Fitting Covariance Models for Theory Generation

Joerg Evermann
Memorial University of Newfoundland, jevermann@mun.ca

Mary Tate
Victoria University, mary.tate@vuw.ac.nz

Follow this and additional works at: https://aisel.aisnet.org/jais

Fitting Covariance Models for Theory Generation

Joerg Evermann
Memorial University of Newfoundland
jevermann@mun.ca

Mary Tate
Victoria University of Wellington
mary.tate@vuw.ac.nz

Abstract

Covariance-based structural equation modeling (CB-SEM) is an increasingly popular technique for analyzing quantitative data in Information Systems research. As such, it is traditionally viewed as a method to test theory, rather than build it. However, many of the theoretical models tested with this technique in IS research show significant differences between the model and the data. This paper shows that as part of the pursuit of model fit, researchers using CB-SEM can provide deeper insights into a phenomenon, allowing us to build theories based on quantitative data.

Keywords: *Structural Equation Modeling, Theory Building, Quantitative Analysis, Data Analysis, Research Methods.*

Fitting Covariance Models for Theory Generation

1. Introduction

Theories are sets of propositions that relate constructs, bounded by a specified context (Bacharach, 1989). Theories provide causal explanations of a phenomenon and testable hypotheses (Bacharach, 1989; Dubin, 1969; Gregor, 2006; Popper, 1968). Covariance-based structural equation modeling (CB-SEM) is a quantitative method that has become increasingly used to test theories in Information Systems. Theoretical constructs and propositions are represented by variables and hypotheses (Bacharach, 1989). Specifically, in CB-SEM, latent variables represent theoretical constructs, and the regression relationships between them represent hypothesized causal propositions between constructs. CB-SEM allows simultaneous testing of multiple relationships and provides a test, using the χ^2 test statistic, of how well the statistical model, representing the theory, fits the observed data.

However, the IS literature frequently ignores the result of this test of model fit. Of 54 CB-SEM studies published in MISQ, ISR, JMIS, and JAIS between 2004 and 2008, only three achieve a well-fitting model, and only one of the remaining 51 studies acknowledges the misfit and makes an effort to identify and discuss the substantive reasons for the misfit. In fact, IS methodologists even recommend ignoring the χ^2 test (Gefen, Straub, & Boudreau, 2000).

As a brief aside, the IS field is not unique in this. For comparison, we examined CB-SEM studies in the Association of Management Journal and Organization Science within the same period. Of 26 CB-SEM studies, only eight show good fit by the χ^2 test, but all studies either explicitly or implicitly (by proceeding to interpret the parameter estimates) claim good fit.

We believe that models that do not fit the observed data are useful, because, given the extensive theory building and data collection effort that goes into any research study, we can learn much from them. Rather than ignoring misfit, researchers need to identify the reasons for the misfit in order to refine and improve the theory on which their models are based: "We need to understand what is problematic if we are to do better next time around" (Hayduk, Cummings, et al., 2007, p.845).

To this effect, the goal of this paper is to demonstrate that the process of model diagnostics and pursuit of model fit can lead to rich and interesting insights into the data, akin to theory building approaches in qualitative inquiry. In doing this, we wish to impress on the reader the importance of paying close attention to instrument development and the selection of measurement indicators for constructs.

The paper proceeds with a general discussion of theory building, followed by a discussion of the use of data in the Information Systems field. We then describe theory (model) testing in CB-SEM and the issues around the χ^2 test of model fit. This is followed by two sections that motivate theory building based on model test failure and describe how this can be done in the CB-SEM context. We then give an illustrative example and show how diagnostic processes can create new theory. We follow with a summary of our approach and a discussion of its limitations. Finally, we conclude the paper with a general discussion.

2. Theories and Theory Building

Fundamentally, science has been argued to operate either according to the inductive method, which abstracts theories from specific data, or the hypothetico-deductive method, which generates testable hypotheses from theories for falsification. Qualitative enquiry frequently uses the inductive method, and qualitative data is frequently viewed as the only source from which to build theories. For example, Shah and Corley (2006) suggest that "theory building often requires the rich knowledge that only qualitative methods can provide" (p.1821) Inductive case studies (Eisenhardt, 1989; Eisenhardt & Graebner, 2007) and the grounded theory approach (Glaser & Strauss, 1967; Urquhart, Lehman, & Myers, 2010) are popular qualitative theory building approaches in the management and information systems literature. By contrast, quantitative survey research tends to be used in association with the hypothetico-deductive method, and it is often assumed that it can only be used for theory testing:

"one of the most important limitations of cross-sectional, survey-based research is that it can *only* be used to test theory" (Shah & Corley, 2006, p.1822, emphasis added).

Quantitative data from surveys or experiments is the dominant form of data collected in the IS field (Palvia et al., 2004; Chen & Hirschheim, 2004) and is primarily used for theory testing in the hypothetico-deductive model of science. Given its importance to our field, the process of theory building in this model has received comparatively little attention. However, it is recognized that theory building is based to a large extent on intuition and other creative processes. For example, the philosopher of science Mario Bunge (1962) suggests,"The invention of hypotheses, the devising of techniques, and the designing of experiments, are clear cases of imaginative operations" (p.80). Similarly, Root-Bernstein (1989) states, "Mastery of facts and techniques alone does not make a scientist. The difference between a technician and a discoverer is imagination" (p.313). And the Nobel laureate Sir Peter Medawar (1969) writes, "Imaginativeness and a critical temper are both necessary at all times" (p.58). Dubin (1969) uses the terms "discovering" and "invention" to describe the creative process of building theories, and in the management literature, Weick (1989) characterizes theory construction as "disciplined imagination."

While creativity forms the basis for theory building in the hypothetico-deductive model, the same writers acknowledge that data is a primary prerequisite for this to occur: "Yet, the fruitful invention and the deep insight … do no emerge *ex nihilo*. … There is no new knowledge that is not somewhat determined by prior knowledge" (Bunge, 1962, p.80). Medawar (1969) echoes this by saying that "the imagination cannot work *in vacuo*: There must be something to be imaginative about, a background of observation … before the exploratory dialogue can begin" (pp.44-45).

The idea of using quantitative data for theory building is also congruent with the grounded theory approach, even though grounded theory is typically understood as a method only for qualitative data analysis. Glaser and Strauss (1967) point out, "There is no fundamental clash between the purposes and capacities of qualitative and quantitative methods or data … each form of data is useful for both verification and generation of theory" (p.17f). Further, they say, *"The freedom and flexibility that we claim for generating theory from quantitative data will lead to new strategies and styles of quantitative analysis, with their own rules yet to be discovered"* (p.186, emphasis in original). The proposal in this paper may be seen as an illustration of such a new analysis strategy.

In summary, the hypothetico-deductive model of science operates predominantly with quantitative data. Its theories are ultimately rooted in data (though not necessarily data that has been formally and explicitly captured, but also in anecdotal evidence or observations), even though the theory building process has not been made as explicit as in the inductive model of science, which operates primarily on qualitative data. However, as we shall argue in the next section, despite the importance of data to theory building, the IS field is "data poor," and this can inhibit the creative process of generating new theory.

3. Data Poverty in the Information Systems Field

In the previous section, we have argued that data is a necessary basis on which to build theories. However, in contrast to the natural sciences, usually held up as the paragon of the hypothetico-deductive model, the social sciences are data poor. Dubin (1969) aptly characterizes this difference:

> *The knowledge base in natural sciences is fundamentally a body of experimental and descriptive fact. The natural scientists fund facts and then demand that their theoretical models make sense out of the accumulating body of data. … By way of contrast, social scientists have tended to accumulate theories and theoretical models. The social scientist funds theory and not data. … The behavioural scientist tends to accumulate belief systems and call this the theory of his field … In the social sciences, our relative indifference to facts has left us with a comfortable and very narrow range of alternative theoretical models with which to deal. The social scientist's imagination, when it feeds only upon its logical capacity to combine and recombine a fixed set of elements, can create only a limited range of alternative theories … I am therefore urging that the research stance toward theory building*

among behavioural scientists be that of constant alertness to the descriptive knowledge of the domain about which they wish to theory. It is facts against which the adequacy of the theory is always tested. It is also facts out of which the theory is developed in the first place. (pp.238-240)

This has been recognized in the Information Systems field, where then editor-in-chief of JAIS, Kalle Lyytinen (2009) writes, "In most cases good data precedes good theory. In the IS field this may come as a surprise to many, as in most cases the role of data is relegated only to the latter part of the research cycle (as an indicator or falsification)" (p.717) and further, "I also argue that data poverty – rather than theory poverty – has created barriers for the development of a unique and strong IS discipline. It has limited the scope and scale of IS research projects and degraded the likelihood of reaching strong results with salience" (p.718).

In summary, while (quantitative) data forms the basis for theory building also in the hypothetico-deductive model, the IS community is relatively data poor, as the data collected are used only for model testing, and not for theory building. We believe that this is partially a consequence of researchers claiming that their theories are confirmed by the data and, thus, no investigation and theory building is necessary. The next section shows that this is not typically the case. As a result, most data from survey-based studies and other quantitative methods, which together make up between one third (Palvia et al, 2004) and two thirds (Chen & Hirscheim, 2004) of all IS research, could and should be exploited for theory building.

4. Theory Testing Using Structural Equation Models

In the last 20 years, quantitative data analysis in the Information Systems field has increasingly applied structural equation modeling techniques. Among survey-based studies, which make up between a quarter (Palvia et al., 2004) and almost half (Chen & Hirscheim, 2004) of all IS research, this is a widely-used technique. In covariance-based structural equation modeling (CB-SEM), theoretical constructs are represented by latent variables, and hypothesized relationships between them are represented by linear regression relationships. CB-SEM is typically used for theory testing within the hypothetico-deductive model of science. It allows the simultaneous testing of multiple relationships and provides a test of how well the statistical model, which represents the theory, fits the observed data.

The test of model fit is a χ^2 test, which tests whether the model, with its constraints and estimated parameters, implies covariances that are within random sampling errors of the observed covariances (Bollen, 1989). For model-implied covariances that are not significantly different from the observed ones, the χ^2 test should be insignificant. However, the result of this test of model fit is frequently ignored in the IS literature. Of 54 studies published in MISQ, ISR, JMIS, and JAIS between 2004 and 2008 that use CB-SEM, only three achieve a non-significant χ^2 test for their model, and only one of the remaining 51 studies acknowledges the misfit and makes an effort to identify and discuss the substantive reasons for the misfit. More seriously, nine studies do not even present the χ^2 statistic, and 39 studies claim good model fit despite a significant χ^2 test, based on a variety of approximate fit indices. In fact, methodologists in the IS field even recommend to ignore the χ^2 test (Gefen et al., 2000).

Ignoring the χ^2 test is tantamount to ignoring evidence that falsifies a theory, and the lack of rigorous model testing has been the subject of impassioned advocacy by psychometric researchers. Hayduk, Cummings, et al. (2007, p.848) warn, "Overlooking indications of potentially huge problems are the kinds of things lawyers will gladly describe as malfeasance, dereliction of responsibility, or absence of due diligence." Researchers instead resort to approximate fit indices and threshold values to argue for well fitting models despite evidence to the contrary: "Where it has all gone badly wrong in this type of SEM model testing is that many investigators have actively avoided the statistical test of fit of their models, in favor of a collection of ad hoc indices which are forced to act like 'tests of fit'" (Barrett, 2007, p.819). It is hard to avoid the conclusion that Barrett (2007, p.819f) could also be describing the IS literature: "Indeed, one gets the feeling that social scientists cannot actually contemplate that most of their models do not fit their data, and so invent new ways of making sure that by referencing some kind of ad hoc index, that tired old phrase 'acceptable approximate fit' may be rolled out as the required rubber stamp of validity."

One of the main issues critics raise about the χ2 test is sample size dependence. However, as Wilks (1938) was careful to point out when developing the test, the test statistic, which is based on a likelihood ratio, is centrally χ2 distributed when the hypothesized model is true. A centrally χ2 distributed test statistic is sample size independent. Only when the hypothesized model is not true is the test statistic non-centrally χ2 distributed, and, therefore, dependent on sample size (Curran, Bollen, Paxton, & Kirby, 2002).

Critics argue that all models are misspecified to some degree, leading to a non-central χ2 test statistic that increases with sample size. For example, Bentler (2007, p.828) suggests, "A model is liable always to be misspecified, and hence to be rejected by any 'exact' test." Similarly, Goffin (2007, p.835) argues, "The models we develop in psychology should virtually never be presumed to contain the whole truth and therefore be subjected to a test of perfect fit." Steiger (2007, p.894) concludes, "Why test a hypothesis that is always false?" We are not this fatalistic and disagree with this sentiment, as do Hayduk, Cummings, et al. (2007), who state in a response to Barrett (2007): "We cannot prevent Barrett from claiming that all his models are detectably wrong in general, but we can encourage everyone to strive for models that are properly specified, not wrong. Observing that at least some wrong models are more assuredly detected by larger samples (because of decreased sampling variability) is good methodological news to those seeking proper models!" (p.844). The sample size dependence of the χ2 test for misspecified models is a desirable feature as, like any other statistical technique, it provides increased statistical power to detect these models as the sample size increases. We believe that this increased power is desirable, as we should be submitting our theories to the best and strongest tests we have available when working in the hypothetico-deductive model of science. Even if it were not possible in principle to construct a true model, it should be incumbent upon researchers to at least strive for exact fit instead of dismissing the test. Barrett (2007) suggests that anything less might be construed either as "simple intellectual laziness on the part of the investigator" or as self-deception about the quality of our theories.

In summary, we believe that the IS community should apply more rigorous standards to model testing. Instead of dismissing the χ2 test, we believe that striving for well-fitting models by using the collected data not only for theory testing, but also for theory building, allows us to make better use of a significant amount of data in our field, turning it from a data poor field, where data are not used to inform new theory but essentially discarded after a single use, into one that is data rich, allowing researchers to build on a body of data for new theory generation.

5. Model Diagnostics for Theory Development

In the CB-SEM literature, researchers have extensively discussed the notion of using data for theory development. For example, McIntosh (2007, p.861) notes the importance of using model diagnostics for theory development, writing that "merely settling for close fit could hinder the advancement of knowledge in a given substantive field, since there is little impetus to seek out and resolve the reasons why exact fit was not attained."

While we argue for more rigorous model testing based on the χ2 test, we do not mean to imply that ill-fitting models should be rejected from publication. Instead, given the careful theory development and extensive data collection effort that forms the basis of every quantitative study, we agree with Hayduk, Cummings, et al. (2007, p.845) who argue,

> *Attentively constructed and theoretically meaningful models that fail ought to be carefully discussed and published… Any area that is unable to openly acknowledge and examine the deficiencies in its current theories is hampered from proceeding toward better theories… If a model fails, the authors should not proceed to discuss the model as if it were 'OK anyway'. They should publish a discussion of 'how the world looks from this theory/model perspective', and their diagnostic investigations of 'how and why this theory/model perspective on the world fails'. We need to understand what is problematic if we are to do better next time around.*

Barrett (2007) makes sensible suggestions about how to deal with this further:

> *If the [distributional] assumptions appear reasonable, … and an author is curious to explore further, then begin examining the residual matrix for clues as to where misfit is occurring, adjust the model accordingly, refit, and proceed in this way to explore minor-adjustment alternative models until either fit is achieved, or where it becomes obvious that something is very wrong with the a priori theory.*

In summary, attending to model fit and acknowledging failure of theoretical models provides the impetus for continued theory development in the way that Hayduk, Cummings, et al. (2007) and Barrett (2007) suggest. Carefully collected data should be assumed to be always right; they cannot be dismissed and deserve to be explained, even when this explanation occurs post-hoc. The next section discusses how model diagnostics and theory building can occur in the context of CB-SEM analyses.

6. Model Diagnostics and Model Modifications in Structural Equation Models

Model diagnostics in a CB-SEM context is very much an art rather than a science. The CB-SEM technique itself provides a number of statistical tools, but their use depends very much on the researcher. The matrix of residuals provides an indication for how well the model has explained each observed covariance, and large and/or systematic residuals can offer clues on how to modify the model for better fit. Modification indices (MI) or Lagrange Multipliers (LM) indicate to what extent the different model constraints (such as omitted or fixed paths) contribute to the ill-fit of the model. Expected parameter changes (EPC) indicate the change of a model parameter when a constraint is relaxed. Traditional recommendations (e.g., Kaplan, 1990) have focused on using MI and EPC to decide whether to free certain parameters. However, neither the residuals nor the MI or EPC should be followed blindly (Bollen, 1990), as they may lead to model parameters that are difficult to interpret (Bollen, 1990), are not useful in situations where two or more model parameters should be constrained (Steiger, 1990), and cannot identify situations where additional latent variables may be required (Hayduk, 1990).

Using traditional CB-SEM aids like MI, EPC, and fit statistics, researchers have developed automatic data-driven model specification search procedures (MacCallum, 1986). Model specification search can begin with the a priori model, a null model, or a saturated model, and non-significant paths are constrained, and constraints with significant modification indices are relaxed. However, as MacCallum (1986) points out, the likelihood of arriving at the true population model using specification search is low, even under the best conditions. Variations on this include heuristic search methods such as Tabu search (Marcoulides, Drezner, & Schumacker, 1998) or Ant Colony optimization (Marcoulides & Drezner, 2003). Automated, data-driven specification searches are widely available in CB-SEM software (Schumacker, 2006). The limitations that apply to the diagnostic aids also apply to the specification searches that are based on them.

A second kind of automatic model specification search is the TETRAD procedure (Scheines, Spirtes, Glymour, Meek, & Richardson, 1998), which has been recommended and seen some recent application in the IS field (Lee, Barua, & Whinston, 1997; Im & Wang, 2007; Liu, 2009). The TETRAD procedure is based on examining the covariance or correlation matrix and identifying sets of indicators that have zero partial correlations (for indicator triplets) or vanishing tetrads (for models with latent variables). The "Purify" algorithm of TETRAD removes cross-loading indicators, leaving uni-dimensional measurement models. With these, the "MIMbuild" algorithm searches for a structural model. The TETRAD procedure produces a set of models that are plausible from the data. These still need to be evaluated using traditional CB-SEM. While TETRAD differs from traditional specification searches, it remains a data-driven procedure.

The process for which we advocate in the present paper is different from data-driven model modification approaches and specification searches. It recognizes Bollen's (1990) recommendation for greater emphasis on theory/substantively driven revisions and use of data-driven modifications only as a last

resort. We agree with Ting's (1998) critical stance toward TETRAD and other data-driven approaches, when he admonishes us: "Do not let any software dictate the course of model selection. … Researchers should take an active role in formulating alternative models rather than looking for a quick fix" (p.163). Hayduk (1990) sums this up nicely:

> *In summary, if one adopts the philosophy that structural equation models are supposed to be prods to sluggish imaginations, sparks that ignite insight, keys that unlock advancement, or hammers that forge progress from burning issues, we will have to do better than merely searching through the list of potentially-freeable coefficients, no matter how diligently and with how much technical sophistication we conduct the search (p.196).*

Hence, instead of applying these statistical techniques, we argued in a previous section that theory building (and, therefore, model improvement) requires an intimate understanding of theory and measures, as well as intuition about possible and plausible alternative theoretical explanations. Hence, our focus in the following sections is not on the statistical diagnostics but on the theoretical constructs and their operationalization in survey items. We follow Hayduk's advice that "thoughtful reconsideration of the relevant theory and data gathering procedures remain the best guide to model revision" (Hayduk, 1987, p.179). Our proposal is similar to a strategy proposed by Jöreskog (1993) and is based on separating the model constructs for individual evaluation and subsequently assessing the fit of the combined model.

It is important to note that "model modifications based on observed discrepancies might be capitalizing on chance sampling fluctuations in the data, improving fit at the expense of theoretical meaningfulness. … Researchers who engage in this exercise … must justify any modifications they make and, preferably, any resulting respecified models should be replicated with independent data" (Markland, 2007, p.856). If no independent sample is available, a split-sample approach should be used (Hayduk, 1987; Jöreskog, 1993). Whichever approach is taken, researchers need to clearly report results: "We believe it is particular important that authors distinguish between results based on estimation of theory based models that were specified prior to analysis of the data and results based on post-hoc modifications of a priori models. … If the eventual model was derived through multiple, sequential modifications of an a priori model, then authors should describe the history of the development of the final model from the a priori model" (Hoyle & Panter, 1995).

7. Illustrative Example - Data

For illustrating how quantitative data can be used to generate theory, we use data from a previous study by Chin, Johnson, and Schwarz (2008) on the Technology Acceptance Model (TAM). We use this study as an illustrative example only. We do not wish to specifically critique Chin et al.'s study but use it for the following reasons. First, TAM is one of the most frequently used theories in IS research, well accepted, and familiar to most researchers. The measurement items have been virtually unchanged over many studies. Second, Chin et al. (2008) present one of the few studies in IS research that publish covariance or correlation matrices to allow readers to independently verify their conclusions and extend their research, as we do here. Third, Chin et al. (2008) claim that their model fits the data well, despite the fact that the $\chi2$ test shows significant discrepancies.

While Chin et al. (2008) provide the covariance matrix, they do not report the fit function. We assume maximum-likelihood (ML) fit, as this is the most commonly used fit function. To verify our assumptions about the estimation and fit methods, we first reproduce the results obtained by Chin et al. (2008). Using the reported sample size of 283, the ML method on a model with reflective indicators and uncorrelated errors reproduces the reported results. We now turn to a more detailed analysis of their first TAM instrument, which employs the items from Davis (1989). The results presented by Chin et al. (2008) show a $\chi2=181$ on df=101 for a significant p-value, indicating that the model does not fit the observed covariance data.

The complete model is shown in Figure 1, and the measurement items are shown in Table 1 (all items measured on 7-point Likert scales). The Technology Acceptance Model specifies three constructs,

perceived ease of use (abbreviated as "eou"), perceived usefulness (abbreviated as "use"), and predicted usage (abbreviated as "lu"). The covariance matrix can be found in Appendix C of (Chin et al., 2008) and is, therefore, not reproduced here.
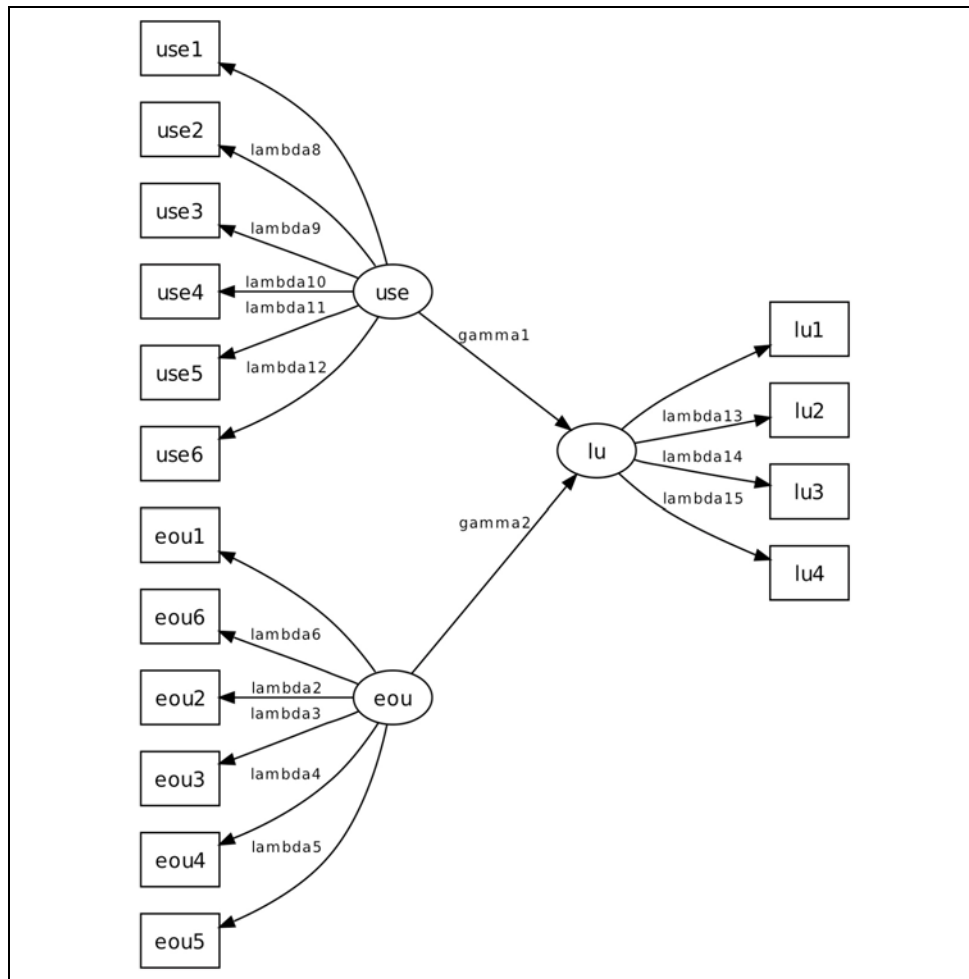

Figure 1. TAM Model

Table 1. Instrument Items from (Chin et al., 2008), All Measured on a 7-Point Likert Scale	
Item	**Wording**
Perceived Ease of Use	
eou1	Learning to operate the (task-related) platform portions of (system) is easy for me
eou2	I find it easy to get the (task-related) portions of (system) to do what I want it to do
eou3	My interaction with the (task-related) portions of (system) has been clear and understandable.
eou4	I find the (task-related) portions of (system) to be flexible to interact with.
eou5	It is easy for me to become skillful at using the (task-related) portions of (system)
eou6	I find the (task-related) portions of (system) easy to use.
Perceived Usefulness	
use1	Using (system) as a (technology type) enables me to (accomplish tasks) more quickly
use2	Using (system) improves my (ability to accomplish task)
use3	Using (system) as a (technology type) increases my productivity

Table 1. Instrument Items from (Chin et al., 2008), All Measured on a 7-Point Likert Scale (continued)	
use4	Using (system) as a (technology type) increases my effectiveness in accomplishing (task)
use5	Using (system) makes it easier to do my (task)
use6	I find (system) useful in my (task completion)
Predicted Usage	
Lu1	If the choice of a (technology type) platform were up to me, it would likely be (system).
Lu2	If I need to (accomplish task) and the choice was up to me, I would expect to use (system) as a (task-related) platform.
Lu3	If asked, I would likely recommend (system) as a (task-related) platform
Lu4	For future (task-oriented) tasks that are totally within my control, I would probably use (system) as a (task-oriented) platform

In the remainder of the text, we present only the graphical models and χ2 test statistics. Complete model specifications and estimation results for all models to allow future critique of this analysis are provided in the appendix. We used the sem package in the open-source R system for our analysis (Fox, 2006).

8. Illustrative Example - Application

We began by estimating the complete model shown in Figure 1, which produced significant misfit ($\chi2=182.25$, df=101, p=0.0000), as reported by Chin et al. (2008). Examining the residuals (Table 2), i.e., the differences between model-implied and observed covariances, showed that the model explains the variances well but the covariances less so. However, there was no single variable that had unusually large residuals with other variables, nor were there any patterns immediately identifiable in the residual matrix (Table 2). Hence, applying the proposal of Jöreskog (1993), we examined each construct individually, to assess whether Perceived Ease of Use, Perceived Usefulness, and Predicted Usage individually fit the factor models that were hypothesized for them.

Table 2. Excerpt of the Matrix of Raw Residuals							
	lu4	lu3	lu2	lu1	eou1	eou2	eou3
lu4	.0000	.0099	-.0070	-.0099	-.0589	.0356	.0435
lu3	.0099	0.000	-.0089	-.0107	-.0828	.0388	.0559
lu2	-.0070	-.0089	.0000	.0204	-.0186	-.0275	.0245
lu1	-.0099	-.0107	.0204	.0000	-.1242	-.0736	-.0245
eou1	-.0589	-.0828	-.0186	-.1242	.0000	.0951	.0153
eou2	.0356	.0388	-.0275	-.0736	.0951	.0000	-.0148
eou3	.0435	.0559	.0245	-.0245	.0153	-.0148	.0000

8.1. Perceived Ease of Use

With free error variances, uncorrelated errors, and one path coefficient constrained for scaling purposes, a six-indicator EOU factor model showed significant misfit ($\chi2=34.693$, df=9, p=0.0001), showing that EOU is not a simple construct as hypothesized (Figure 2).

Instead of focusing primarily on statistical clues such as residuals and modification indices, we began to critically analyze the EOU item wording in detail. Table 1 shows that, rather than being homogeneous in meaning, and interchangeable, i.e., being questions about the same underlying phenomenon, they are in fact different. Item EOU4 deals with flexibility, item EOU3 deals with clarity and understandability, items EOU1 and EOU5 are concerned with the ease of learning and mastering

of a system, and only EOU2 and EOU6 deal directly with ease of use. Therefore, we began modeling EOU2 and EOU6 as two items of a single EOU factor (Figure 3). To achieve identification, we chose to constrain error variances to 20 percent of the item variances. We defer the motivations for such constraints to the discussion section.
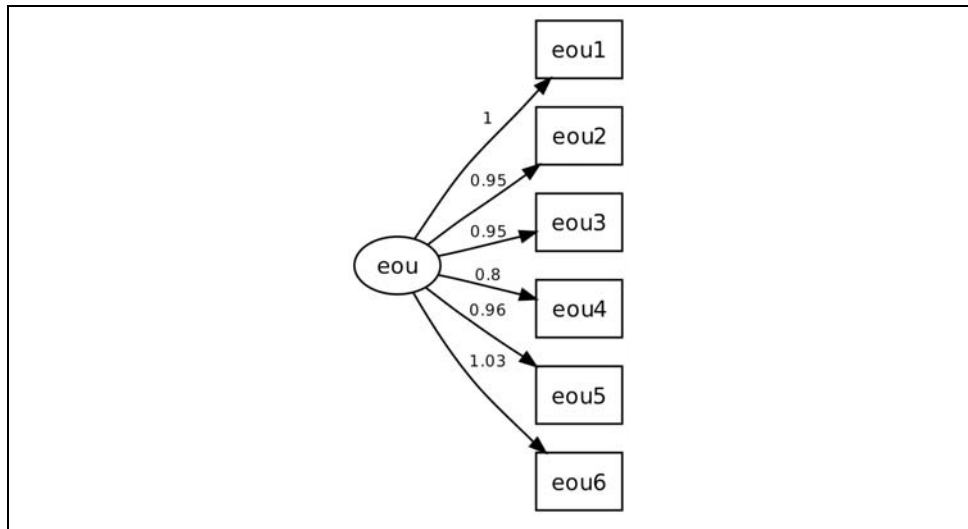
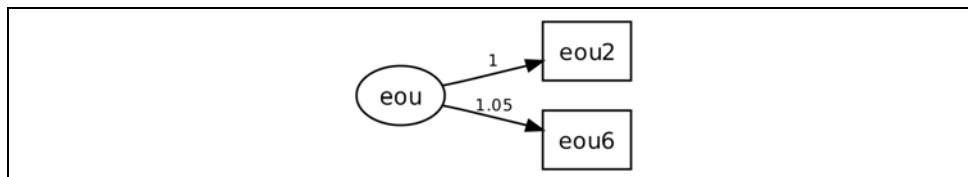

Figure 2. EOU Factor Model (χ2=34.693, df=9, p=0.0001)

Figure 3. EOU Items (χ2=0.1951, df=1, p=0.6587)

Based on the item wording, EOU4 expresses Perceived Flexibility of the system, rather than Perceived Ease of Use. While a system that is perceived to be flexible may be perceived to be easy to use, the two are not the same. Hence, we model Perceived Flexibility as one cause (among other possible ones) of Perceived Ease of Use. Similarly, EOU3 expresses clarity and understandability, which may also lead to a system that is perceived to be easy to use, but, again, the constructs are not identical. We again constrain the error variances of the indicator variables at 20 percent of the item variances. The model is shown in Figure 4, and the estimation showed good fit (χ2=4.1967, df=3, p=0.241). Note that in the estimation results below, flexibility is not significantly related to ease of use. This is plausible, as a system that is flexible may offer more options to the user, making it more complex and, hence, harder to use, thus negating any inherent increase in ease of use by flexibility. A flexible system may also provide multiple ways of achieving the same result, possibly confusing the novice user and making it appear harder to use, further diminishing any positive effect that flexibility might have.
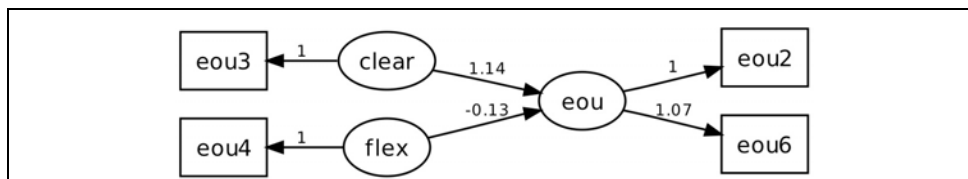

Figure 4. EOU, Flexibility, and Clarity (χ2=4.1967, df=3, p=0.2410)

Having established the ease of use latent as above, we must take care to prevent interpretational confounding. This is a situation in which, due to free parameters, a latent takes on new meaning by virtue of significantly different path coefficients to other variables. The path coefficient from EOU to EOU6 differs in the two estimations above (1.0479 versus 1.068587), and we need to test whether the difference is significant. Restricting the coefficient in the latter model to its value in the former allows a χ2 difference test with one degree of freedom, which, in this case, is non-significant (Δχ2=0.1657).

The item wording suggests that items EOU1 and EOU5 are related to the ease of learning a system, rather than the ease of using a system. Again, the constructs may be related, but they are not identical. In fact, the direction of causality is not even clear: If a system is easy to use, it may well be perceived to also be easy to learn. It is also plausible that if a person perceives a system to be easy to learn, she will also believe the system to be easy to use. Adding the Learnability construct with the indicators EOU1 and EOU5 into the model in Figure 4 proved to be unsuccessful. Modeling it as an independent cause of EOU showed misfit (χ2=47.657, df=10, p=0.0000), as did modeling it as a consequence only of EOU (χ2=184.71, df=12, p=0.0000). Realizing that Perceived Flexibility and Perceived Clarity might influence Learnability in the same way that they impact Perceived Ease of Use, we modeled it as a consequence of both Perceived Flexibility and Perceived Clarity (Figure 5). While showing improvement, this model also did not fit our data (χ2=37.309, df=10, p=0.0001).

Figure 5. EOU, Learnability, Flexibility, and Clarity (χ2=37.309, df=10, p=0.0001)
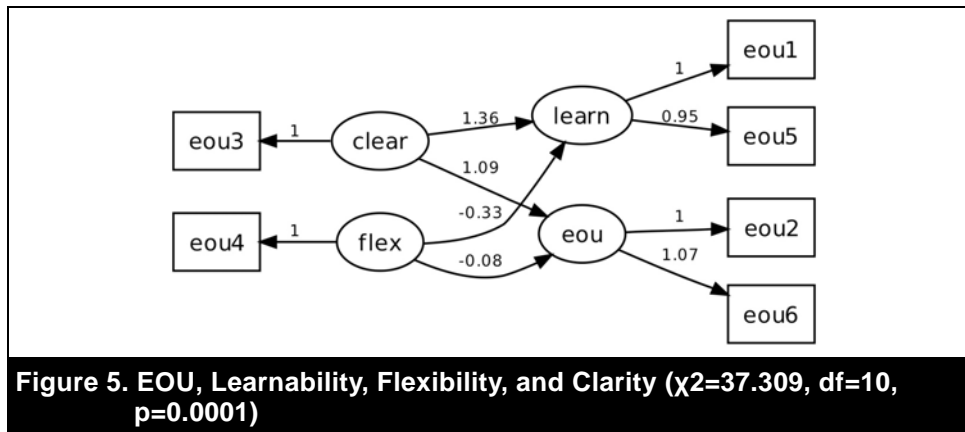
Closely examining the residuals for these three models yielded no clues as to what caused the model misfit, as large residuals occurred between differing and unrelated variables. We finally examined whether items EOU1 and EOU5 (under the assumption of 20 percent error variance) are even caused by the same latent. The model showed significant misfit (χ2=7.742, df=1, p=0.0054). This leads us to question the validity of items EOU1 and EOU5 (or the reliability of the items, if larger error variance would yield a fitting model). While their wording suggests they are related, and are not direct indicators of Perceived Ease of Use, none of the plausible models fit the data.

8.2. Perceived Usefulness

We applied the same approach to the Perceived Usefulness construct that we used for Perceived Ease of Use. An initial factor model with six indicators, free error variances, uncorrelated errors, and one path coefficient constrained for scaling yielded significant misfit (Figure 6) (χ2=32.607, df=9, p=0.0002).

While the items for Perceived Usefulness appear more homogeneous than the items for EOU, we can still make out distinct groups. Items USE1, USE2, and USE3 deal with efficiency, performance, and productivity, i.e., with amount of work per time unit. On the other hand, USE4 deals with effectiveness, which is quite distinct from efficiency. USE6 is the most immediate and direct measure of usefulness. The only problematic item is USE5, which deals with making the task easier. While this is clearly

related to usefulness, effectiveness, and efficiency, the causal relationships are not immediately clear. A system can be effective without making the task any easier, or it can make the task easier, but not be effective. Similarly, efficiency and easing of the task are not necessarily related. Separating efficiency, effectiveness, and usefulness leads to the model in Figure 7 (free error variances where possible, uncorrelated errors, one path coefficient constrained for scaling each latent), which fits the observed data ($\chi2=7.6693$, df=4, p=0.1045). The separation of effectiveness and efficiency is also supported in the literature (Son, Kim, & Riggins, 2006).

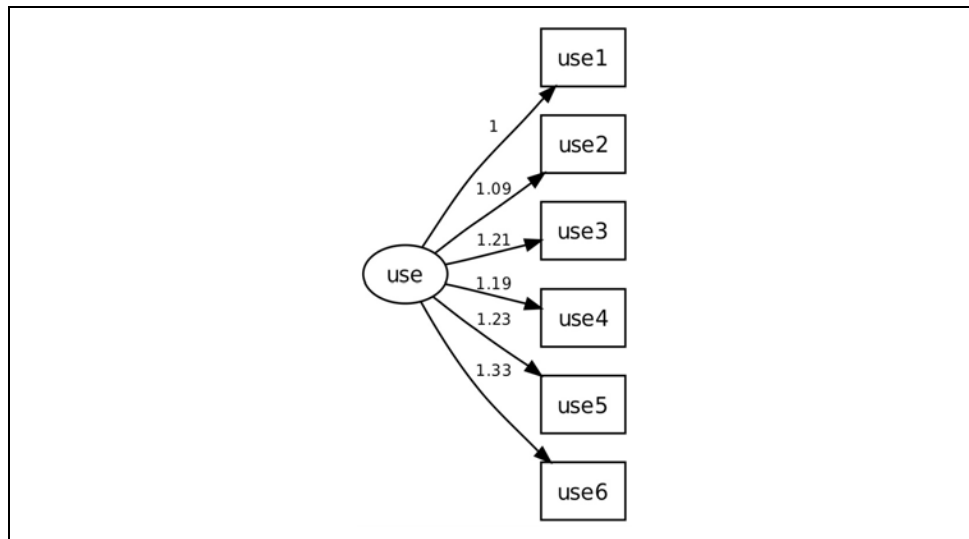**Figure 6. Usefulness Factor Model ($\chi2=32.607$, df=9, p=0.0002)**
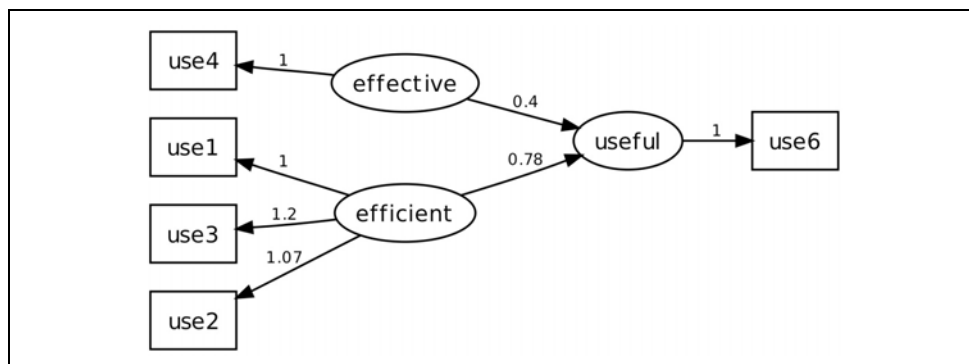
Figure 7. Usefulness Model ($\chi2=7.6693$, df=4, p=0.1045)

8.3. The Exogenous Latents

At this stage we combine the two exogenous constructs, Perceived Ease of Use and Perceived Usefulness. While the initial TAM results suggest that Perceived Ease of Use influences Perceived Usefulness (Davis, 1989), Chin et al. (2008) model them as merely exogenously co-varying. As we have additional constructs available, i.e., Perceived Clarity and Perceived Flexibility, which are both exogenous, we hypothesize that increases in Perceived Flexibility might also cause increases in Perceived Usefulness, irrespective of the Perceived Ease of Use. As we have seen above (Figure 5), a flexible system can plausibly be argued to be harder to use than an inflexible one. However, it is also plausible that a flexible system is more useful, as it is likely to be able to accomplish more tasks or allow more efficient ways to perform the task. This could serve as an explanation for the effects of EOU on Usefulness suggested by Davis (1989). Hence, a first model to be tested is shown in Figure 8. Again, error variances are free where possible, errors are uncorrelated, and all four exogenous latents are allowed to co-vary. This model fits the data well ($\chi2=28.756$, df=22, p=0.1520).
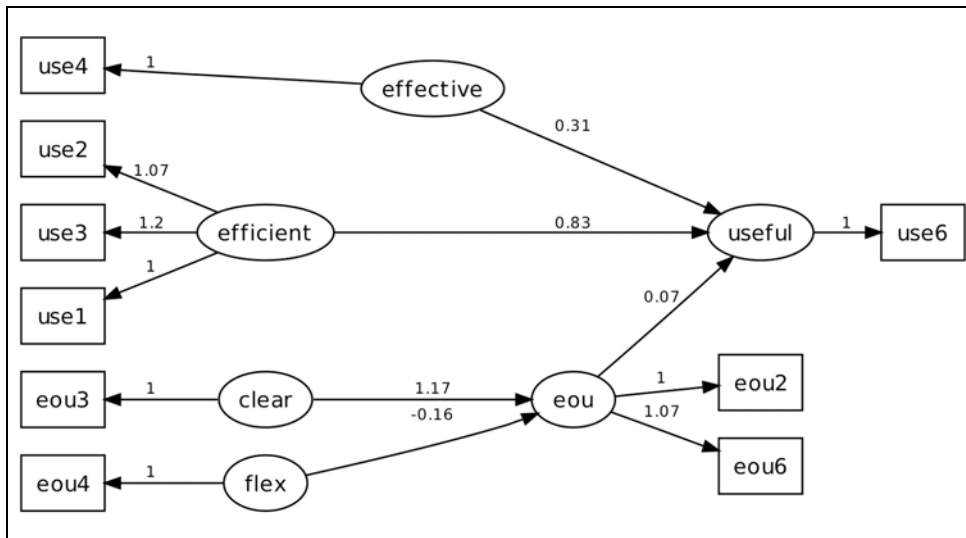
Figure 8. Exogenous Latents (χ2=28.756, df=22, p=0.1520)

To assess interpretational confounding, we conduct χ2 difference tests by restricting each path coefficient in Figure 8 to its value in Figure 5 or Figure 7. All χ2 difference tests are non-significant, showing that path coefficients, and hence, the meaning of the latent constructs, remain unchanged.

8.4. Predicted Usage

The final construct, Predicted Usage, is represented by an endogenous latent variable with four hypothesized indicators. Examining the indicators closely shows that LU3 is slightly different, as it deals with recommendations for others, rather than the subject's own future intended usage. However, a test of the hypothesized four-indicator factor model, with free error variances, uncorrelated errors, and one path coefficient fixed for scaling showed good fit (χ2=1.6517, df=2, p=0.4379), so we used this sub-model for the next step.
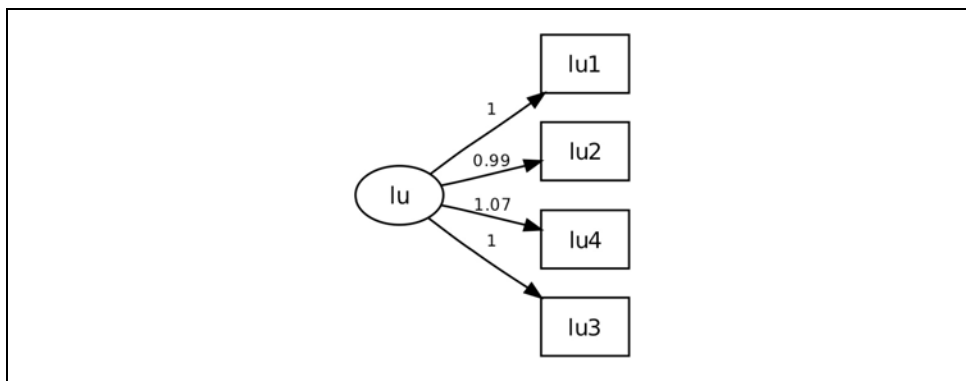
Figure 9. Predicted Usage Factor Model (χ2=1.6517, df=2, p=0.4379)

8.5. Full Model

We are now ready to assemble the full model, based on the model in Figure 8. The TAM theory suggests that both Perceived Usefulness and Perceived Ease of Use cause future predicted usage of a system (Chin et al., 2008; Davis, 1989), so we introduce these two paths. We estimate the model again with errors free where possible, uncorrelated errors, and correlated exogenous latents. The model fits the data (χ2=67.338, df=55, p=0.1228). This model is shown in Figure 10.
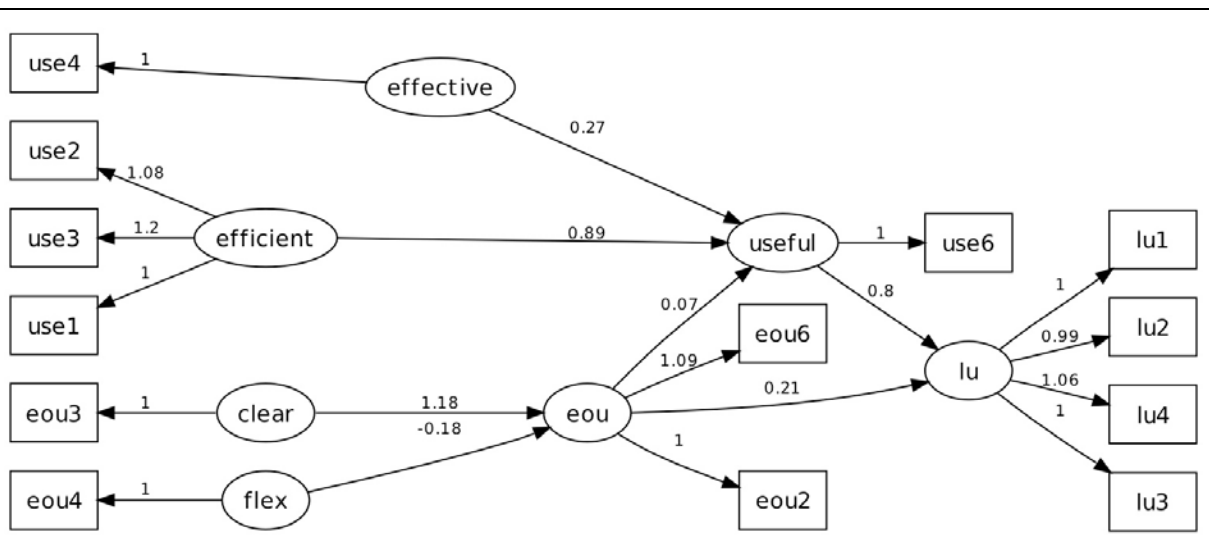
Figure 10. Full Model (χ2=67.338, df=55, p=0.1228)

We are now in a position to try to add the "uncooperative" items USE5, EOU1, and EOU5 into the model. At this stage, there are more possibilities for these items to fit into a theoretically plausible model. Additionally, because of the larger model, adding these items adds more degrees of freedom, which might permit a better fit, statistically. We begin by adding USE5 as another indicator of Perceived Usefulness (Figure 11). The χ2 statistic shows that the model does not fit the data (χ2=89.837, df=67, p=0.0328). This confirms the earlier problems with USE5 and supports the conclusion that USE5 is not an indicator of Perceived Usefulness.
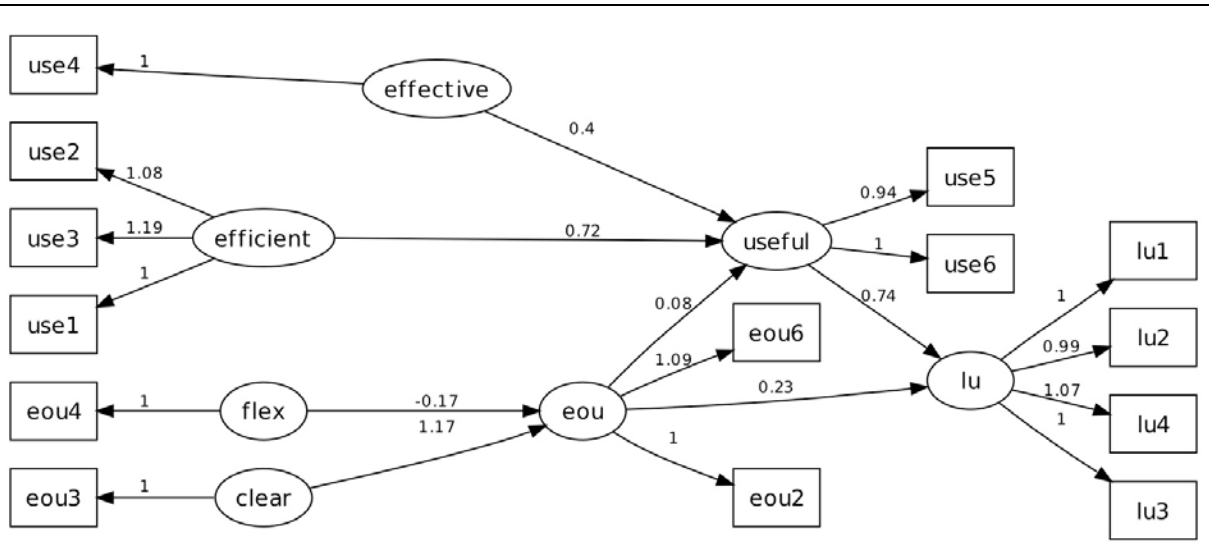
Figure 11. Full Model with USE5 (χ2=89.837, df=67, p=0.0328)

Instead of trying to fit USE5, our attempt to add EOU5 as an indicator of Learnability, which in turn is caused by Perceived Ease of Use, shows more promise (χ2=80.204, df=67, p=0.1292) and is shown in Figure 12. Adding EOU1 as another indicator of Learnability reduces model fit to unacceptable levels (χ2=123.05, df=79, p=0.0011). Adding both EOU1 and USE5 (because of the combined degrees of freedom they contribute) also does not lead to improvement (χ2=151.85, df=94, p=0.0001). Hence, the most complex plausible model that we can fit to the data is that in Figure 12.
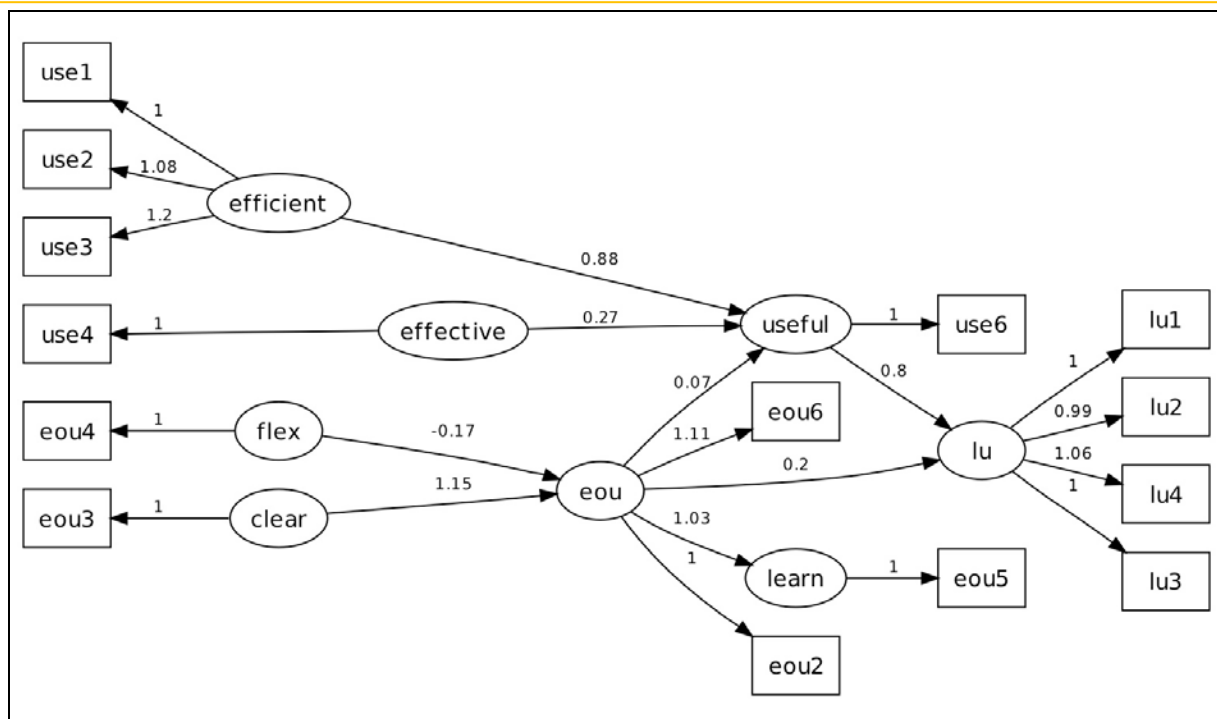
Figure 12. Full Model with EOU5 (χ2=80.204, df=67, p=0.1292)

Because we have been unable to fit all observed variables, our model has fewer degrees of freedom than the one presented in Chin et al. (2008). Hence, it may be argued that the good fit is due to low statistical power. While power estimates in structural equation modeling are problematic due to the lack of an alternative hypothesis (Saris & Satorra, 1993) and ideally require a simulation study (Muthén & Muthén, 2002), indicative power calculations using the method of MacCallum, Browne, and Sugawara (1996) show that a model with 67 degrees of freedom and a sample size of 283 has a power of .943 to detect deviations from exact fit.

Our model makes a number of assumptions reflected in model constraints, necessitated by model identification requirements. Such constraints should ideally be based on prior theory or previous empirical results. We chose to constrain error variances to 20 percent of the item variances, based on a number of reasons. A previous study by Chin and Todd (1995) published error estimates for the Ease of Use indicator that average 18 percent, while their re-analysis of data from Adams, Nelson, and Todd (1992) yields average error estimates of 21 percent. Finally, the data we use here yields an average of 23 percent error variance in the full model where we can freely estimate it (Figure 1). While individual error estimates range between 14 percent and 36 percent, the use of a 20 percent average removes model and data set idiosyncrasies. Second, the error variance "quantifies your assessment of how similar or dissimilar your concept is to the best indicator" (Hayduk, 1996). A low error variance suggests that much of the variance in the indicator is caused by the construct, i.e., the indicator is a good reflection of the construct. We believe that a small error variance of 20 percent is defensible because our item wordings are strongly and directly related to the meaning of the latent constructs. Finally, we chose small error variances because, as Miles and Shevlin (2007) point out, the larger the error variances, the easier it is to fit the model. Intuitively, with large error variances, the structural and measurement models have relatively less indicator variance to explain than with smaller error variances.

9. Summary and Limitations

In this paper, we presented a proposal to recognize mis-fit of models in CB-SEM and to use the collected data for theory building. In summary, we have used the following steps:

Table 3. Summary of Analysis Steps		
Step		**Notes**
1	Test a priori model	This is a necessary unbiased test of the theory. However, it is unlikely that the initial model will fit well. The lack of fit provides the motivation for the following steps.
2	Examine residuals and modification indices	In a large model with dozens of manifest variables, it is unlikely that systematic residuals or patterns of residuals are obvious. These statistical tools will be of limited use.
3	Assess the measurement for each construct	For each construct, we examine residuals and modification indices, but focus on the meaning conveyed in the wording of the survey instrument. It is likely that survey items for a given construct have slightly different meaning, despite careful generation. These different meanings reflect different constructs, represented by additional latent variables.
4	Combine models of exogenous constructs into CFA model	It is likely that this model fits the data well, as there should be no further measurement problems, based on the analysis in step 3.
5	Combine models of exogenous and endogenous constructs into full structural model	If the a priori theoretical model is sound and complete, this model should fit the data well.

There are some limitations to this process. First, applying the proposal by Jöreskog (1993), we focus primarily on the measurement models for individual constructs, rather than the structural model. We do this because the individual models are easier to understand and diagnose, and the indicator wording is an important criterion in suggesting model modifications.

Second, the process we have presented will lead to incremental improvement but neglects the possibility of ground-breaking re-conceptualization of the phenomenon. For example, we retain the usual TAM constructs instead of introducing different constructs to explain the same data. To some degree, this is a result of our focus on face validity of the indicators with respect to the constructs. A given set of indicators necessarily limits the set of plausible constructs it can be argued to reflect.

Third, a significant $\chi 2$ test result may be caused by issues other than the model itself, e.g., violation of distributional assumptions. The default maximum-likelihood estimator, and consequently, the $\chi 2$ test statistic, requires multivariate normality, and departures from normality can result in a biased test statistic. Corrections, such as the Satorra-Bentler (1994) or Yuan-Bentler (2000) robust estimation methods or bootstrapping methods, are available but require the complete data set, not just the covariance matrix. For our example, Chin et al. (2008) do not report a test for multivariate normality, but their use of Maximum-Likelihood estimation without corrections suggests that this requirement is satisfied for this particular case.

Another issue affecting model fit may be causal heterogeneity, the situation where the sample is actually composed of sub-samples of different populations in which different causal processes operate. While techniques for establishing multi-group invariance are well known (e.g., Evermann, 2010), the identification of possible sub-samples within the sample relies heavily on the researcher's knowledge of the sample characteristics and theory base. In our example, we work only with the covariance matrix, so that we cannot identify any sub-samples, e.g., based on demographic descriptives. If sample characteristics are suspected to have an impact on the causal processes represented in the structural equation model (measurement and/or structural paths), Step 3 in Table 3 can be carried out separately for each suspected sub-sample. If the structure of the resulting models is identical ("configural equivalence" (Evermann, 2010)) the invariance of model parameters can be tested using a multi-group model (e.g., Evermann, 2010). The groups among which invariance can be established may be combined for step 4. A similar process for step 4 might combine some or all groups for step 5. Finally, a similar process of invariance testing might combine some or all groups after step 5. Just like the model modification should not be a mechanistic step driven by statistics, but instead be based on plausible causal explanations, so should the identification of sub-samples be

based on plausible causal explanations. In this way, different model structures or different parameter estimates between groups can point toward important boundary conditions for the theory, as represented by the variables used for sample stratification.

Finally, we note that the presented approach to theory building is intended to be based on an existing theoretical model. It is rare that exploratory survey data should be collected entirely without a guiding theory. Were such data to exist, theory building might begin by examining the item wordings of the survey instrument and identifying immediate causal constructs.

10. Discussion and Conclusion

The goal of this paper was to demonstrate that, in trying to fit a CB-SEM model using a strict test of model fit, we can generate theory. We believe that our final model in Figure 12 has achieved this. Through constant comparison of theory and data (Glaser & Strauss, 1967), we have identified five new theoretical constructs -- "perceived effectiveness," "perceived efficiency," "perceived clarity of interaction," "perceived system flexibility," and "perceived learnability"-- that are linked in a nomological theory network to the existing TAM constructs. These constructs and their relationships deserve, and enable, further theoretical attention and empirical investigation. They can also provide links to related fields of research, such as learning and educational theories ("perceived learnability") or human-computer interaction ("perceived clarity," "perceived flexibility"). Additionally, our model explains PEoU and PU as endogenous variables. The additional constructs are not only theoretically interesting, they are also practically important, as they provide more specific guidance to interventions to improve the usage of information systems. In Dubin's (1969) terms, we have built theory by seeking intervening variables, a process that starts with the "admission that the starting theoretical model is inadequate and must be supplemented" (p.81). Finally, even the "discarded" items EOU1 and USE5 are theoretically useful, because "the theory builder is often well advised to inquire about the data that researchers collect but subsequently exclude from their research analysis. These data may be mined for important insights about new units" (Dubin, 1969, p.84).

Related to our primary goal, we wish to emphasize the importance of instrument construction and the lack of attention paid to it. We believe that careless construction of an instrument, and neglect of interrelationships between subtle aspects of the survey items, are a frequent cause of problems with model fit. Our illustrative example showed that the basic TAM theory was supported and the model modifications are related to the indicators and driven by the item wordings.

Related to our pursuit of model fit, we have demonstrated that it is possible to fit realistically sized CB-SEM models with realistic sample sizes and our final model in Figure 12 has non-significant χ2 fit. We believe that the IS community needs to adopt stronger measures of model fit. In the Popperian model of science (1968) on which the hypothetic-deductive paradigm is based, we should be looking to falsify or reject our theories using the best and strongest tests available. It is unlikely that a, however well-conceived, a priori model fits well and can be tentatively accepted. However, it is unhelpful when journal editors and reviewers insist on authors presenting "the" model and showing that it is correct. In the face of such inflexible requirements, it is perhaps not surprising that authors grasp at every straw to argue that their model fits well, despite a significant χ2 statistic. While we agree that an a priori model should be tested without "peeking at the data," this is only sensible if, as a community, we are prepared to deal fairly with the outcome. As Hayduk, Cummings, et al. (2007, p.845) note, "We argue that attentively constructed and theoretically meaningful models that fail ought to be carefully discussed and published." We believe that the process of *post-hoc* model diagnostics shown here should, in a briefer form, be part of the discussion section of every paper. This discussion has a potentially important role to play in refining existing theory and building new theory. Our community is not well served if we stick our collective heads in the sand and simply ignore the fact that most of our models fail to fit the data we have collected.

10.1. The Epistemological Status of the Generated Model

As we introduced each new construct into the model, we emphasized an at least plausible reason for the relationships in our models. Post-hoc model modifications should be defensible on theoretical grounds, and should not be data-driven. We do not propose or condone "fishing expeditions" in the data, e.g., by blindly chasing modification indices. Model changes should at least be plausible, if not fully supported by existing theories. For example, an early study by Segars and Grover (1993) also suggested a separate effectiveness construct. However, their findings were rightly criticized by Chin and Todd (1995) as being based on a statistically driven specification search instead of theoretical considerations and the meaning of survey items, as we have done here. The process advocated in the present paper is clearly different from automatic, data-driven model specification approaches based purely on statistical grounds (MacCallum, 1986; Scheines et al., 1998). Instead, as we have argued earlier, theory building must be based on researcher's intuition and knowledge of theory and measures: "Thoughtful reconsideration of the relevant theory and data gathering procedures remain the best guide to model revision" (Hayduk, 1987, p.179).

It is important to note that "model modifications based on observed discrepancies might be capitalizing on chance sampling fluctuations in the data, improving fit at the expense of theoretical meaningfulness. … Researchers who engage in this exercise … must justify any modifications they make and, preferably, any resulting respecified models should be replicated with independent data" (Markland 2007, p.856). We believe that we have justified our model modifications as, at the very least, plausible, and have at times also referred to established literature. By focusing on item wording rather than statistical information, we believe that our model remains theoretically meaningful. Thus, while our result is intended to point the way for future research, it is only a tentative model and must still be replicated with independently collected data in a future study.

Besides independent verification in future studies or a split-sample approach (Hayduk, 1987; Jöreskog, 1993), another way to mitigate this issue is to use the proposed approach not on single-study data but on pooled data from multiple studies, i.e., for meta-analyses. Different meta-analytic techniques for CB-SEM exist, typically based on a pooled covariance or correlation matrix (Cheung & Chan, 2005). This avoids the problem of capitalizing on sample specific idiosyncrasies. On the other hand, multiple samples of different populations may lead to causal heterogeneity, the situation where the causal processes differ between sampled populations. Thus, care must be taken when selecting the studies for such a meta-analysis.

The cautions against over-fitting the model to idiosyncrasies in the data mirror those given in qualitative inductive theory building. For example, Eisenhardt (1989, p.547) cautions, "The risks are that the theory describes a very idiosyncratic phenomenon or that the theorist is unable to raise the level of generality of the theory." We do not believe that quantitative theory building as proposed here is in any way different from that of qualitative inductive theory building and that researchers need to be explicit about the limitations in either case. The epistemological status of our final model is similar to that of inductively generated theories. It is unsupported by evidence other than what it is generated from. It has, thus, the status of untested theory and must be tested against an independently collected sample. In this, its status is no different to a theory that is inductively developed from qualitative data (Eisenhardt, 1989; Eisenhardt & Graebner, 2007).

Finally, researchers need to take care to clearly separate the presentation and test of the a priori model from the discussion of post-hoc modifications. Hoyle and Panter (1995) recommend that "a straightforward means of distinguishing between predicted and discovered findings in reports of SEM results is to relegate presentation of the latter to a separate, clearly labeled section" (p.173). This section should contain the complete "history" of model modifications from the a priori model to the final model, similar to what we have presented here, but perhaps in briefer form.

10.2. Chasing Data or Chasing Theory?

While we do not propose or condone data-driven modification, we also note that dismissing post-hoc model modification as "fishing" in the data is an over-simplification. We believe that such

modifications admit two possible results. First, the resulting model may, in fact, be generalizable to different populations and samples and improves on existing theory while it maintains the same theoretical boundary conditions. Alternatively, the resulting model may not be generalizable to other populations but improves on existing theory by making it more specific within narrower boundary conditions. In either case, the researcher engaging in post-hoc model modification must clearly and explicitly discuss the boundaries within which he or she expects the model to be valid. If the boundaries are narrower, there needs to be an explicit discussion of how the specific sample characteristics affect the model. While we believe that our modified model remains valid in the same boundary conditions, one could also argue that characteristics of the sample chosen by Chin et al. (2008), undergraduate students in a computer course, somehow affect the model. For example, because computer courses train students to be analytic thinkers, the students may have recognized the differences between efficiency and effectiveness in a more pronounced way.

In the previous section, we briefly sketched an application of our proposal when a plausible case for heterogeneous sub-samples can be made. While it is difficult to identify heterogeneous sub-samples in a single study, (and impossible when working only with the covariance matrix as we do here), a meta-analytic approach that includes multiple studies would lend itself naturally to the identification of different models that hold for specific samples (and populations) and, therefore, the identification of boundary conditions of the theory.

10.3. The Parsimony of the Generated Model

The generated theory lacks the parsimony of the original TAM model. This lack of parsimony is an issue that has long been recognized in the qualitative theory building literature. As Eisenhardt (1989, p.847) notes, "There is a temptation to build theory which tries to capture everything. The result can be theory which is very rich in detail, but lacks the simplicity of overall perspective." In general, researchers and philosophers of science have recognized that there is a trade-off between parsimony, accuracy, and generality (Weick, 1979). This idea was first proposed by Thorngate (1976), who called it the "Impostulate of Theoretical Simplicity" and proposed that "it is impossible for an explanation of social behaviour to be simultaneously general, simple, and accurate" (p.126) and that "general, accurate theories of social behaviour must necessarily be complex" (p.127). Social scientists have noted that the quest for parsimoniousness is often the cause of a gap between models and the complexities of the real world. "[Although social scientists] strive for theories that are simultaneously parsimonious, highly general, and therefore applicable to a wide range of phenomena, yet precise enough to imply rejectable hypotheses, it does not appear possible…to achieve simultaneously all three of these ideal characteristics…my own position is that of the three, parsimony is the most expendable" (Blalock, 1982, p.28). The philosopher of science Mario Bunge writes in a book titled "The myth of simplicity" (1963) that "if practical advice is wanted as a corollary, let it be this: Occam's Razor, like all razors, should be handled with care to prevent beheading science in an attempt to shave off some of its pilosities. In science, as in the barber shop, better alive and bearded than cleanly shaven and dead" (p.115).

In CB-SEM analysis, this preference for a complex but well-fitting model has been expressed by Hayduk, Cummings, et al. (2007), who caution, "If the model χ2 test detects a causally mis-specified model, the biased estimates … become impotent and unconvincing" (p.845). In general, the parameter estimates from a mis-specified model should not be substantively interpreted. Estimates are only correct for the true model, and model fit statistics give no indication of the bias in the estimates. For example, we do not know whether the TAM structural coefficient estimates are correct if the model does not fit. Having fitted a model, albeit a more complex one, we can compare the coefficient estimates to those of the simpler model. While they are similar for the PEoU to IU path (0.2 in Figures 10 and 12 versus 0.22 reported by Chin et al. (2008)), they differ for the PU to IU path (0.8 in Figures 10 and 12 versus 0.62 reported by Chin et al. (2008)).

10.4. Measurement Validity and Reliability

The additional constructs in the generated model necessarily result in fewer indicators for each construct; in the extreme case, only a single indicator is left. Thus, reliability metrics such as Cronbach's alpha (Cortina, 1993) or internal consistency (Raykov, 1997) are not applicable, and the validity of the indicators is established through the overall fit of the model. A valid indicator is one that is properly causally specified. A well-fitting model supports the conclusion that all regression paths represent causal paths. Hence, the indicators can be assumed to be valid (Hayduk, Pazderka-Robinson, et al., 2007). In addition to the argument by model fit, we believe that the indicators in the final model show increased face validity. The indicators are related to constructs in a direct and obvious way, as the new constructs take their meaning directly from the indicators on which they are based. For example, flexibility is assessed in our final model with just one indicator, EOU4 ("I find the (task-related) portions of (system) to be flexible to interact with"). Hence, the flexibility construct could be defined as the flexibility of the system interaction when performing a task.

We note that this is not an operationist definition whereby the meaning of the concept is for all times defined by its single indicator (Grace, 2001; Feest, 2005). Instead, future studies may wish to devise additional indicators, and ideally different measurement techniques altogether (for example, Webb, Campbell, Schwartz, Sechrest, & Grove, 1981) for such constructs in order to allow for traditional estimates of reliability and validity. As Little, Lindenberger, and Nesselroade (1999) illustrate, multiple indicators also improve the efficiency of estimating the constructs. Examples of additional item wordings for perceived flexibility, similar to EOU4, may be: "The (system) provides different ways to achieve (task)" or "The (system) allows me to choose my preferred way to perform (task) from among multiple options." When developing additional items, great care must be taken to ensure that all items are identical in meaning, i.e., they represent different ways of asking the same questions. This is the basis of traditional validity and reliability assessment and avoids a repeat of the process outlined here, where model fit issues were centered on divergent meaning of survey items.

10.5. Recommendations

We close with some recommendations for researchers, reviewers, and editors. For researchers, we believe that the IS research field needs to pay increased attention to model fit and not disregard important evidence that our models are a poor fit to data. For reviewers and editors, we believe that there is value in publishing and discussing models that do not fit, if they are theoretically well motivated, with sensible indicators and data that has been carefully collected. However, we should insist that authors honestly discuss and diagnose model fit issues. Clinging to ill-fitting factor models and using data only once to test a single set of hypotheses are not helpful in driving theory in our field forward.

Our recommendations to acknowledge failing models, understand why they fail, and use that understanding constructively echo a recent call for the "pursuit of failure" in a related discipline. In the introduction, we briefly showed that the problems with model fit are not unique to the IS field. Similarly, a call for more rigorous theory testing has recently been voiced in organizational sciences where Gray and Cooper (2010) suggest that "our field [organizational science] seems to privilege corroborating over disconfirming evidence" (p.622) and argue that "organizational studies … would benefit from a stronger focus on theory development via the pursuit of failure" (p.621).

In conclusion, we believe the data has much more to offer than a test of a single model, model test results should not be dismissed and the data deserve to be explained. It is incumbent upon us not to waste the enormous effort that we spend on theorizing and data collection.

References

Adams, D. A., Nelson, R. R. & Todd, P. A. (1992). Perceived usefulness, ease of use and usage of information technology: A replication. *MIS Quarterly*, *16*(2), 227-247.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *The Academy of Management Review*, 14(4), 496–515.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5), 815–824.

Bentler, P. M. (2007). On tests and indices for evaluating structural equation models. *Personality and Individual Differences*, *42*(5), 825-829.

Blalock, H. (1982). *Conceptualization and measurement in the social sciences*. Beverly Hills, CA: SAGE Publications.

Bollen, K. (1989) *Structural equations with latent variables*. New York, NY: John Wiley and Sons.

Bollen, K. (1990). A comment on model evaluation and modification. *Multivariate Behavioral Research*, *25*(2), 181-185.

Bunge, M. A. (1962). *Intuition and Science*. Englewood Cliffs, NJ: Prentice-Hall.

Bunge, M. A. (1963). *The myth of simplicity: Problems of scientific philosophy*. Englewood Cliffs, NJ: Prentice-Hall.

Chen, W. S. & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal*, *14*(3), 197-235.

Cheung, M. W. L. & Chan, W. (2005). Meta-analytic structural equation modeling – A two-stage approach. *Psychological Methods, 10*(1), 40-64.

Chin, W. W, Johnson, N., & Schwarz, A. (2008). A fast form approach to measuring technology acceptance and other constructs. *MIS Quarterly, 32*(4), 687–703.

Chin, W. W. & Todd, P. A. (1995). On the use, ease of use, and usefulness of structural equation modeling in MIS research: A note of caution. *MIS Quarterly, 19*(2), 237-246

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Curran, P. J., Bollen, K. A., Paxton, P., & Kirby, J. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte-Carlo study. *Multivariate Behavioral Research, 37*(1), 1-36.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.

Dubin, R. (1969). *Theory Building*. New York, NY: Free Press.

Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, *14*(4), 532–550.

Eisenhardt, K. M. & Graebner, M.E. (2007) Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, (50)1, 25-32.

Evermann, J. (2010). Multiple-group analysis using the sem package in the R system. *Structural Equation Modeling*, *17*(4), 677-702.

Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences*, *41*(2), 131-149.

Fox, J. (2006). Structural equation modeling with the sem package in R. *Structural Equation Modeling*, 13(3), 465–486.

Gefen, D., Straub, D. W., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for research practice. *Communications of the AIS*, *4*(7), 1–70.

Glaser, B. G. & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, NY: Aldine Publishing Company.

Goffin, R. D. (2007). Assessing the adequacy of structural equation models: Golden rules and editorial policies. *Personality and Individual Differences*, *42*(5), 831–839.

Grace, R. (2001). On the failure of operationism. *Theory and Psychology*, *11*(1), 5-33.

Gray, P. H. & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods*, 13(4), 620-643.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611–642.

Hayduk, L. A. (1987). *Structural Equation Modeling with LISREL – Essentials and Advances*. Baltimore, MD: Johns Hopkins University Press.

Hayduk, L. A. (1990). Should model modifications be oriented toward improving data fit or encouraging creative and analytical thinking? *Multivariate Behavioral Research*, *25*(2), 193-196.

Hayduk, L. A. (1996). *LISREL Issues, Debates, and Strategies*. Baltimore, MD: Johns Hopkins University Press.

Hayduk, L. A., Cummings, G. C., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! One, two, three - testing the theory in structural equation models! *Personality and Individual Differences*, *42*(5), 841–850.

Hayduk, L. A., Pazderka-Robinson, H., Cummings, G. C., Boadu, K., Verbeek, E. L. & Perks, T. A. (2007). The weird world, and equally weird measurement models: Reactive indicators and the validity revolution. *Structural Equation Modeling*, *14*(2), 280-310.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling – concepts, issues, and applications* (pp.158-176). Thousand Oaks, CA: Sage Publications.

Im, C. & Wang, J. (2007). A TETRAD-based approach for theory development in information systems research. *Communications of the Association for Information Systems*, *20*(1), 322-345.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Ed.), *Testing Structural Equation Models* (pp.294-316). Newbury Park, CA: Sage Publications.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendations. *Multivariate Behavioral Research*, 25(2), 137-155.

Lee, B., Barua, A. & Whinston, A. B. (1997). Discovery and representation of causal relationships in MIS research: A methodological framework. *MIS Quarterly*, *21*(1), 109-136.

Little, T. D., Lindenberger, U. & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators Are bad and "bad" indicators are good. *Psychological Methods*, *4*(2), 192-211.

Liu, L. (2009). Technology acceptance model: A replicated test using TETRAD. *International Journal of Intelligent Systems*, *24*(12), 1230-1242.

Lyytinen, K. (2009). Data matters in IS theory building. *Journal of the Association for Information Systems*, *10*(10), 715-720.

Maccallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*(1), 107-120.

Maccallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149.

Marcoulides, G. A. & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling*, *10*(1), 154–164.

Marcoulides, G. A., Drezner, Z. & Schumacker, R. E. (1998). Model specification searches in structural equation models using Tabu search. *Structural Equation Modeling*, *5*(4), 365-376.

Markland, D. (2007). The golden rule is that there are no golden rules: A commentary on Paul Barett's recommendations for reporting model fit in structural equation modeling. *Personality and Individual Differences*, *42*(5), 851–858.

Mcintosh, C. N. (2007). Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences*, *42*(5), 859–867.

Medawar, P. B. (1969). *Induction and Intuition in Scientific Thought*. Philadelphia, PA: American Philosophical Society.

Miles, J. & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences*, *42*(5), 869–874.

Muthén, L. K. & Muthén, B. O. (2002). How to use a Monte-Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620.

Palvia, P., Leary, D., Mao, E., Midha, V., Pinjani, P. & Salam, A. F. (2004). Research methodologies in MIS: An update. *Communications of the AIS*, *14*, 526-542.

Popper, K. (1968). *The Logic of Scientific Discovery*. London, UK: Hutchinson & Co.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173-184.

Root-Bernstein, R. S. (1989). *Discovering*. Cambridge, MA: Harvard University Press.

Saris, W. E. & Sotorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Ed.) *Testing Structural Equation Models* (pp.181-204). Newbury Park, CA: Sage Publications.

Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, A. & C. C. Clogg (Ed.) *Latent variables analysis – Applications for developmental research* (pp.399-419). Thousand Oaks, CA: Sage Publications.

Scheines, R., Spirtes, P., Glymour, C., Meek C. & Richardson, T. (1998). The TETRAD project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, *33*(1), 65-117.

Schumacker, R. E. (2006). Conducting specification searches with Amos. *Structural Equation Modeling*, *13*(1), 118–129.

Segars, A. H. & Grover, V. (1993). Re-examining perceived ease-of-use and usefulness: A confirmatory factor analysis. *MIS Quarterly*, *17*(4), 517-525.

Shah, S. K. & Corley, K. G. (2006). Building better theory by bridging the quantitative-qualitative divide. *Journal of Management Studies*, *43*(8), 1821-1835.

Son, J. Y., Kim, S. S. & Riggins, F. J. (2006). Consumer adoption of net-enabled infomediaries: Theoretical explanations and an empirical test. *Journal of the Association for Information Systems*, *7*(7), 473-508.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–898.

Thorngate, W. (1976). Possible Limits on a Science of Social Behavior. In L. H. Strickland, F. E. Aboud & K. J. Gergen (Ed.) *Social Psychology in Transition* (pp.121-139). New York, NY: Plenum Press.

Ting, K. (1998). The TETRAD approach to model respecification. *Multivariate Behavioral Research*, *33*(1), 157-164.

Urquhart, C., Lehmann, H. & Myers, M. D. (2010). Putting the 'theory' back in grounded theory: Guidelines for grounded theory studies in information systems. *Information Systems Journal*, *20*(4), 357-381.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L. & Grove, J. B. (1981). *Nonreactive Measures in the Social Sciences* (2nd ed.). Boston, MA: Houghton Mifflin Company.

Weick, K. E. (1979). *The Social Psychology of Organizing* (2nd ed.). Reading, MA: Addison-Wesley Publishing Company.

Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*(4), 516-531.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60-62.

Yuan, K. H. & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*(1), 165-200.

Appendix. Model Specifications and Estimation Results

In the sem package in R, models are specified by one line for each path. Regression paths are specified with -> while covariance paths are specified by <->. The second entry on each line is the name of the parameter or NA if the parameter is fixed. The third entry is the value of a fixed parameter or the starting value for a free parameter. Covariance paths for observed variables represent error covariances; covariance paths for endogenous latents represent disturbance covariances. Comments begin with a #. For further details see (Fox, 2006).

Model Specification and Results for Figure 3

```
# The EOU construct
eou -> eou2, NA, 1
eou -> eou6, lambda1, 1
eou <-> eou, phi1, NA
eou2 <-> eou2, NA, 0.36
eou6 <-> eou6, NA, 0.4
```

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
Lambda1	1.0479	0.046374	22.5972	0	eou6 <--- eou
phi1	1.4793	0.154257	9.5896	0	eou <--> eou

Model Specification and Results for Figure 4

```
# The EOU part
eou -> eou2, NA, 1
eou -> eou6, lambda1, 1.05
eou <-> eou, zeta1, NA
eou2 <-> eou2, NA, 0.36
eou6 <-> eou6, NA, 0.40

# Perceived Clarity
clear -> eou3, NA, 1
eou3 <-> eou3, NA, 0.34
clear <-> clear, phi2, NA

clear -> eou, gamma1, NA

# Perceived Flexibility
flex -> eou4, NA, 1
eou4 <-> eou4, NA, 0.32
flex <-> flex, phi3, NA
flex -> eou, gamma2, NA

# Covariances
flex <-> clear, phi1,NA
```

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda1	1.068587	0.046081	23.1892	0.0000e+00	eou6 <--- eou
zeta1	-0.049852	0.048739	-1.0228	3.0638e-01	eou <--> eou
phi2	1.420000	0.148683	9.5505	0.0000e+00	clear <--> clear
gamma1	1.144443	0.205550	5.5677	2.5809e-08	eou <--- clear
phi3	1.328000	0.138944	9.5578	0.0000e+00	flex <--> flex
gamma2	-0.133897	0.207013	-0.6468	5.1776e-01	eou <--- flex
phi1	1.251000	0.125622	9.9585	0.0000e+00	clear <--> flex

Model Specification and Results for Figure 5

```
# The EOU construct
eou -> eou2, NA, 1
eou -> eou6, lambda1, 1.05                 # Covariances
eou <-> eou, zeta1, NA                      flex <-> clear, phi1, NA
eou2 <-> eou2, NA, 0.36
eou6 <-> eou6, NA, 0.40                     # Learnability
                                            learn <-> learn, zeta2, NA
# Perceived clarity                         learn -> eou1, NA, 1
clear -> eou3, NA, 1                         learn -> eou5, lambda5, NA
eou3 <-> eou3, NA, 0.34                      eou1 <-> eou1, NA, 0.42
clear <-> clear, phi2, NA                    eou5 <-> eou5, NA, 0.4
clear -> eou, gamma1, NA
                                            # Causal relationships
# Perceived flexibility                     flex -> learn, gamma3, NA
flex -> eou4, NA, 1                          clear -> learn, gamma4, NA
eou4 <-> eou4, NA, 0.32
flex <-> flex, phi3, NA
flex -> eou, gamma2, NA
```

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda1	1.0709785	0.045459	23.55942	0.0000e+00	eou6 <--- eou
zeta1	-0.0354484	0.025560	-1.38687	1.6548e-01	eou <--> eou
phi2	1.4283542	0.147325	9.69527	0.0000e+00	clear <--> clear
gamma1	1.0916229	0.127883	8.53609	0.0000e+00	eou <--- clear
phi3	1.3281374	0.138984	9.55601	0.0000e+00	flex <--> flex
gamma2	-0.0838928	0.134661	-0.62299	5.3329e-01	eou <--- flex
phi1	1.2497466	0.125631	9.94774	0.0000e+00	clear <--> flex
zeta2	0.0097682	0.041329	0.23635	8.1316e-01	learn <--> learn
lambda5	0.9485977	0.041987	22.59262	0.0000e+00	eou5 <--- learn
gamma3	-0.3288684	0.185190	-1.77585	7.5758e-02	learn <--- flex
gamma4	1.3584442	0.175915	7.72216	1.1546e-14	learn <--- clear

Model Specification and Results for Figure 6

```
use -> use1, NA, 1                          use1 <-> use1, delta1, NA
use -> use2, lambda2, 1                     use2 <-> use2, delta2, NA
use -> use3, lambda3, 1                     use3 <-> use3, delta3, NA
use -> use4, lambda4, 1                     use4 <-> use4, delta4, NA
use -> use5, lambda5, 1                     use5 <-> use5, delta5, NA
use -> use6, lambda6, 1                     use6 <-> use6, delta6, NA
use <-> use, phi1, NA
```

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda2	1.08573	0.074727	14.5293	0.0000e+00	use2 <--- use
lambda3	1.20927	0.081221	14.8887	0.0000e+00	use3 <--- use
lambda4	1.19246	0.078119	15.2647	0.0000e+00	use4 <--- use
lambda5	1.23345	0.089563	13.7718	0.0000e+00	use5 <--- use
lambda6	1.32992	0.086171	15.4334	0.0000e+00	use6 <--- use
phi1	0.60581	0.080701	7.5069	6.0618e-14	use <--> use
delta1	0.40919	0.039626	10.3262	0.0000e+00	use1 <--> use1
delta2	0.39187	0.039016	10.0437	0.0000e+00	use2 <--> use2
delta3	0.42510	0.043382	9.7990	0.0000e+00	use3 <--> use3
delta4	0.35057	0.037084	9.4534	0.0000e+00	use4 <--> use4
delta5	0.58132	0.056803	10.2339	0.0000e+00	use5 <--> use5
delta6	0.36451	0.041210	8.8450	0.0000e+00	use6 <--> use6

Model Specification and Results for Figure 7

Perceived efficiency
efficient -> use1, NA, 1
efficient -> use2, lambda2, 1
efficient -> use3, lambda3, 1   # Perceived usefulness
efficient <-> efficient, phi1, NA   useful -> use6, NA, 1
use1 <-> use1, delta1, NA   useful <-> useful, zeta1, 1
use2 <-> use2, delta2, NA   use6 <-> use6, NA, 0.30
use3 <-> use3, delta3, NA

               # Causal relationships
Perceived effectiveness   efficient -> useful, gamma1, NA
effective -> use4, NA, 1   effective -> useful, gamma2, NA
use4 <-> use4, NA, 0.24
effective <-> effective, phi2, NA
effective <-> efficient, phi3, 1

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda2	1.07423	0.071373	15.0511	0.0000e+00	use2 <--- efficient
lambda3	1.20188	0.077912	15.4261	1.1546e-14	use3 <--- efficient
phi2	0.63996	0.082885	7.7211	0.0000e+00	efficient <--> efficient
delta1	0.37504	0.039071	9.5988	0.0000e+00	use1 <--> use1
delta2	0.36750	0.039643	9.2701	0.0000e+00	use2 <--> use2
delta3	0.38656	0.043796	8.8264	0.0000e+00	use3 <--> use3
phi2	0.97200	0.102150	9.5155	0.0000e+00	effective <--> effective
phi3	0.72728	0.077438	9.3918	0.0000e+00	efficient <--> effective
zeta1	0.13732	0.045717	3.0037	2.6670e-03	useful <--> useful
gamma1	0.78095	0.273660	2.8537	4.3210e-03	useful <--- efficient
gamma2	0.39921	0.220848	1.8076	7.0667e-02	useful <--- effective

Model Specification and Results for Figure 8

The EOU construct
eou -> eou2, NA, 1
eou -> eou6, lambda1, 1
eou <-> eou, zeta1, NA   # Perceived effectiveness
eou2 <-> eou2, NA, 0.36   effective -> use4, NA, 1
eou6 <-> eou6, NA, 0.40   use4 <-> use4, NA, 0.24
               effective <-> effective, phi5, NA
# Perceived clarity   effective <-> efficient, phi6, 1
clear -> eou3, NA, 1
eou3 <-> eou3, NA, 0.34   # Perceived usefulness
clear <-> clear, phi2, NA   useful -> use6, NA, 1
clear -> eou, gamma1, NA   useful <-> useful, zeta2, 1
               use6 <-> use6, NA, 0.30
# Perceived flexibility
flex -> eou4, NA, 1   # Causal relationships
eou4 <-> eou4, NA, 0.32, NA   efficient -> useful, gamma3, NA
flex <-> flex, phi3, NA   effective -> useful, gamma4, NA
flex -> eou, gamma2, NA   eou -> useful, gamma5, NA
flex <-> clear, phi1, NA
               # Exogenous covariances
Perceived efficiency   flex <-> efficient, phi7, NA
efficient -> use1, NA, 1   flex <-> effective, phi8, NA
efficient -> use2, lambda2, 1   clear <-> efficient, phi9, NA
efficient -> use3, lambda3, 1   clear <-> effective, phi10, NA
efficient <-> efficient, phi4, NA
use1 <-> use1, delta1, NA
use2 <-> use2, delta2, NA
use3 <-> use3, delta3, NA

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda1	1.066844	0.045994	23.19538	0.0000e+00	eou6 <--- eou
zeta1	-0.047657	0.049361	-0.96548	3.3431e-01	eou <--- eou
phi2	1.420703	0.148752	9.55080	0.0000e+00	clear <--> clear
gamma1	1.166287	0.206443	5.64943	1.6098e-08	eou <--- clear
phi3	1.327058	0.138964	9.54962	0.0000e+00	flex <--> flex
gamma2	-0.158570	0.207667	-0.76358	4.4512e-01	eou <--- flex
phi1	1.254636	0.125566	9.99181	0.0000e+00	clear <--> flex
lambda2	1.073017	0.071159	15.07922	0.0000e+00	use2 <--- efficient
lambda3	1.198657	0.077625	15.44154	0.0000e+00	use3 <--- efficient
phi4	0.640068	0.082960	7.71541	1.1990e-14	efficient <--> efficient
delta1	0.374932	0.039052	9.60075	0.0000e+00	use1 <--> use1
delta2	0.369047	0.039516	9.33912	0.0000e+00	use2 <--> use2
delta3	0.391364	0.043708	8.95405	0.0000e+00	use3 <--> use3
phi5	0.974642	0.102297	9.52758	0.0000e+00	effective <--> effective
phi6	0.729410	0.077559	9.40463	0.0000e+00	efficient <--> effective
zeta2	0.133213	0.045175	2.94881	3.1900e-03	useful <--> useful
gamma3	0.834740	0.292963	2.84930	4.3816e-03	useful <--- efficient
gamma4	0.309456	0.227977	1.35740	1.7465e-01	useful <--- effective
gamma5	0.067931	0.045673	1.48735	1.3692e-01	useful <--- eou
phi7	0.511513	0.075680	6.75889	1.3905e-11	efficient <--> flex
phi8	0.637746	0.092410	6.90123	5.1552e-12	effective <--> flex
phi9	0.495305	0.074123	6.68225	2.3531e-11	efficient <--> clear
phi10	0.562385	0.089211	6.30402	2.9003e-10	effective <--> clear

Model Specification and Results for Figure 9

lu -> lu1, NA, 1   lu1 <-> lu1, delta1, 0.40
lu -> lu2, lambda1, 1   lu2 <-> lu2, delta2, 0.36
lu -> lu4, lambda2, 1   lu4 <-> lu4, delta3, 0.42
lu -> lu3, lambda3, 1   lu3 <-> lu3, delta4, 0.40
lu <-> lu, phi8, NA

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda1	0.98939	0.042050	23.5289	0.0000e+00	lu2 <--- lu
lambda2	1.06626	0.043979	24.2449	0.0000e+00	lu4 <--- lu
lambda3	0.99786	0.042552	23.4502	0.0000e+00	lu3 <--- lu
phi8	1.58006	0.165904	9.5240	0.0000e+00	lu <--> lu
delta1	0.40493	0.043440	9.3216	0.0000e+00	lu1 <--> lu1
delta2	0.32828	0.037447	8.7666	0.0000e+00	lu2 <--> lu2
delta3	0.31460	0.038691	8.1311	4.4409e-16	lu4 <--> lu4
delta4	0.32969	0.037618	8.7640	0.0000e+00	lu3 <--> lu3

Model Specification and Results for Figure 10

```
# The EOU construct
eou -> eou2, NA, 1
eou -> eou6, lambda1, 1
eou <-> eou, zeta1, NA
eou2 <-> eou2, deltaeou2, 0.36
eou6 <-> eou6, deltaeou6, 0.40
# Perceived clarity
clear -> eou3, NA, 1
eou3 <-> eou3, NA, 0.34
clear <-> clear, phi2, NA
clear -> eou, gamma1, NA
# Perceived flexibility
flex -> eou4, NA, 1
eou4 <-> eou4, NA, 0.32, NA
flex <-> flex, phi3, NA
flex -> eou, gamma2, NA
flex <-> clear, phi1, NA

# Perceived efficiency
efficient -> use1, NA, 1
efficient -> use2, lambda2, 1
efficient -> use3, lambda3, 1
efficient <-> efficient, phi4, NA
use1 <-> use1, delta1, NA
use2 <-> use2, delta2, NA
use3 <-> use3, delta3, NA

# Perceived effectiveness
effective -> use4, NA, 1
use4 <-> use4, NA, 0.24
effective <-> effective, phi5, NA
effective <-> efficient, phi6, 1
```

```
# Perceived usefulness
useful -> use6, NA, 1
useful <-> useful, zeta2, 1
use6 <-> use6, deltause6, 0.30

# Casual relationships
efficient -> useful, gamma3, NA
effective -> useful, gamma4, NA
eou -> useful, gamma5, NA

# Exogenous covariances
flex <-> efficient, phi7, NA
flex <-> effective, phi8, NA
clear <-> efficient, phi9, NA
clear <-> effective, phi10, NA

# Causal relationships
eou -> lu, gamma6, NA
useful -> lu, gamma7, NA

# Predicted usage construct
lu -> lu1, NA, 1
lu -> lu2, lambda7, 1
lu -> lu4, lambda8, 1
lu -> lu3, lambda9, 1
lu <-> lu, phi50, NA
lu1 <-> lu1, deltalu1, 0.40
lu2 <-> lu2, deltalu2, 0.36
lu4 <-> lu4, deltalu3, 0.42
lu3 <-> lu3, deltalu4, 0.40
```

Parameter Estimates

	Estimate	Std Error	z value	Pr(>|z|)	
lambda1	1.089881	0.049595	21.97582	0.0000e+00	eou6 <--- eou
zeta1	-0.062519	0.050980	-1.22635	2.2007e-01	eou <--- eou
deltaeou2	0.431831	0.047618	9.06867	0.0000e+00	eou2 <--> eou2
deltaeou6	0.352191	0.046959	7.49999	6.3727e-14	eou6 <--> eou6
phi2	1.419199	0.148732	9.54196	0.0000e+00	clear <--> clear
gamma1	1.179635	0.212421	5.55329	2.8034e-08	eou <--- clear
phi3	1.326353	0.138971	9.54409	0.0000e+00	flex <--> flex
gamma2	-0.183003	0.213274	-0.85806	3.9086e-01	eou <--- flex
phi1	1.256735	0.125539	10.01071	0.0000e+00	clear <--> flex
lambda2	1.080303	0.071361	15.13857	0.0000e+00	use2 <--- efficient
lambda3	1.204050	0.077735	15.48920	0.0000e+00	use3 <--- efficient
phi4	0.635372	0.082601	7.69210	1.4433e-14	efficient <--> efficient
delta1	0.379627	0.038907	9.75736	0.0000e+00	use1 <--> use1
delta2	0.364484	0.038589	9.44536	0.0000e+00	use2 <--> use2
delta3	0.389874	0.042826	9.10378	0.0000e+00	use3 <--> use3
phi5	0.973554	0.102237	9.52248	0.0000e+00	effective <--> effective
phi6	0.726194	0.077345	9.38902	0.0000e+00	efficient <--> effective
zeta2	0.016422	0.050580	0.32468	7.4543e-01	useful <--> useful
deltause6	0.415821	0.060199	6.90744	4.9349e-12	use6 <--> use6
gamma3	0.887276	0.259145	3.42386	6.1738e-04	useful <--- efficient
gamma4	0.271192	0.199119	1.36196	1.7321e-01	useful <--- effective
gamma5	0.065920	0.046114	1.42949	1.5286e-01	useful <--- eou
phi7	0.512962	0.075466	6.79721	1.0666e-11	efficient <--> flex
phi8	0.641682	0.092317	6.95086	3.6307e-12	effective <--> flex

phi9	0.489540	0.073684	6.64374	3.0582e-11	efficient <--> clear
phi10	0.557394	0.089074	6.25767	3.9076e-10	effective <--> clear
gamma6	0.205960	0.065753	3.13231	1.7344e-03	lu <--- eou
gamma7	0.795834	0.081594	9.75359	0.0000e+00	lu <--- useful
lambda7	0.986071	0.041914	23.52587	0.0000e+00	lu2 <--- lu
lambda8	1.064321	0.043661	24.37685	0.0000e+00	lu4 <--- lu
lambda9	1.000093	0.042114	23.74735	0.0000e+00	lu3 <--- lu
phi50	0.655979	0.078837	8.32066	0.0000e+00	lu <--> lu
deltalu1	0.402543	0.042545	9.46161	0.0000e+00	lu1 <--> lu1
deltalu2	0.336317	0.037043	9.07917	0.0000e+00	lu2 <--> lu2
deltalu3	0.318425	0.037697	8.44689	0.0000e+00	lu4 <--> lu4
deltalu4	0.320249	0.036215	8.84303	0.0000e+00	lu3 <--> lu3

About the Authors

Joerg EVERMANN is an Associate Professor of Information Systems with the Faculty of Business Administration at Memorial University of Newfoundland, Canada. He received his PhD from the University of British Columbia. Prior to coming to Memorial University, he was a lecturer at Victoria University of Wellington, New Zealand. Joerg's current interests are in empirical aspects of data integration and in research methodology. His work has been published in high-quality international journals such as *Information Systems Journal, Information Systems, IEEE Transactions on Software Engineering,* and *IEEE Transactions on Knowledge and Data Engineering*. His work on research methodology is published in *Structural Equation Modeling - An Interdisciplinary Journal, the Journal of the AIS*, and in the proceedings of the International Conference on Information Systems.

Mary TATE is a Senior Lecturer in Information Systems at Victoria University of Wellington, New Zealand. She has many years IT industry experience, including Service Delivery Manager for Victoria University's IT department, and managing the Internet channel for a major bank, before taking up an academic position in 2001. Mary has more than 40 peer-reviewed publications including journal articles, book chapters and conference presentations; and has won awards for research and teaching. Mary has a strong interest in research methods, and has been involved in research methods track at ICIS and other leading conferences and theory-building workshops.