwise) that the exogenous latent constructs are correlated through shared measurement error while the endogenous constructs are not.

[20] Choice of validation technique is affected to an extent by whether the constructs being tested are formative or reflective [Blalock, 1969]. The types of measurements and scales employed are different depending on whether the measures are reflective of their constructs or formative. Suppose, for instance, the construct "firm performance". It could be measured *formatively* by: (1) an index that compared the pricing of the firm to that of its competitors, (2) revenue generated per employee, and (3) a ratio comparing the IT performance of the business unit with its industrial group. These measures form the construct, but do not really reflect it. A set of measures that *does* reflect its construct would be the perception of a CIO about the strategic value of IT in the firm, measured by four questions with similar low to high semantic anchors. Only constructs that rely on reflective measures need to establish factorial validity since formative measurements may not be highly correlated.

[21] There is one exception to this: when dealing with directly observed variables, LISREL estimates a set of linear regressions among constructs that are composed of one formative directly observed variable [Jöreskog and Sörbom, 1989].

[22] In first-generation regression models, researchers must first establish that the measurement model is correct, typically using a factor analysis to establish convergent and discriminant validity, and then use internal reliability techniques, such as Cronbach's α, to assess construct reliability. Once these validities have been established, researchers combine these observed variables into latent variables, usually through the creation of index values, ignoring the fact that some measurement items may carry more weight than others and ignoring non-common variance. Only then do researchers estimate the specified causation paths between the latent variables – but only one at a time and, again, ignoring non-model specific variance. Testing paths to more than one dependent variable at a time can be accomplished in MANOVA, of course, but this approach is restricted somewhat by the requirement for categorical independent variables.

[23] The total variance of a measurement item is composed of three elements: common, specific, and error variance. Common variance is the variance that reflects the latent construct; it is typically shared with other measurement items. Error variance is variance that is added to the

item due to imperfect measurement. Specific variance is variance that is associated with the unique item alone. First-generation regression models consider only the common variance; LISREL examines all three [Hair et al., 1998].

[24] Other than, of course, the circumstance where there are multiple measures and LISREL requires that one of the item loadings be fixed at 1.0.

[25] Construct reliability is calculated as : $(\Sigma \text{ (std loadings)})^2 / (((\Sigma \text{ (std loadings)})^2 + \Sigma \text{ (std errors)})$

[26] Hair et al. [1998], while recommending that the p-value of the χ^2 should be > .05 also note that "… but .1 or .2 should be exceeded before non significance is confirmed" (p. 654).

[27] Researchers should be aware that some feel that this ratio, like the χ^2 itself, has been entirely discredited as a meaningful statistic.

[28] NFI is calculated as $(\chi^2_{null} - \chi^2_{proposed}) / \chi^2_{null}$

[30] AVE is calculated as: $\Sigma \lambda^2 \; / \; (\Sigma\lambda^2 + \Sigma \text{ Var}(\varepsilon))$

[31] NFI in this case would be calculated as: $\delta = ((\chi^2_{Mo}) - (\chi^2_{Mn})/(\chi^2_{Mo})$

where M_o is the original model and M_n the nested model.

[32] The f^2 statistic is calculated as follows:

$$f^2 = \frac{R^2_{\text{revised-model}} - R^2_{\text{original-model}}}{1 - R^2_{\text{original-model}}}$$

[32] The variance of a calculated variable is a function of the observed variables it is built from [Freund, 1982].

[33] Typically, the p-value in LISREL is set to .05. Thus, when more than 20 comparisons are made, as would be the case in an interaction effect involving more than 3 values, there is a high probability of randomly getting a significant difference.

[34] For a detailed discussion on the nature of causation and why temporal precedence and isolation can never be truly established, see Bollen [1989].

[35] Unless the measurement error is known, at least 2, and preferably at least 3 observed variables should be used for each latent variable in covariance-based SEM [Anderson and Gerbing, 1988].

[36] NFI in this case would be calculated as: $\delta = ((\chi^2_{Mo}) - (\chi^2_{Mn})/(\chi^2_{Mo})$

where M_o is the original model and M_n the nested model.

[37] The f^2 statistic is calculated as follows:

$$f^2 = \frac{R^2_{\text{revised-model}} - R^2_{\text{original-model}}}{1 - R^2_{\text{original-model}}}$$

REFERENCES

Note: Readers are advised to seek out the expanded version of relevant methodological citations on research validities found at the ISWorld Endnote Research Libraries site:

http://www.business.auckland.ac.nz/msis/staff/F.Tan/ISWorld/endnote.htm**.**

This site also presents Endnote libraries for TAM/Diffusion Theory that the reader may find helpful**.**

Adams, D. A., R. R. Nelson, and P. A. Todd (1992) "Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication," *MIS Quarterly* (16) 2, pp. 227-248.

Ajzen, I. and M. Fishbein (1980) *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice Hall.

Anderson, J. C. and D. W. Gerbing (1988) "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin* (103) 3, pp. 411-423.

Anderson, J. C., D. W. Gerbing, and J. E. Hunter (1987) "On the Assessment of Unidimensional Measurement: Internal and External Consistency, and Overall Consistency Criteria," *Journal of Marketing Research* (24), pp. 432-437.

Bagozzi, R. P. and H. Baumgartner (1994) “The Evaluation of Structural Equation Models and Hypothesis Testing,” in R. P. Bagozzi (Ed.) *Principles of Marketing Research*, Cambridge, MA: Blackwell, pp. 386-422.

Bagozzi, R. P. and C. Fornell (1982) “Theoretical Concepts, Measurement, and Meaning,” in , vol. 2 C. Fornell (Ed.) *A Second Generation of Mulivariate Analysis*: Praeger, pp. 5-23.

Barclay, D., R. Thompson, and C. Higgins (1995) "The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use an Illustration," *Technology Studies* (2) 2, pp. 285-309.

Bentler, P. M. (1990) "Comparative Fit Indexes in Structural Models," *Psychological Bulletin* (107) 2 (June), pp. 238-246.

Bentler, P. M. and D. G. Bonett (1980) "Significance Tests and Goodness of Fit in the Analysis of Covariance Structures," *Psychological Bulletin* (88) 3, pp. 588-606.

Blalock, H. M. (1969) *Theory Construction: From Verbal to Mathematical Formulations*. Englewood Cliffs, NJ: Prentice-Hall.

Bollen, K. A. (1989) *Structural Equations with Latent Variables*. New York: John Wiley and Sons.

Bullock, H. E., L. L. Harlow, and S. A. Mulaik (1994) "Causation Issues in Structural Equation Modeling Research," *Structured Equation Modeling* (1) 3, pp. 253-267.

Campbell, D. T. (1960) "Recommendations for APA Test Standards Regarding Construct, Trait, Discriminant Validity," *American Psychologist* (15) August, pp. 546-553.

Chau, P. Y. K. (1996) "An Empirical Assessment of a Modified Technology Acceptance Model," *Journal of Management Information Systems* (13) (2), pp. 185-204.

Chin, W. W. (1998a) "Issues and Opinion on Structural Equation Modeling," *MIS Quarterly* (22) 1 (March), pp. vii-xvi.

Chin, W. W. (1998b) “The Partial Least Squares Approach to Structural Equation Modeling,” in G. A. Marcoulides (Ed.) *Modern Methods for Business Research*, London, pp. 295-336.

Chin, W. W. and A. Gopal (1995) "Adoption Intention in GSS: Relative Importance of Beliefs," DATA BASE *for Advances in Information Systems* (26) 2&3, pp. 42-64.

Chin, W. W. and P. A. Todd (1995) "On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS Research: A Note of Caution," *MIS Quarterly* (19) 2 (June), pp. 237-246.

Cohen, J. (1977) *Statistical Power Analysis for the Behavioral Sciences*, Revised Edition. New York: Academic Press.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition. Hillsdale, NJ: L. Erlbaum Associates.

Cohen, P., J. Cohen, J. Teresi, M. Marchi et al. (1990) "Problems in the Measurement of Latent Variables in Structural Equation Causal Models," *Applied Psychological Measurement* (14), pp. 183-196.

Cook, T. D. and D. T. Campbell (1979) *Quasi Experimentation: Design and Analytical Issues for Field Settings*. Chicago: Rand McNally.

Davis, F. D. (1989) "Perceived Usefulness, Perceived Ease of Use and User Acceptance of Information Technology," *MIS Quarterly* (13) 3 (September), pp. 319-340.

Davis, F. D. and R. P. Bagozzi (1992) "What Do Intention Scales Measure?" *The Journal of General Psychology* (119), pp. 391-407.

Davis, F. D., R. P. Bagozzi, and P. R. Warshaw (1989) "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35) 8 (August), pp. 982-1003.

Doll, W. J., A. Hendrickson, and X. Deng (1998) "Using Davis's Perceived Usefulness and Ease-of-Use Instruments for Decision Making: A Confirmatory and Multigroup Invariance Analysis," *Decision Sciences* (29) 4, pp. 839-869.

Dubin, R. (1976) “Theory Building in Applied Areas,” in *Handbook of Industrial and Organizational Psychology*, Chicago: Rand McNally College Publishing Co., pp. 17-26.

Fenech, T. (1998) “Using Perceived Ease of Use and Perceived Usefulness to Predict Acceptance of the World Wide Web,” in *Computer Networks*, vol. 30, pp. 629-630.

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

Freund, J. E. (1982) *Mathematical Statistics*. New York: Prentice Hall.

Fromkin, H. L. and S. Streufert (1976) “Laboratory Experimentation,” in B. Dunnette (Ed.) *Handbook of Industrial and Organizational Psychology*, Chicago: Rand McNally College Publishing Company, pp. 415-465.

Gefen, D. (1997) Building Users' Trust in Freeware Providers and the Effects of this Trust on Users' Perceptions of Usefulness, Ease of Use and Intended Use. Dissertation, Georgia State University.

Gefen, D. (2000) “Lessons Learnt from the Successful Adoption of an ERP: The Central Role of Trust,” in S. D. Zanakis, G. and C. Zopounidis (Eds.) *Recent Developments and Applications in Decision Making*: Kluwer Academic.

Gefen, D. and M. Keil (1998) “The Impact of Developer Responsiveness on Perceptions of Usefulness and Ease of Use: An Extension of the Technology of the Technology Acceptance Model,” DATA BASE *for Advances in Information Systems* (29), pp. 35-49.

Gefen, D. and D. Straub (2000) "The Relative Importance of Perceived Ease-of-Use in IS Adoption: A Study of e-Commerce Adoption," *JAIS* (forthcoming).

Gefen, D. and D. W. Straub (1997) "Gender Differences in Perception and Adoption of E-Mail: An Extension to the Technology Acceptance Model," *MIS Quarterly* (21) 4 (December), pp. 389-400.

Gerbing, D. W. and J. C. Anderson (1988) "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research* (25) May, pp. 186-192.

Hair, J. F., Jr., R. E. Anderson, R. L. Tatham, and W. C. Black (1992) *Multivariate Data Analysis with Readings*, 4th edition. Englewood Cliffs, NJ: Prentice Hall.

Hair, J. F., Jr., R. E. Anderson, R. L. Tatham, and W. C. Black (1995) *Multivariate Data Analysis with Readings*, 4th edition. Englewood Cliffs, NJ: Prentice Hall.

Hair, J. F., Jr., R. E. Anderson, R. L. Tatham, and W. C. Black (1998) *Multivariate Data Analysis with Readings,* 5th Edition. Englewood Cliffs, NJ: Prentice Hall.

Hanushek, E. A. and J. E. Jackson (1977) *Statistical Methods for Social Scientists*. Orlando, FL USA: Academic Press.

Hendrickson, A. R., P. D. Massey, and T. P. Cronan (1993) "On the Test-Retest Reliability of Perceived Usefulness and Perceived Ease of Use Scales," *MIS Quarterly* (17) 2 (June), pp. 227-230.

Igbaria, M. I., Juhani Maragahh, Hazem (1995) "Why do individuals use computer technology? A Finnish case study," *Information & Management* (29) 5, pp. 227-238.

Jöreskog, K. G. and D. Sörbom (1983) *LISREL: Analysis of Linear Structural RElations by the Method of Maximum Likelihood*, 2nd Edition. Chicago: National Educational Resources.

Jöreskog, K. G. and D. Sörbom (1989) *LISREL7: A Guide to the Program and Applications*, 2nd edition. Chicago: SPSS Inc.

Karahanna, E. and D. W. Straub (1999) "The Psychological Origins of Perceived Usefulness and Perceived Ease-of-Use," *Information & Management* (35), pp. 237-250.

Karahanna, E., D. W. Straub, and N. L. Chervany (1999) "Information Technology Adoption across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs," *MIS Quarterly* (23) 2, pp. 183-213.

Keil, M., P. M. Beranek, and B. R. Konsynski (1995) "Usefulness and Ease of Use: Field Study Evidence Regarding Task Considerations," *Decision Support Systems* (13) 1, pp. 75-91.

Mathieson, K. (1991) "Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior," *Information Systems Research* (2) 3 (September), pp. 173-191.

Montazemi, A. R. C., D. A. Gupta, K. M. (1996) "An Empirical Study of Factors Affecting Package Selection," *Journal of Management Information Systems* (13) 1, pp. 89-105.

Moore, G. C. and I. Benbasat (1991) "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2) 3 (September), pp. 192-222.

Neter, J., W. Wasserman, and M. H. Kutner (1990) *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design*, 3rd edition. Boston, MA: Irwin.

Nunnally, J. C. (1967) *Psychometric Theory*. New York: McGraw-Hill.

Nunnally, J. C. (1978) *Psychometric Theory*, 2nd edition. New York: McGraw-Hill.

Nunnally, J. C. and I. H. Bernstein (1994) *Psychometric Theory, Third Edition*. New York: McGraw-Hill.

Peter, J. P. (1979) "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research* (16) 1 (February), pp. 6-17.

Premkumar, G. and M. Potter (1995) "Adoption of Computer Aided Software Engineering (CASE) Technology: An Innovation Adoption Perspective," Database (26) 2&3(May/August), pp. 105-123.

Ridings, C. and D. Gefen (2000) "Applying TAM to A Parallel Systems Conversion Strategy," *Journal of Information Technology Theory & Application* (2) 2, pp. http://peffers.ba.ttu.edu/jitta/journal/volume2_2/volume2_2p1.htm.

Rose, G. and D. Straub (1998) "Predicting General IT Use: Applying TAM to the Arabic World," *Journal of Global Information Management* (6) 3, pp. 39-46.

Sambamurthy, V. and W. W. Chin (1994) "The Effects of Group Attitudes toward Alternative GDSS Designs on the Decision-making Performance of Computer-Supported Groups," *Decision Science* (25) 2, pp. 215-239.

Segars, A. H. (1997) "Assessing the Unidimensionality of Measurement: A Paradigm and Illustration within the Context of Information Systems Research," *Omega* (25) 1 (February), pp. 107-121.

Segars, A. H. and V. Grover (1993) "Re-Examining Perceived Ease of Use and Usefulness: A Confirmatory Factor Analysis," *MIS Quarterly* (17) 4 (December), pp. 517-525.

Steiger, J. H., A. Shapiro, and M. W. Browne (1985) "On the Multivariate Asymptotic Distribution of Sequential Chi-square Statistics," *Psychometrica* (50pp. 253-264.

Stelzl, I. (1986) "Changing a Causal Hypothesis without Changing the Fit: Some Rules for Generating Equivalent Path Models," *Multivariate Behavioral Research* (21), pp. 309-331.

Straub, D. W. (1989) "Validating Instruments in MIS Research," *MIS Quarterly* (13) 2, pp. 147-169.

Straub, D. W. (1990) "Effective IS Security: An Empirical Study," *Information Systems Research* (1) 3, pp. 255-276.

Straub, D. W. (1994) "The Effect of Culture on IT Diffusion: E-Mail and FAX in Japan and the U.S.," *Information Systems Research* (5) 1 (March), pp. 23-47.

Straub, D. W., M. Keil, and W. Brennan (1997) "Testing the Technology Acceptance Model across Cultures: A Three Country Study," *Information & Management* (33), pp. 1-11.

Straub, D. W., M. Limayem, and E. Karahanna (1995) "Measuring System Usage: Implications for IS Theory Testing," *Management Science* (41) 8 (August), pp. 1328-1342.

Szajna, B. (1994) "Software Evaluation and Choice: Predictive Validation of the Technology Acceptance Instrument," *MIS Quarterly* (17) 3, pp. 319-324.

Szajna, B. (1996) "Empirical Evaluation of the Revised Technology Acceptance Model," *Management Science* (42) (1), pp. 85-92.

Taylor, S. and P. A. Todd (1995a) "Assessing IT usage: The role of prior experience," *MIS Quarterly* (19) 4, pp. 561-570.

Taylor, S. and P. A. Todd (1995b) "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research* (6) 2, pp. 144-176.

Thompson, R., D. W. Barclay, and C. A. Higgins (1995) "The Partial Least Squares Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration," *Technology Studies: Special Issue on Research Methodology* (2) 2 (Fall), pp. 284-324.

Thompson, R. L., C. A. Higgins, and J. M. Howell (1991) "Personal Computing: Toward a Conceptual Model of Utilization," *MIS Quarterly* (15) 1 (March), pp. 125-142.

Venkatesh, V. (1999) "Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation," *MIS Quarterly* (23) 2, pp. 239-260.

Venkatesh, V. and F. D. Davis. (1994) “Modeling the Determinants of Perceived Ease of Use.” *International Conference on Information Systems, Vancouver, British Columbia, 1994*, pp. 213-227.

Venkatesh, V. and F. D. Davis (1996) "A Model of the Antecedents of Perceived Ease of Use: Development and Test," *Decision Sciences* (27) 3 (Summer), pp. 451-481.

APPENDIX A

TAM STUDIES

Study	Subjects
[Davis, 1989] (Study 1)	Knowledge workers
[Davis, 1989] (Study 2)	MBA students
[Davis et al., 1989] (after 1 hour)	MBA students
[Davis et al., 1989] (after 14 weeks)	MBA students
[Mathieson, 1991]	Undergraduate students
[Moore and Benbasat, 1991]	Knowledge workers
[Thompson et al., 1991]	Knowledge workers
[Davis and Bagozzi, 1992] (Study 1)	MBA students
[Davis and Bagozzi, 1992] (Study 2)	MBA students
[Adams et al., 1992] (Study 1)	Knowledge workers
[Adams et al., 1992] (Study 2)	Knowledge workers
[Hendrickson et al., 1993]	Undergraduate students
[Segars and Grover, 1993]	Adams et al.'s (1992) data
[Hendrickson et al., 1993]	Undergraduate students
[Sambamurthy and Chin, 1994]	Knowledge workers
[Sambamurthy and Chin, 1994]	Undergraduate students
[Venkatesh and Davis, 1996]	Undergraduate students
[Straub, 1994]	Knowledge workers
[Szajna, 1994]	MBA students
[Chin and Gopal, 1995]	Knowledge workers
[Premkumar and Potter, 1995]	Knowledge workers
[Straub et al., 1995] (Model 1)	Knowledge workers
[Straub et al., 1995] (Model 2)	Knowledge workers
[Keil et al., 1995]	Knowledge workers
[Taylor and Todd, 1995b]	Students
[Taylor and Todd, 1995a]	Students
[Igbaria, 1995]	MBA students
[Montazemi, 1996]	Knowledge workers
[Chau, 1996] (Study 1)	Administrative/clerical staff
[Chau, 1996] (Study 2)	Administrative/clerical staff
[Szajna, 1996] (Study 1: pre-implementation)	Graduate business students
[Szajna, 1996] (Study 2: post-implementation)	Graduate business students
[Gefen and Straub, 1997]	Knowledge workers in airline industry
[Straub et al., 1997]	Knowledge workers in airline industry
[Gefen, 1997]	MBA students
[Gefen and Keil, 1998]	Knowledge workers
[Doll et al., 1998]	Undergraduate students
[Fenech, 1998]	Undergraduate students
[Rose and Straub, 1998]	Knowledge workers
[Karahanna and Straub, 1999]	Knowledge workers
[Karahanna et al., 1999] (Study 1)	Knowledge workers
[Karahanna et al., 1999] (Study 2)	Knowledge workers
[Venkatesh, 1999]	Knowledge workers
[Gefen, 2000]	Knowledge workers
[Ridings and Gefen, 2000]	Knowledge workers
[Gefen and Straub, 2000]	MBA Students

APPENDIX B

INSTRUCTIONS TO SUBJECTS AND INSTRUMENTATION

INSTRUCTIONS:

As part of an ongoing study on Internet use, we would be grateful if you could devote 10 minutes to completing this instrument.

1. Please logon to the Internet and access www.travelocity.com
2. Use the Web-site to search for a flight to Heathrow Airport (London) next month.
3. Then, please fill in the instrument below.

Please circle the appropriate category:

Gender	M , F
Age group	15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 50-54, 55-59, 60-64, 65-69, above 70
What language do you speak at home (English, Italian, Hindi, Cantonese, etc.)?	
Have you ever bought products on the World Wide Web	Yes, No
How many times have you used Travelocity.com?	
Have you given your credit card number on the Web?	Yes, No

Please indicate your agreement with the next set of statements using the following rating scale:

1	2	3	4	5	6	7
Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree

Code*	Item	Agree Disagree
EOU1	Travelocity.com is easy to use.	1 2 3 4 5 6 7
EOU2	It is easy to become skillful at using Travelocity.com.	1 2 3 4 5 6 7
EOU3	Learning to operate Travelocity.com is easy .	1 2 3 4 5 6 7
EOU4	Travelocity.com is flexible to interact with .	1 2 3 4 5 6 7
EOU5	My interaction with Travelocity.com is clear and understandable .	1 2 3 4 5 6 7
EOU6	It is easy to interact with Travelocity.com.	1 2 3 4 5 6 7
PU1	Travelocity.com is useful for searching and buying flights .	1 2 3 4 5 6 7
PU2	Travelocity.com improves my performance in flight searching and buying.	1 2 3 4 5 6 7
PU3	Travelocity.com enables me to search and buy flights faster.	1 2 3 4 5 6 7
PU4	Travelocity.com enhances my effectiveness in flight searching and buying.	1 2 3 4 5 6 7
PU5	Travelocity.com makes it easier to search for and purchase flights.	1 2 3 4 5 6 7
PU6	Travelocity.com increases my productivity in searching and purchasing flights.	1 2 3 4 5 6 7
IUSE1	I am very likely to buy books from Travelocity.com.	1 2 3 4 5 6 7
IUSE2	I would use my credit card to purchase from Travelocity.com.	1 2 3 4 5 6 7
IUSE3	I would not hesitate to provide information about my habits to Travelocity.	1 2 3 4 5 6 7

Thank You!

* Students did not receive the item codes****.

APPENDIX C

EXAMINING NESTED MODELS IN SEM

In covariance-based SEM, examining nested models is accomplished by comparing the χ^2 statistic of the original model with the χ^2 of a "nested" model. Generally speaking, a model M_2 is nested within another model M_1 (i.e., $M_2 < M_1$) if it contains exactly the same constructs and if its freely estimated parameters are a subset of those estimated in M_1. If the difference in χ^2 between the two models is insignificant given the difference in degrees of freedom between the models, then the additional path in the "nested" model does not significantly improve the model. In such a case, the parsimonious, theoretical model should be chosen. Comparing models in this manner can be used for causation paths (β and γ), item loadings (λ), and correlation (Φ and Ψ).

Anderson and Gerbing [1988] suggest using this method to assess a theoretical model by estimating five nested plausible alternative model specifications. The five models are: (1) a *saturated model* (M_s) that links all constructs; (2) a *null model* M_n that contains no paths among the constructs; (3) a *theoretical model* M_t representing the theoretical model to be tested; (4) a *constrained model* M_c that constrains theoretically defensible paths in M_t; and (5) a *unconstrained model* M_u that frees theoretically defensible paths in M_t. These five structural models represent a nested sequence of: $M_n < M_c < M_t < M_u < M_s$. The null model of the Generic Theoretical Network from Figure 5 is presented in Figure 6; the saturated model is presented in Figure 7.

The four tests required to examine the five nested models are asymptotically independent [Steiger et al., 1985], each test examining a no difference null hypothesis between two nested structural models. However, since the χ^2 statistic depends on sample size, trivial differences between the two nested models can cause a significant difference in the χ^2 [Anderson and Gerbing, 1988, Bentler and Bonett, 1980]. In order to overcome this problem, the NFI (Normed Fit Index) statistic comparing a nested model M_n with an original model M_o should be used [Bentler and Bonett, 1980].[34] Ranging from 0 to 1, this index represents the increment in fit obtained in evaluating two hierarchical step-up models. It should be noted, though, that any nested model comparison is applicable only for the comparison of models that differ only in one path [Anderson and Gerbing, 1988], in a manner analogous to stepwise linear regression.

Nested model comparison is also available in PLS [Thompson et al., 1995], although not through examining the difference of significance in χ^2 values. In PLS, the significance of a nested model containing an additional path is examined by comparing the R^2 of the revised model with that of the original model using an f^2 statistic.[35] The additional path can be considered as having a small, medium, or large effect if f^2 is above .02, .15 or .35, respectively [Chin, 1998b], as in Cohen's [1988] analysis of power in linear regression. Unlike LISREL and linear regression, however, PLS cannot be set to automatically perform a stepwise analysis.

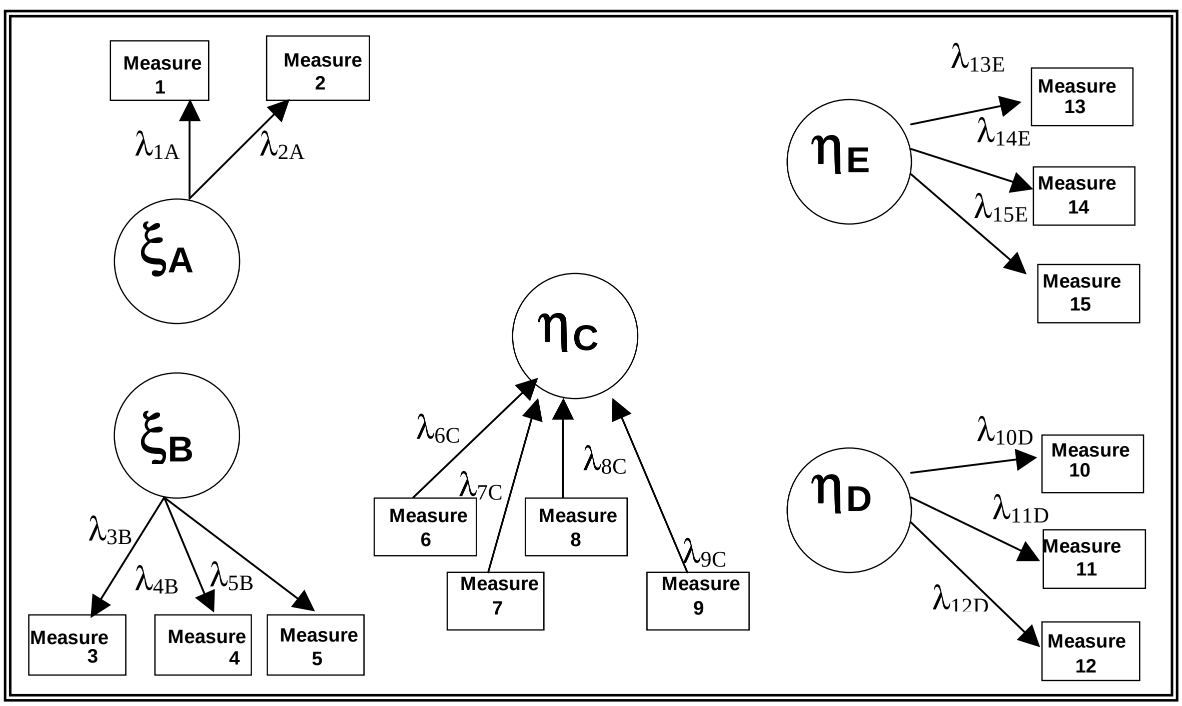


Figure 6. Null Model of the Generic Theoretical Network

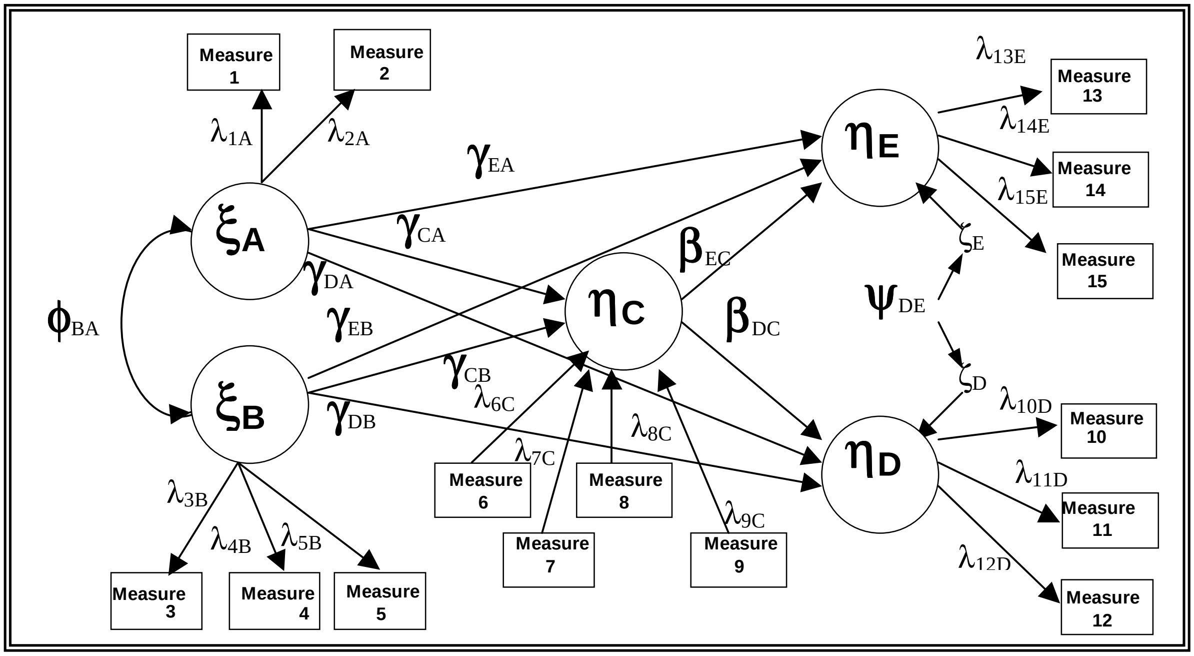


Figure 7. Saturated Model of the Generic Theoretical Network

What would nested models look like in the TAM running example? There are theoretical reasons for both specifying a path between EOU and IUSE and for not specifying this path. Gefen and Straub [2000] present empirical evidence that the significance of this relationship depends on the intrinsic or extrinsic nature of the task for which the IT is being used. If this theoretical refinement were tested with nested models, then the path would be specified in a theoretical model and then unspecified (constrained or removed) in a nested model. With an additional path specified over the theoretical model, a third, less constrained model could be easily imagined where both EOU and a variable like SPIR impact IUSE.

While there has been little nested model testing in TAM studies (see Karahanna and Straub [1999] for an example of its employment, however), there have been numerous explorations along this vein in IS research in general (see Table 13). Nested models allow the IS researcher to see where the model can be theoretically improved, which is particularly important in TAM research.

GLOSSARY

This glossary presents three types of terms that are used in this article:

1. Statistical
2. TAM constructs
3. Other terminology

Both abbreviations and specialized terms are included.

STATISTICAL TERMS:

- **AGFI:** Adjusted Goodness of Fit Index. Within covariance-based SEM, statistic measuring the fit (adjusted for degrees of freedom) of the combined measurement and structural model to the data.

- **AMOS:** A covariance-based SEM, developed by Dr. Arbuckle, Published by SmallWarters and marketed by SPSS as a statistically equivalent tool to LISREL. Details are available at http://www.spss.com/amos/ .

- **ANOVA:** Univariate analysis of variance. Statistical technique to determine, on the basis of one dependent measure, whether samples are from populations with equal means.

- **AVE:** Average Variance Extracted. Calculated as $(\Sigma\lambda_i^2)/((\Sigma\lambda_i^2) + (\Sigma(1-\lambda_i^2))$, the AVE measures the percent of variance captured by a construct by showing the ratio of the sum of the variance captured by the construct and measurement variance.

- **CFA:** Confirmatory Factor Analysis. A variant of factor analysis where the goal is to test specific theoretical expectations about the structure of a set of measures.

- **Construct validity**: One of a number of subtypes of validity that focuses on the extent to which a given test is an effective measure of a theoretical construct.

- **Cronbach's alpha:** Commonly used measure of reliability for a set of two or more construct indicators. Values range between 0 and 1.0, with higher values indicating higher reliability among the indicators.

- **DV:** Dependent Variable. Presumed effect of, or response to, a change in the independent variable(s).

- **EQS:** A covariance-based SEM developed by Dr. Bentler and sold by Multivariate Software, Inc. EQS provides researchers with the ability to perform a wide array of analyses, including linear regressions, CFA, path analysis, and population comparisons. Details are available at http://www.smallwaters.com/.

- **Equivalence of Models:** When two or more models produce exactly the same fit indexes in LISREL making model interpretation based on statistics alone problematic. This can easily happens in LISREL when changing the direction of an assumed causation or changing a causation path (β) into a shared correlation (ψ).

- **Endogenous construct:** Construct that is the dependent or outcome variable in at least one causal relationship. In terms of a path diagram, there are one or more arrows leading into the endogenous construct.

* **Exogenous construct:** Construct that acts only as a predictor or "cause" for other constructs in the model. In terms of a path diagram, the exogenous constructs have only causal arrows leading out of them and are not predicted by any other constructs in the model.

* **F statistic** (F-ratio): tests the hypothesis that the amount of explained variation is greater than that explained by chance alone. The F statistic is calculated as the ratio of the sum of squared error explained by the model divided by its degrees of freedom to the sum of squared error about the average divided by its degrees of freedom. This provides the ratio of the variance of the prediction errors. When employed in the procedure entitled ANOVA, the obtained value of F provides a test for the statistical significance of the observed differences among the means of two or more random samples.

* **f^2:** A statistic used to assess whether a change in R-square is substantive between nested models in PLS in which an additional path is added.

* **Factor analysis:** A statistical approach that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factor).

* **First generation statistical techniques:** A general term relating to correlation based analyses methods that preceded LISREL and PLS. These methods include linear regression, ANOVA, MANOVA, etc. These technique require researchers to analyze the item loadings on the latent variables separately from the linkage of the independent variables to the dependent variable.

* **Formative variables:** Observed variables that "cause" the latent variable, i.e., represent different dimensions of it.

- **GFI:** Goodness of Fit Index. Within covariance-based SEM, statistic measuring the absolute fit (unadjusted for degrees of freedom) of the combined measurement and structural model to the data.

- **Heteroscedasticity:** Unequal variance among the measurement items.

- **Holistic analysis:** Analysis combining both structural and measurement models.

- **IV:** Independent Variable. Presumed cause of any change in a response or dependent variable(s).

- **Latent variable:** Research construct that is not observable or measured directly, but is measured indirectly through observable variables that reflect or form the construct.

- **Linear models:** A systematic relationship between two variables that can be described by a straight line.

- **Linear regression:** A linear regression uses the method of least squares to determine the best equation describing a set of x and y data points.

- **LISREL:** A procedure for the analysis of **LI**near **S**tructural **REL**ations among one or more sets of variables and variates. It examines the covariance structures of the variables and variates included in the model under consideration. LISREL permits both confirmatory factor analysis and the analysis of path models with multiple sets of data in a simultaneous analysis.

- **Loading (Factor Loading):** Weighting which reflect the correlation between the original variables and derived factors. Squared factor loadings are the percent of variance in an observed item that is explained by its factor.

- **LOGIT:** Special form of regression in which the criterion variable is a non-metric, dichotomous (binary) variable.

- **MANOVA:** Multivariate analysis of variance. Statistical technique that can be used to simultaneously explore the relationship between several categorical independent variables and two or more metric dependent variables.

- **Measurement model:** Sub-model in structural equation modeling that (1) specifies the indicators for each construct, and (2) assesses the reliability of each construct for estimating the causal relationships.

- **Multicollinearity:** Extent to which an independent variable varies with other independent variables. Excessively high multicollinearity challenges the statistical assumption that the independent variables are truly independent of each other. Some techniques, such as PLS, are distribution-free and do not make the assumption of independence. Linear regression assumes low or no multicollinearity and provides a VIF statistic to assess its extent. LISREL assumes that all the IVs are independent of each other, at once.

- **Nested models:** Models that utilize the same constructs, but differ in terms of the number or types of causal relationships represented. When they differ by only one causal path, they are said to be “nested” in one another.

- **NFI:** Normed Fix Index. Within covariance-based SEM, statistic measuring the normed difference in χ^2 between a single factor null model and a proposed multi-factor model.

- **Observed indicator / variables:** Observed value used as an indirect measure of a concept or latent variable that cannot be measured or observed directly.

- **Over Fitting:** Ex-post facto "adjustments" of the research model to the data: customizing the research model to sample-specific correlations. The resulting model represents the data but is not adequate for hypotheses testing. One way of handling this type of hindsight analysis is by splitting the data into two datasets. Building the model based on one dataset and then testing the hypotheses on the other [Cliff, 1983].

- **Parallel correlational patterns (see Unidimensionality):** Additional correlations between measurement items that are not reflected in a factor analysis or in the measurement model. For example, if items A1, A2, A3 and A4 load together on the same factor in a factor analysis but, additionally, A1 and A2 are highly correlated to each other in another dimension that is not captured in the factor analysis. Confirmatory factor analysis in LISREL can detect such cases.

- **PLS:** Partial Least Squares. A second generation regression model that combines a factor analysis with linear regressions, making only minimal distribution assumptions.

- **PCA:** Principal Components Analysis. Statistical procedure employed to resolve a set of correlated variables into a smaller group of uncorrelated or orthogonal factors.

- **Polychoric correlation:** Measure of association employed as a replacement for the product-moment correlation when both variables are ordinal measures with three or more categories. LISREL usually assumes that the correlation matrix being analyzed is a Pearson matrix of interval or ratio data. If the correlations are non-parametric, adjustments in the LISREL model have to be made and a WLS estimation, rather than ML, should be used.

- **Reflective variables:** Observed variables that "reflect" the latent variable and as a representation of the latent variable should be unidimensional and correlated.

- **Reliability**: Extent to which a variable or set of variables is consistent in what it is intended to measure. If multiple measurements are taken, the reliable measures will all be very consistent in their values.

- **R-square or R^2:** Coefficient of determination. Measure of the proportion of the variance of the dependent variable about its mean that is explained by the independent variable(s). R-square is derived from the F statistic. This statistic is usually employed in linear regression analysis and PLS.

- **RMR:** Root Mean Square Residual. Within covariance-based SEM, statistic assessing the residual variance of the observed variables and how the residual variance of one variable correlates with the residual variance of the other items.

- **Second generation data analysis techniques:** Techniques enabling researchers to answer a set of interrelated research questions in a single, systematic, and comprehensive analysis by modeling the relationships among multiple independent and dependent constructs simultaneously.

* **SEM:** Structural Equation Modeling. Multivariate technique combining aspects of multiple regression (examining dependence relationships) and factor analysis (representing unmeasured concepts with multiple variables) to estimate a series of interrelated dependence relationships simultaneously.

* **SMC:** Squared Multiple Correlation. Explained variance of each latent variable. Used in LISREL, similar to R-square in regression.

* **Statistical conclusion validity:** Type of validity that addresses whether appropriate statistics were used in calculations that were performed to draw conclusions about the population of interest.

* **Stepwise linear regression:** Regression model that is developed (and run) in stages where new independent variables are added to the regression model one at a time in a decreasing order of increased R-square so long as the resulting increase in the F statistic is still significant.

* **Structural model:** Set of one or more dependence relationships linking the model constructs. The structural model is most useful in representing the interrelationships of variables between dependence relationships.

* **Structural relationships:** Linkages between research constructs (or variables) that express the underlying structure of the phenomenon under investigation. Sometimes referred to as "paths." Structural relationships are often represented as hypotheses in the research design.

* **Unidimensionality:** Similar to the concept of reliability, a unidimensional construct is one in which the set of indicators has only one underlying trait or concept in common.

TAM CONSTRUCTS:

- **EOU:** Ease Of Use.
- **IUSE:** Intentions to Use.
- **PU:** Perceived Usefulness.
- **SPIR:** Social presence-information richness.
- **TAM:** Technology Acceptance Model.

OTHER TERMINOLOGY

- **GPA:** Grade Point Average.
- **IT:** Information Technology.
- **Case study:** Research method involving the intense examination of a single unit (person, group, or organization) by the researcher, where no independent variables are manipulated nor confounding variables controlled.
- **Field study:** Research method involving non-experimental inquiries occurring in natural systems. Researchers using field studies cannot manipulate independent variables or control the influence of confounding variables.

* **Field experiment:** Research method involving the experimental manipulation of one or more variables within a naturally occurring system and subsequent measurement of the impact of the manipulation on one or more dependent variables.

* **Free simulation experiment:** A form of experimentation in which the IVs are not manipulated in order to examine independent variables - dependent variables relationships, but are allowed to move freely over their natural range. Subjects are all presented with identical experimental tasks and respond to these tasks with freely chosen choices.

* **Laboratory experiment:** Research method taking place in a setting especially created by the researcher for the investigation of the phenomenon. Within a laboratory experiment, the researcher has control over the independent variable(s) and the random assignment of research participants to various treatment and non-treatment conditions.

* **Travelocity.com:** Travel site on the Internet providing secure online reservation capabilities for air, car, hotel and vacation reservations, plus access to a vast database of destination and other travel information. http://www.travelocity.com

ABOUT THE AUTHORS

Marie-Claude Boudreau (gs04mcb@panther.gsu.edu) is an assistant professor in the MIS Department at the University of Georgia. She received a Diplôme d'Enseignement Supérieur Spécialisé from l'École Supérieure des Affaires de Grenoble, an M.B.A. from l'Université Laval, and a Ph.D. from Georgia State University. Her current research investigates the consequences of information systems in organizations. She has published in *Information Systems Research, MIS Quarterly, The Academy of Management Executive* and *Information Technology & People*.

David Gefen (gefend@drexel.edu) is an Assistant Professor of MIS at Drexel University, where he teaches Strategic Management of IT, Database Analysis and Design, and Programming languages at the MBA level. He received his Ph.D. degree in CIS from Georgia State University and a Master of Sciences from Tel-Aviv University. His research specialization is in IT adoption, the Internet, culture and gender effects, and e-trust. His current research interests focus on psychological and relational processes involved in the successful implementation of technological innovations. His research findings are published or forthcoming in leading academic and professional journals, including the *MIS Quarterly, DATA BASE for Advances in Information Systems, Omega: the International Journal of Management Science, Journal of the Association for Information Systems,* and the *Journal of Information Technology Theory & Application.* Dr. Gefen is also the author of several encyclopedia articles on IT adoption and IT security, and a book chapter on trust and ERP adoption.

Detmar W. Straub (dstraub@gsu.edu) is the J. Mack Robinson Distinguished Professor of Information Systems at Georgia State University, Detmar has conducted research in the areas of e-Commerce, computer security, technological innovation, and international IT studies. He holds a DBA in MIS from Indiana and a PhD in English from Penn State. He has published over 80 papers in journals such as *Management Science, Information Systems Research, MIS Quarterly, Organization Science, Communications of the ACM, Journal of MIS, Information & Management, Communications of the AIS, Academy of Management Executive,* and *Sloan Management Review.* He is currently an Associate Editor for *Management Science* and *Information Systems Research.* Former Co-Editor of DATA BASE *for Advances in Information Systems* and an Associate Editor and Associate Publisher for *MIS Quarterly*, he has consulted widely in industry in the computer security area as well as in the areas of e-Commerce and technological innovation. He teaches courses at Georgia State in the areas of: Electronic Commerce Strategy, IT Strategies for Management, Systems Integration and IT Outsourcing, International IT Policies and Issues, and Computer Security Management.

Copyright ©2000, by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from ais@gsu.edu

Communications of the Association for Information Systems

ISSN: 1529-3181

EDITOR
Paul Gray
Claremont Graduate University

AIS SENIOR EDITORIAL BOARD

Henry C. Lucas, Jr. Editor-in-Chief New York University	Paul Gray Editor, CAIS Claremont Graduate University	Phillip Ein-Dor Editor, JAIS Tel-Aviv University
Edward A. Stohr Editor-at-Large New York University	Blake Ives Editor, Electronic Publications Louisiana State University	Reagan Ramsower Editor, ISWorld Net Baylor University

CAIS ADVISORY BOARD

Gordon Davis University of Minnesota	Ken Kraemer University of California at Irvine	Richard Mason Southern Methodist University
Jay Nunamaker University of Arizona	Henk Sol Delft University	Ralph Sprague Universityof Hawaii

CAIS EDITORIAL BOARD

Steve Alter University of San Francisco	Tung Bui University of Hawaii	Christer Carlsson Abo Academy, Finland	H. Michael Chung California State University
Omar El Sawy University of Southern California	Jane Fedorowicz Bentley College	Brent Gallupe Queens University, Canada	Sy Goodman University of Arizona
Ruth Guthrie California State University	Chris Holland Manchester Business School, UK	Jaak Jurison Fordham University	George Kasper Virginia Commonwealth University
Jerry Luftman Stevens Institute of Technology	Munir Mandviwalla Temple University	M.Lynne Markus Claremont Graduate University	Don McCubbrey University of Denver
Michael Myers University of Auckland, New Zealand	Seev Neumann Tel Aviv University, Israel	Hung Kook Park Sangmyung University, Korea	Dan Power University of Northern Iowa
Maung Sein Agder College, Norway	Margaret Tan National University of Singapore, Singapore	Robert E. Umbaugh Carlisle Consulting Group	Doug Vogel City University of Hong Kong, China
Hugh Watson University of Georgia	Dick Welke Georgia State University	Rolf Wigand Syracuse University	Phil Yetton University of New South Wales, Australia

ADMINISTRATIVE PERSONNEL

Eph McLean AIS, Executive Director Georgia State University	Jennifer Davis Subscriptions Manager Georgia State University	Reagan Ramsower Publisher, CAIS Baylor University