

WHEN IMPRECISE STATISTICAL STATEMENTS BECOME PROBLEMATIC: A RESPONSE TO GOODHUE, LEWIS, AND THOMPSON¹

George A. Marcoulides

Graduate School of Education, University of California, Riverside, Sproul Hall 1207,
Riverside, CA 92521 U.S.A. {george.marcoulides@ucr.edu}

Wynne W. Chin

Department of Decision and Information Systems, C. T. Bauer College of Business, University of Houston,
Houston TX 77204-6021 U.S.A. {wchin@uh.edu} and
Sogang University, 2 Shinsu-dong, Mapo-gu, Seoul, 121-742 KOREA

Carol Saunders

Department of Management, University of Central Florida, Orlando, FL 32816-1400 U.S.A. {csaunders@bus.ucf.edu}

We wish to thank the editors of MISQ for inviting this commentary in response to the Issues and Opinions piece by Goodhue, Lewis, and Thompson (hereinafter GLT). We must admit we were somewhat surprised when we received a lengthy Issues and Opinions paper that responded to our slightly over four page (published) Foreword to the 2009 *MIS Quarterly* Special Issue on PLS (Marcoulides et al. 2009). Our short Foreword was merely designed to provide an overview of the special issue and alert the *MISQ* audience to the importance of making correct comparisons between PLS and other statistical modeling techniques.

We were even more surprised that the Issues and Opinions paper displays imprecise statements and attenuates concerns about questionable comparisons of PLS with other statistical techniques. This matter is particularly disconcerting when one reflects on the statements provided by Professor Schneeweiss (see Appendix B3 in Goodhue et al. 2012a) when asked by GLT whether PLS and other techniques can be legitimately compared. Our interpretation differs from that of GLT in that Professor Schneeweiss' statements seem rather to caution

GLT that it all depends on what is meant by comparing the different methods and that comparing different results arising from different entities does not make much sense, since the methods are to begin with dissimilar. It seems that GLT failed to recognize that their recommended procedure is akin to comparing apples to oranges and we therefore caution readers about using their paper as the basis or justification for future Monte Carlo comparison studies.

Often applied researchers do not pay sufficient attention to the stochastic assumptions underlying particular statistical models. Such lack of attention to espoused statistical theory can also divert focus from precise statistical statements, analyses, and applications, by purporting to do what cannot be legitimately done with the particular recommended approach and even encouraging others to engage in making comparisons that offer little value. In many ways, this lack of attention and precision is reminiscent of Karl A. Fox's (1980, p. 33) argument that "between substantive research and statistical theory there is a long distance, and even some hostility."

As was the case in the Goodhue, Lewis, Thompson (2006) HICSS article and in Goodhue, Lewis and Thompson (2012b), we believe that their suggested model comparisons

¹Ron Cenfetelli was the accepting senior editor for this paper. Geneviève Bassellier served as the associate editor.

are simply not correct (which, as we elaborate in later sections, are due to incorrect parameterizations). In what follows, we address some of our main concerns with the GLT piece, primarily because modeling is currently enjoying such widespread popularity and it would be irresponsible on our part not to voice our serious concerns about an approach that could become part of the structural equation modeling (SEM) literature (we note the term SEM is used here generically and interchangeably to refer to path analytic models with latent variables, covariance-based models, or simply latent variable models).²

While our comments here may strike some readers as critical, our concerns are genuine and pertinent to promote the development and enrichment of the concepts and practices that emerge from any empirical research using such modeling techniques. Unfortunately, statistical issues and assumptions that may appear incidental to an applied researcher's substantive ideas can and often do become stumbling blocks that invalidate their models. We hope to at least alert such users to the right course of action in this specific modeling research paradigm.

We begin our response with a discussion of some key definitions and what we mean by the terms *parameterization* and *correct parameterization*.³ Then, we collectively address the five interrelated issues identified by GLT in the main body of their Issues and Opinions piece and elaborated on further in Appendix A. In this section we show how imprecise statements in GLT's piece can lead to problems, statistically speaking. Given that the title of the GLT paper indicates that it is a response to our Foreword, we believe we should primarily focus our discussion viz our original Foreword. Nonetheless we feel compelled to point out our concerns regarding the assumptions and proposed process presented by GLT. Along the way we also offer some guidance to resources that are readily available in the literature outlining how one can conduct legitimate comparisons and how to

indicate limitations in making comparisons. Our intent in this response is to share with you our **major** concerns with the GLT piece. It is not our intent, however, to provide a lengthy tutorial on how to conduct legitimate comparisons among statistical techniques, how to specify the correct parameterization of models, or to detail all inaccuracies present in the GLT piece.

Accuracy and Precision: The Five Issues

Issue 1: Incorrect Parameterization

In order to ensure precision of notation and definitions, we begin by offering some basic modeling terminology. It is important to distinguish between the definitions of a population parameter and parameterization. A parameter for a specific population of interest is a quantity or statistical measure that, for a given population, is fixed and that is used as the value of a variable in some general distribution or frequency function to make it descriptive of that population. Thus parameters are population quantities (e.g., like the mean or variance), which characterize a population distribution on a variable of interest. In other words, a population parameter can be viewed as a numerical summary of the population.

On the other hand, parameterization is merely the specification of the parameters of a model. The term is very commonly used in the SEM literature. For example, the commercially available program Mplus (Muthén and Muthén 2010) uses the command "PARAMETERIZATION = " to enable researchers to specify a variety of models according to a particular model structure. In a particular confirmatory factor analytic model when a "delta" parameterization is used, scale factors are allowed to be parameters in the model but residual variances are not, whereas when a "theta" parameterization is used, residual variances are allowed to be parameters in the model but scale factors are not. None of the definitions offered by GLT in their Issues and Opinions article deal with the specification of the parameters of a model. Hence, we consider them inappropriate and recognize that we need to elaborate here on the definition of parameterization, as well as offer some additional details on basic modeling terminology.

Let us consider \mathbf{x} as a stochastic p -vector of observed variables with population variance-covariance (or correlation) matrix Σ . Now, let us further consider the free parameters in a proposed model be contained in the q -vector θ . A structural equation model then implies a certain parameterization $\Sigma(\theta)$ of the variance-covariance matrix of the observed variables.

²A major reason for the frequent use of SEM is that it allows one to posit complex multivariable relationships among observed and latent variables whereby direct and indirect (mediated) effects are straightforwardly evaluated along with indexes of their estimation precision. Another major reason appears to be the availability of simple to use computer programs that require very little technical knowledge of the statistical models underpinning the modeling techniques. Unfortunately, methods and notions that are widely available and represented in user-friendly software packages risk the tendency to be quickly abused. Indeed, as indicated by Cortina (2002), structural equation modeling may be the best example of this phenomenon.

³We note that the word *parameterization* is spelled throughout using American English, but can also be alternatively spelled using British English as *parametrization* (see Dijkstra 1983, p. 71).

The null hypothesis specifically states that the structural equation model is correctly specified. This signifies that there are parameter values such that the model implied variance–covariance matrix equals the population variance–covariance matrix. Mathematically, this can be simply written as $H_0: \Sigma = \Sigma(\theta)$, for some θ . In other words, we say that the model holds if there exists a parameter value θ_0 such that $\Sigma = \Sigma(\theta_0)$. For correctly specified models, a minimum distance estimator of θ_0 can be obtained (e.g., a maximum likelihood estimator), which is asymptotically distributed according to a χ^2 distribution. A structural equation model is then said to hold if there exist values for the free parameters such that the model implied variance–covariance matrix equals the population variance–covariance matrix Σ of the observed vector x . Consequently, in any attempted data simulation studies, the correct parameterization $\Sigma(\theta)$ of the variance–covariance matrix of the observed variables is essential. Any deviations beyond those expected by estimating the model with respect to the sample covariance matrix (which introduces sampling variability) implies differential parameterization that is independent of the estimation method used (**Issue 1**).

So what might go wrong if one were to differentially parameterize a structural equation model? As an example, let us first consider a very special case in which the three path models displayed in Figure 1 are differentially parameterized (for complete details, see Hershberger 2006; Hershberger and Marcoulides forthcoming). We note that for simplicity these models only contain three observed variables, but could easily be expanded to more complicated models including ones with latent variables. The parameters for each of the three proposed models in the q -vector θ set are respectively as follows:

$$\theta_1 = \begin{bmatrix} b_{21} = 0.470 \\ b_{32} = 0.361 \\ s_{\varepsilon 1}^2 = 2.000 \\ s_{\varepsilon 2}^2 = 1.358 \\ s_{\varepsilon 3}^2 = 1.265 \end{bmatrix}, \theta_2 = \begin{bmatrix} 0.522 \\ 0.433 \\ 1.500 \\ 1.509 \\ 1.518 \end{bmatrix}, \theta_3 = \begin{bmatrix} 0.522 \\ 0.361 \\ 1.800 \\ 1.509 \\ 1.265 \end{bmatrix}$$

As it turns out, and despite the different parameterizations of the three models, they all have the exact same variance–covariance matrix implied by the parameter estimates. Such models are commonly referred to in the structural equation modeling literature as equivalent models (and there can potentially be an infinite number of such equivalent models, as initially introduced to the IS discipline by Chin (1998a; see Raykov and Marcoulides [2001, 2007] and references therein for details). We hasten to note that model equivalence is not defined by the data, but rather by an algebraic equivalence between model parameters. In other words, if a researcher were to use a different data set, the resulting implied

variance–covariance matrices estimated from each of the three models could be the same even though the estimated path estimates may differ. As a result, in such cases of model equivalence, the values for the tests of model fit (e.g., chi-square) will always be identical (see Hershberger and Marcoulides forthcoming).

This highlights one of our concerns with GLT’s narrow focus targeting mainly on comparing path estimates. Their Figure 1 may mislead readers into believing there can only be one correct path estimate when this is not necessarily true if equivalent models exist. Correct parameterization of an SEM model involves all parameter estimates in a holistic manner as they connect to one another to generate the implied variance–covariance matrix. But in cases not involving equivalent models, differentially parameterized models will lead to a variety of results (i.e., different parameter estimates in θ , different variance–covariance matrices, different values of tests of model fit, etc.).

As a further example, let us consider the situation of performing a canonical correlation analysis using either covariance or partial least squares estimation. Figure 2 shows three different model specifications, but all reflect the same canonical correlation analysis. Model 4 follows the MIMIC representation of the canonical model for covariance estimation (Bagozzi et al. 1981, p. 444) whereas Model 5 is an equivalent using PLS. Finally, Model 6 is the standard PLS representation for a canonical correlation (Chin, 1998b, p. 307). As in the previous example, independent of the data set being analyzed, all three models are equivalent and can be used to obtain the same estimates (i.e., the canonical correlation, canonical weights, and predictor variate cross loadings). Yet, the model specifications for path parameters to be estimated are not the same. Note that Model 4 has one construct, Model 5 has four constructs, and Model 6 has two constructs. The PLS Model 6 provides estimates of the canonical correlation and variate weights and loadings. You need to multiply the canonical correlation with the weights from the first variate C1 in Model 6 to match the path estimates in Model 5. To obtain the canonical correlation for Models 4 and 5, the square root of the first eigenvalue can be calculated as

$$\sqrt{\Lambda_y' R_{yy}^{-1} \Lambda_y}$$

where y represents the Vint indicators.

If we were to follow the procedure advocated by GLT, we run into problems. In order to obtain an estimate of the canonical correlation between two variate weights, GLT would have us

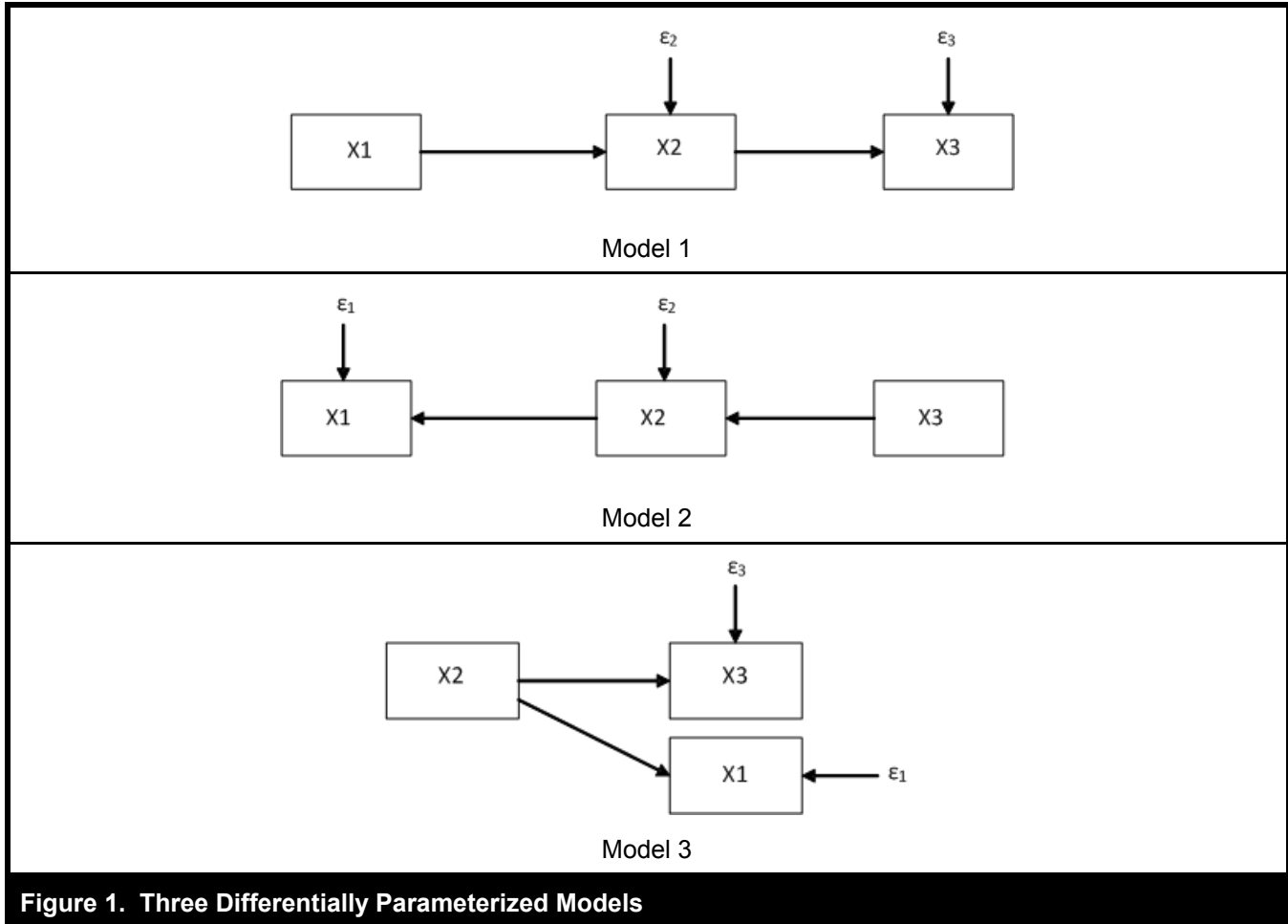


Figure 1. Three Differentially Parameterized Models

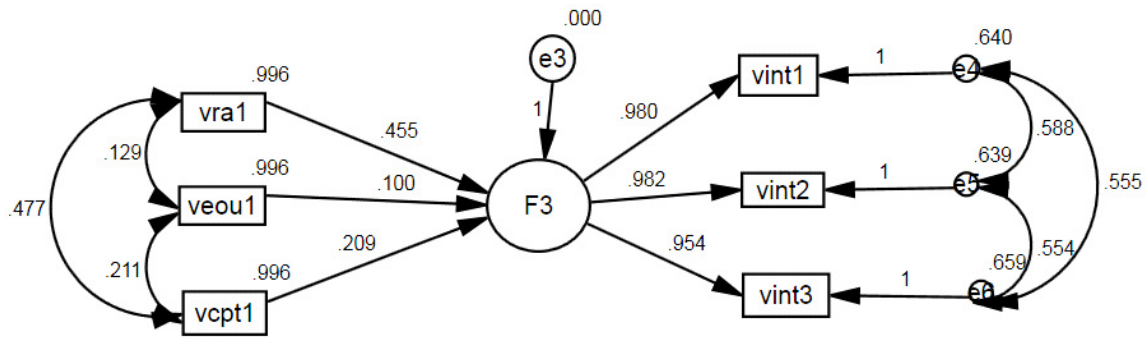
explicitly model two constructs and specify a “measurement model” to represent the same “underlying reality.” GLT state that the first three boxes leading to the choosing of a statistical technique are essentially the same regardless of which statistical technique is ultimately used; that all three techniques assume the same underlying reality (more on this issue later), the same research model, and the same data collected. Only then, GLT argue, can we compare and contrast specific path estimates and their significance (p. 4). Yet, as depicted in our Models 4 and 5, we do not need to explicitly model two underlying constructs and the path between them to obtain the canonical correlation. In fact, due to identification constraints, SEM cannot be used to analyze Model 6.

Issue 2: Comparisons to Regression Are Trivial

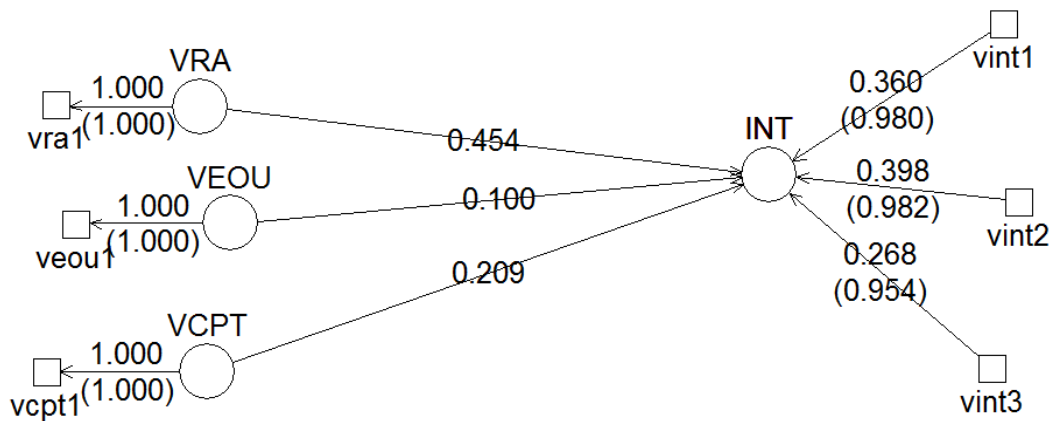
Now let us move onward to the situation where differentially parameterized models yield different results to further high-

light our central concerns with the GLT piece. As indicated in our Foreword, any comparison of the performance of multiple regression relative to either PLS or SEM is trivial since it is well known that an analysis of the same data and model based on a single regression equation using these approaches will always yield identical results (**Issue 2**). This is because regression analysis is a first generation technique that is subsumed under the second generation techniques of PLS and SEM (e.g., Chin 1998b, pp 296-297). Therefore, GLT’s inclusion of the second columns in Tables 1 and 2 and similar boxes in Figure 1 to highlight regression analysis as on par with PLS and SEM is unnecessary since one could mathematically expect the results to be the same.

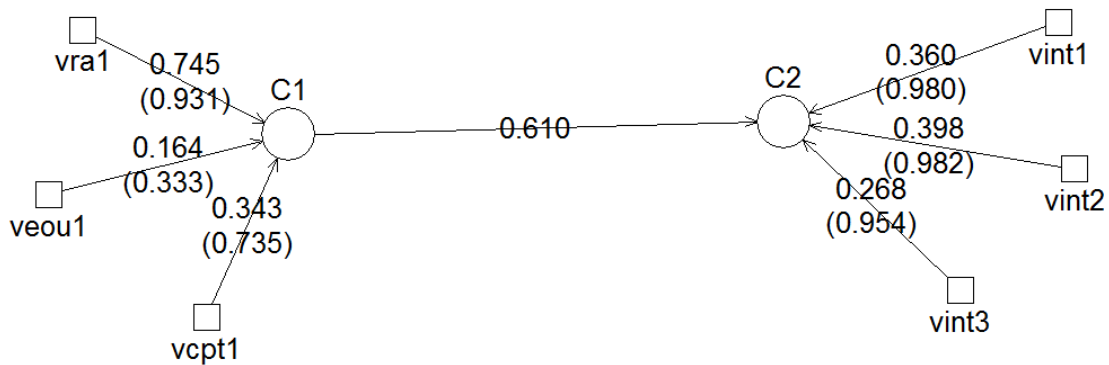
The key question is how composites are formed prior to conducting the regression analysis. Once the decision is made on how to similarly create composite scores, running these scores using PLS, SEM, or regression will end up with exactly the same estimates.



Model 4. Multiple Indicator/Multiple Cause (MIMIC) Mode for Canonical Correlation Analysis Using AMOS



Model 5. MIMIC Model for Canonical Correlation Analysis Using PLS-Graph



Model 6. Standard Model Specification for Canonical Correlation Analysis Using PLS-Graph

Figure 2. Three Different Models Using Covariance ML Estimation (Model 4) and Partial Least Squares (Models 5 and 6). For Models 5 and 6, indicator variate correlates are in parentheses.

Are there cases where these methods might actually yield different results? The answer is a definitive yes! It can occur when you differentially parameterize the measurement model. Such is the case with the model GLT presented and adapted from Goodhue et al. (2006). For the regression model, GLT indicated that they “chose the most common approach used in practice: equal indicator weights” (p. 712), versus the composite scores estimated by PLS or path estimates by LISREL (i.e., where the factors are presented for ξ_1, ξ_2, ξ_3 , etc.). If one were to run analyses as just defined, you would likely get very different coefficient estimates. Does that imply that the methods differ in their estimation? Absolutely not! Any observed differences would merely be a function of the differentially parameterized models being analyzed (i.e., unit weights for regression, PLS weights for PLS, and covariance path estimates for LISREL). If you were to select unit weights for all three techniques, the results would once again be identical.

Continuing on this point, for ease of presentation and simplicity, let us focus on regression analysis (although the same argument could readily be made for any of the other modeling techniques examined). In their Table 2, GLT indicated that when using the regression approach for each supposed construct “**indicator weights must be prespecified – often set to equal.** Or can use factor weights” (p. 711). They opted not to use exploratory factor analysis to determine the appropriate weights for each indicator because they considered the use of equal weights to be “the most common approach used in practice” (p. 10). Thus, to make our point we use both of these types of construct scores in a simple regression analysis (we return in more detail to the issue of construct measurement in a later section). The regression analysis examines an outcome variable and the predictive capabilities of construct scores determined on the basis of five indicators and a realistic sample size of $n = 120$.

Using the above two types of construct scores in a regression analysis using SPSS, leads to the output provided in Figures 3 and 4. As can be readily seen by examining the output presented in Figures 3 and 4, the values of the R square, the standard error of estimate, the beta weight, the t-value, the significance test statistic, the intercept value (the constant), etc. are all the same! However, there is one value that is different: the value of the regression coefficient estimate (and its standard error) for the construct scores obtained when taking the straight average of the indicators (XAVERAGE) which are of equal weights is quite different than that of the regression coefficient estimate (and its standard error) obtained using the factor analytic approach (XWEIGHTS). We note that everything else related to this estimated parameter is the same for both types of construct scores.

So now we ask, does that imply that the two regression methods differ in their estimation? Absolutely not! Both models were examined using the same simple linear regression technique. Any observed differences are merely a function of the differentially parameterized models being analyzed. In this case, a parameterization using the straight average of the indicators as construct scores versus using exploratory factor analysis to determine the appropriate weights for each indicator and then computing construct scores led to the differences in the results. It is obvious that the same argument could readily be made when making comparisons between any of the other techniques.

Now take our earlier canonical correlation case as another example. If we use a straight average to create two variates and submit them to SEM, PLS, and regression analyses, we obtain exactly the same correlation estimate of 0.58 for all three techniques. If we instead differentially weight each indicator consistent with the first principal component, the estimate changes to 0.587. But the change is identical for all three analyses. Thus, the fourth box in Figure 1 and step 1 in Figure 4 as proposed by GLT is problematic because they confound the estimation procedure with differences in how the measures are parameterized.

Issue 3: Distinguish Between Latent Constructs and Composite Variables

We argue that it is important to distinguish between latent constructs and composite variables when performing legitimate comparisons. GLT’s comparison problem might simply be a function of the so-called formative versus reflective (composite versus latent variables) measurement debate (see Hardin and Marcoulides 2011 and references therein), particularly because many of these publications are believed to have misinformed readers due to the lack of theory underlying formative measurement and a misinterpretation of the early psychometric literature (**Issue 3**). GLT readily acknowledged that “statistical techniques using composites and those using latent variables are quite distinct” (p. 705). Unfortunately, and despite the overwhelming statistical evidence provided in the literature (see Hardin et al. 2011; Hardin and Marcoulides 2011; Marcoulides et al. 2009; Treiblmaier et al. 2010), GLT then obfuscate matters by stating that “both composites and latent variables are intended to represent the same things: theoretical constructs that are not directly observable” (p. 711). This statement makes no distinction between composites and latent variables. This lack of statistical precision and accuracy is especially problematic in the case of formative measurement.

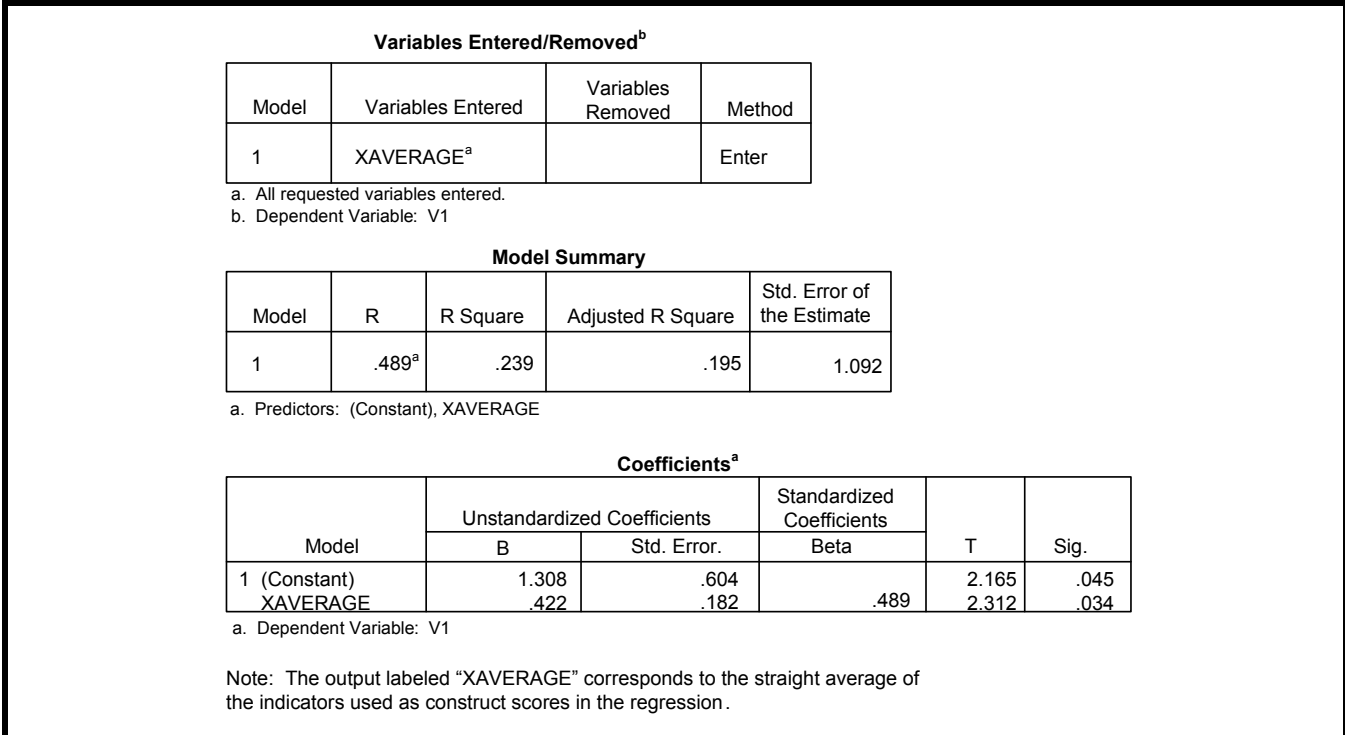


Figure 3. Regression

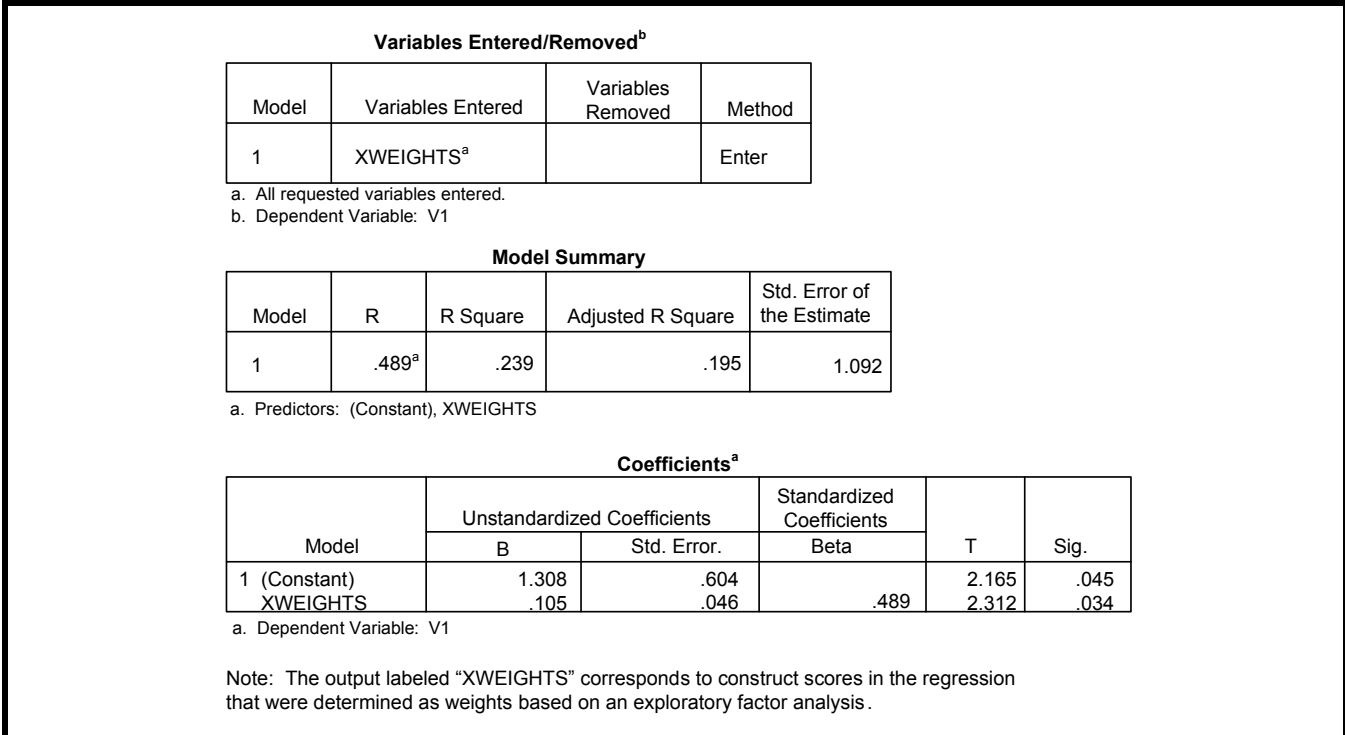


Figure 4. Regression

Hardin and Marcoulides (2011, p. 2) noted that the statistical ideology behind a latent variable is quite elementary:

If a latent variable underlies a set of observed variables or indicators, then conditioning the indicators on the latent variables makes those indicators statistically or conditionally independent (also referred to as local independence; see Raykov and Marcoulides 2011).

Latent variable measurement then concerns the process of ensuring that local independence is satisfied for a selected set of observed variables or indicators and this can be done via the use of a model such as a common factor model. The common factor model then stipulates that the correlations among the observed variables can be explained by their regression on the latent variable (this guarantees that the observed variables are independent after conditioning on the latent variable).

In formative measurement, the relationship between the observed variables and the composite is reversed, whereby the composite is regressed on its observed variables. This can be readily denoted following Treiblmaier et al.'s (2010) and McCallum and Browne's (1993) recommended convention of distinguishing between composite variables as \mathcal{F} and latent variables as F . Thus, in models with \mathbf{x} observed variables, formative models would be said to have $\mathbf{x} \rightarrow \mathcal{F}$ paths, whereas reflective models have $F \rightarrow \mathbf{x}$. We note that extensive discussions on this topic have recently been offered by Bagozzi (2011), Bollen (2011), Chin (2010), Diamantopoulos (2011), Edwards (2011), Hardin et al. (2011), Hardin and Marcoulides (2011), and Treiblmaier et al. (2010), so there really is no need to rehash the complete contents of these papers here. We recommend these papers to any researchers needing both a historical and thorough understanding of the debate in the literature. Suffice it to say, however, that operationalizing formative models as closely matched common factor equivalents does not eliminate the issue that they are differentially parameterized models (in other words, an $\mathbf{x} \rightarrow \mathcal{F}$ path is not the same as a $\mathbf{x} \rightarrow F$ path). Although a new methodology for unambiguously implementing an F that will closely approximate an \mathcal{F} has recently been proposed by Treiblmaier et al. Thus, a comparison between PLS and SEM based on models generated through such a methodology would indeed be of interest (although research still needs to be done to determine the actual degree of correspondence between \mathcal{F} and F). Implementing this approximation, however, requires a two-step approach that splits the determinate part of the formative composite into two or more composites and then models them as latent variables (i.e.,

common factors, which can then theoretically also be placed into any much larger latent variable modeling framework).

The new approach offered by Treiblmaier et al. operationalizing formative models as closely matched common factor equivalents was not utilized in the GLT study. In fact, neither was any other appropriate approximation method. What was done is a simple substitution of estimates for the F s. And although it is the case that in PLS estimation, substitution of estimates for F is routinely done, there are well-known and clear consequences (see complete details in Treiblmaier et al.), not the least of which that "not all parameters will be estimated consistently" (Dijkstra 2010, p. 37). Professor Dijkstra also added that "PLS replaces latent variables by proxies who can 'never' represent them exactly....So the parametrization for the proxies is incorrect" (Dijkstra, e-mail to Authors, January 26, 2010).

Finally, we should mention our concern with the rather rigid assumption provided by GLT in the first box of their Figure 1 process. Here GLT state that all researchers begin with the positivist goal of uncovering an underlying reality and seem to imply that this objective is true for all three techniques. We are less sanguine regarding this position. In the case of PLS, for example, Dijkstra (2010, p. 23) argues that PLS is better suited for constructing composites "that extract information from high-dimensional data in a predictive, useful way." In agreement, Chin (2010) noted that eschewing the "true" model for prediction focus can be a rationale for the use of PLS and is "more akin to the American philosophical perspective of pragmatism" (p. 668) than a pure positivist perspective. Schneeweiss (1991), in fact, clearly states that PLS does not necessarily impose GLT's stipulation that all techniques assume underlying "theoretical constructs that are not directly observable" (p. 705). Specifically, Schneeweiss (1991, p. 155) notes that

The PLS model is defined on a set of jointly distributed random variables by partitioning this set into disjoint subsets and by specifying dependencies between these sets. No further assumptions are required. The PLS parameters and latent variables can be defined so as to represent these dependencies in a concise fashion. So long as the iterative PLS procedure defining these entities converges, they are well defined, irrespective of any specific model structure, in particular irrespective of whether a LISREL model pertains or not.

As for regression, it does not take a position on how composites are formed. Rather, variables are simply deployed

(calculated composites or otherwise) and the dependent variable is studied as a function of other independent variables. Thus, it is unclear why GLT state that “regression assumes that each construct has a knowable value that is a composite of its equally weighted indicator scores...[and that] regression techniques require estimating construct scores” (p. 709). Our search through the entire 703 pages of Cohen et al. (2003), as a check, failed to uncover such prescriptions.

Issue 4: The Ratio of the Largest Eigenvalue to the Sum of the Squared Loadings and Issue 5: The Number of Indicators

As indicated by Marcoulides et al. (2009) and McDonald (1996), consistent estimates will be obtained when the number of indicators goes to infinity (**Issue 5**), in practical situations this does not often occur and it did not occur in the Goodhue et al. (2006) Monte Carlo comparison study. We note that to date the actual degree of correspondence between F and \hat{F} in practice has yet to be determined, so this is one area where more research is clearly needed, but, we would suggest, not by using the methods suggested by GLT.

Mathes (1993) also showed that PLS can be regarded as providing approximate estimates of a very specific common factor model. This notion was furthermore emphasized by Schneeweiss (1993), who indicated that the two types of approaches are related to each other and in some specific situations can come quite close to each other (for further details, see equations 9 through 12 in Schneeweiss). As indicated in our Foreword (**Issue 4**),

the key to governing the closeness of PLS to SEM latent variables for a particular block is the ratio of the largest eigenvalue of the error covariance matrix to the sum of squared loadings. In situations where this ratio, or by the model specified, is made small (e.g., path coefficients and loadings), estimates obtained from PLS and SEM will be very close to each other or approximately equal (p. 173).

The fact that one can obtain approximate estimates in PLS has been known for decades (see also McDonald 1996; Tenenhaus 2007 and references therein). Tenenhaus (2007) even provided a complete table (Table 9) of the approximations for a very specific parameterized model case (.206 versus .199, .163 versus .173, etc), which he called the Fornell case: when all the coefficient weights or loadings relative to a “latent variable” are of the same sign and the observed variables are of comparable order of magnitude (see also Fornell and Bookstein, 1982). So the differences

between the approaches as a function of approximations are quite well known. Tenenhaus even showed the differences between parameter values for the Fornell model estimated by a so-called LISREL type model (they actually used the AMOS program to obtain the estimates) and PLS (Customer Expectation → Perceived Quality =.545 for PLS estimates versus .856 for LISREL). They note that the differences between PLS and LISREL estimates of a causal model come from the order in which model parameters and latent variable or composites are calculated, and from the constraints on these. This is because in PLS the reflective scheme assumed for the latent variable is inverted (this was also pointed out by Marcoulides et al.).

A study that appears to be prominently referenced by GLT, albeit inappropriately as an exemplary study for comparing the efficacy of PLS with that of regression and/or CB-SEM, is one by Hwang et al. (2010). Thus, in order to set the record straight we also provide a synopsis of this study and address the issue of the viability of their comparisons. Hwang et al (pp. 701-702) clearly recognize the differences between the approaches in terms of model specification and parameter estimation ahead of any analyses conducted and indicate that

this leads to the specification of different sets of model parameters for latent variables (i.e., factor means and/or variances in covariance structure analysis versus component weights in partial least squares)...The algebraic formulations underlying the three approaches seem to result in substantial difference in the procedures of parameter estimation.

They go on to point out again that the “approaches estimate different sets of model parameters....Thus, in this study we evaluate and report the recovery of the estimates of a common set of parameters” (p. 703). **They conclude by readily acknowledging their inability to provide correctly parameterized comparisons among the approaches** and indicate that “we generated simulated data on the basis of covariance structure analysis...we adopted the procedure because it was rather difficult to arrive at an impartial way of generating synthetic data for all three approaches” (p. 710). These statements carefully stipulate the precise conditions of their analysis and fully acknowledge the limitations of their “comparisons” between apples and oranges.

Comparing Apples with Oranges

Can one really compare apples with oranges? Certainly one can, but why would such a comparison be interesting or make sense? Is it to determine the difference between a green apple

(which might even be that color due to its type) with a green orange (which is most likely that color because it has not yet ripened)? Sure one can! But what insight would one gain from such a comparison? Is it to show that green apples can sometimes be eaten, whereas green oranges likely cannot? If so, fine. But ultimately any intent to compare apples with oranges must first acknowledge that one is comparing different though related things (“we all different, but in the end, we all fruit,” Gus Portokalos, “My Big Fat Greek Wedding,” 2002). The same holds for comparisons between differentially parameterized models: it is like comparing apples with oranges!

As indicated in our Foreword, “in summary, it should be clear to the IS research community that comparison of PLS to other methods cannot and should not be applied indiscriminately.” As a discipline, we need to compare apples with apples and oranges with oranges. “Ignoring any of the above issues could lead to incorrect conclusions or lead to overstating the importance of outcomes observed in a study” (Marcoulides et al. 2009, p. 174). In their abstract, GLT claimed their Issues and Opinions piece was written to give “an overview of the process of comparison research with a focus on what is required to make those comparisons legitimate” (p. 703). Their justification was based on the one instance where our Foreword had “already been used by at least one reviewer as justification for recommending rejection of a comparison paper submitted to a top-tier IS journal” (p. 704). GLT admit that they “sought to understand what MCS meant by the term ‘correct parameterization.’ [Yet,] ultimately, we were unable to determine what specifically MCS were suggesting” (p. 704).

Our intent in our short Foreword was not, as GLT suggest, to “provide clear guidance on how to conduct ‘legitimate’ comparisons” (p. 703), but rather to provide an overview of the papers in the Special Issue and to highlight a problem common to a number of the submissions that were ultimately not published. Strategies for conducting legitimate comparisons are readily available in the literature (e.g., Treiblmaier et al. 2010). In this response we have clarified what we meant by correct parameterization and we have addressed our main concerns about the GLT piece.

Acknowledgments

This research was partially supported by Sogang Business School’s World Class University Program (R31-20002) funded by the Korea Research Foundation and the Sogang University Research Grant of 2011.

References

- Bagozzi, R. P. 2011. “Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations,” *MIS Quarterly* (35:2), pp. 261-292.
- Bagozzi, R. P., Fornell, C., and Larcker, D. F. 1981. “Canonical Correlation Analysis as a Special Case of Structural Relations Model,” *Multivariate Behavioral Research* (16), pp. 437-454.
- Bollen, K. 2011. “Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models,” *MIS Quarterly* (35:2), pp. 359-372.
- Chin, W. W. 1998a. “Issues and Opinions on Structural Equation Modeling,” *MIS Quarterly* (22:1), pp. vii-xvi.
- Chin, W. W. 1998b. “The Partial Least Squares Approach for Structural Equation Modeling,” in *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Lawrence Erlbaum Associates, pp. 295-336.
- Chin, W. W. 2010. “How to Write Up and Report PLS Analyses,” in *Handbook of Partial Least Squares: Concepts, Methods, and Applications*, V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Heidelberg, Germany: Springer, pp. 655-690.
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Mahwah, NJ: Lawrence Erlbaum Associates.
- Cortina, J. M. 2002. “Big Things Have Small Beginnings: An Assortment of ‘Minor’ Methodological Misunderstanding,” *Journal of Management* (28), pp. 339-363.
- Diamantopoulos, A. 2011. “Incorporating Formative Measures into Covariance-Based Structural Equation Models,” *MIS Quarterly* (35:2), pp. 335-358.
- Dijkstra, T. 1983. “Some Comments on Maximum Likelihood and Partial Least Squares Methods,” *Journal of Econometrics* (22), pp. 67-90.
- Dijkstra, T. 2010. “Latent Variables and Indices: Herman Wold’s Basic Design and Partial Least Squares,” in *Handbook of Partial Least Squares: Concepts, Methods, and Applications: Computational Statistics* (Volume 2), V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Heidelberg, Germany: Springer, pp. 23-46.
- Edwards, J. R. 2011. “The Fallacy of Formative Measurement,” *Organizational Research Methods* (14:2), pp. 370-388.
- Fornell, C., and Bookstein, F. 1982. “Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory,” *Journal of Marketing Research* (19), pp. 440-452.
- Fox, K. A. 1980. “Philosophical Implications of a System of Social Accounts,” *Philosophics* (25), pp. 33-54.
- Goodhue, D., Lewis, W., and Thompson, R. 2006. “PLS, Small Sample Size, and Statistical Power in MIS Research,” in *Proceedings of the 39th Annual Hawaii Conference on System Sciences*, Los Alamitos, CA: IEEE Computer Society Press.
- Goodhue, D., Lewis, W., and Thompson, R. 2012a. “Comparing PLS to Regression and LISREL: A Response to Marcoulides, Chin, and Saunders,” *MIS Quarterly* (36:3), pp. 703-716.

- Goodhue, D., Lewis, W. and Thompson, R. 2012b. "Does PLS Have Advantages for Small Sample Size or Non-Normal Data?," *MIS Quarterly* (36:3), pp. 981-1001.
- Hardin, A., Chang, J., Fuller, M., and Torkzadeh, G. 2011. "Formative Measurement and Academic Research: In Search of Measurement Theory," *Educational and Psychological Measurement* (71:2), pp. 281-305.
- Hardin, A., and Marcoulides, G. A. 2011. "A Commentary on the Use of Formative Measurement," *Educational and Psychological Measurement* (17), pp. 1-12.
- Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., and Hong, S. 2010. "A Comparative Study of Parameter Recovery of the Three Approaches to Structural Equation Modeling," *Journal of Marketing Research* (67), pp. 699-712.
- Hershberger, S. L. 2006. "The Problem of Equivalent Models," in *Structural Equation Modeling: A Second Course*, G. R. Hancock and R. O. Mueller (eds.), Greenwich, CT: Information Age Publishing, pp. 13-41.
- Hershberger, S. L., and Marcoulides, G. A. Forthcoming. "The Problem of Equivalent Models," in *Structural Equation Modeling: A Second Course* (2nd ed.), G. R. Hancock and R. O. Mueller (eds.), Greenwich, CT: Information Age Publishing.
- MacCallum, R. C., and Browne, M. W. 1993. "The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues," *Psychological Bulletin* (114), pp. 533-541.
- Marcoulides, G. A., Chin, W., and Saunders, C. 2009. "A Critical Look at Partial Least Squares Modeling," *MIS Quarterly* (33:1), pp. 171-175.
- Marcoulides, G. A., and Saunders, C. 2006. "Editor's Comments: PLS: A Silver Bullet?," *MIS Quarterly* (30:2), pp. iii-x.
- McDonald, R. P. 1996. "Path Analysis with Composite Variables," *Multivariate Behavioral Research* (31), pp. 239-270.
- Mathes, H. 1993. "Global Optimization Criteria of the PLS-Algorithm in Recursive Path Models with Latent Variables," in *Statistical Modelling and Latent Variables*, K. Haagen, D. J. Bartholomew, and M. Deistler (eds.), Amsterdam: Elsevier, pp. 229-248.
- Muthén, L. K., and Muthén, B. O. 2010. *Mplus User's Guide*, Los Angeles: Muthén and Muthén.
- Raykov, T., and Marcoulides, G. A. 2001. "Can There Be Infinitely Models Equivalent to a Given Covariance Structure Model?," *Structural Equation Modeling* (8), pp. 142-149.
- Raykov, T., and Marcoulides, G. A. 2007. "Equivalent Structural Equation Models: A Challenge and Responsibility," *Structural Equation Modeling* (14), pp. 695-700.
- Raykov, T., and Marcoulides, G. A. 2011. *Introduction to Psychometric Theory*, New York: Routledge.
- Schneeweiss, H. 1991. "Models with Latent Variables: LISREL Versus PLS," *Statistica Neerlandica* (45:2), pp. 145-155.
- Schneeweiss, H. 1993. "Consistency at Large in Models with Latent Variables," in *Statistical Modelling and Latent Variables*, K. Haagen, D.J. Bartholomew, and M. Deistler (eds.), Amsterdam: Elsevier, pp. 299-320.
- Tenenhaus, M. 2007. "Between PLS Path Modeling and ULS-SEM," paper presented at the International Symposium on PLS and Related Methods, Oslo, Norway, September.

- Treiblmaier, H., Bentler, P. M., and Mair, P. 2010. "Formative Constructs Implemented via Common Factors," *Structural Equation Modeling* (18:1), pp. 1-17.

About the Authors

George A. Marcoulides is a professor of Research Methods & Statistics in the Graduate School of Education and in the Interdepartmental Graduate Program in Management (IGPM) in the A. Gary Anderson Graduate School of Management at the University of California, Riverside. He was previously a professor of Statistics in the Department of Information Systems and Decision Sciences at California State University, Fullerton. He has also been a visiting professor at the University of California, Los Angeles, the University of California, Irvine, the University of Geneva, and the University of London. He has served as a consultant to numerous government and state agencies, companies in the United States and abroad, and to various national and multinational corporations. He has coauthored or coedited 15 books and edited volumes. His research contributions have received best paper awards from the Academy of Management, the Decision Sciences Institute, and the American Educational Research Association—University Council for Educational Administration. He is a Fellow of the American Educational Research Association, a Fellow of the Royal Statistical Society, and a member of the Society of Multivariate Experimental Psychology. He is currently editor of the journals *Structural Equation Modeling* and *Educational and Psychological Measurement*, editor of the Quantitative Methodology Book Series, and on the editorial board of numerous other scholarly journals.

Wynne W. Chin is the C. T. Bauer Professor of Decision and Information Sciences at the University of Houston and the World Class University Professor in the Department of Service Systems Management and Engineering at Sogang University. He received AB, MS, MBA, and Ph.D. degrees in Biophysics, Biomedical/Chemical Engineering, Business, and Computers and Information from the University of California, Berkeley, Northwestern University, and the University of Michigan, respectively. He is the developer of PLS-Graph, the first graphical based software dating back to 1990 to perform partial least squares analysis and used by more than 8,000 researchers worldwide. Wynne's research includes sales force automation, IT service, adoption, outsourcing, acceptance, satisfaction, group cohesion and negotiation, and psychometric and casual modeling issues. Wynne is on the editorial board of *Structural Equation Modeling*, *Journal of Information Technology*, and *IEEE Transaction of Management*, and previously served as coeditor of *Data Base* and on the boards of *Journal of AIS*, *Information Systems Research*, and *MIS Quarterly*. Wynne currently resides in Sugar Land, TX, with his wife Kelly, his two daughters—Christina (23 years old) and Angela (18 years old)—and his dog Goodgirl (72 dog years).

Carol Saunders is currently professor of Management at the University of Central Florida and Schoeller Senior Fellow at Friedrich-Alexander-Universität Erlangen-Nürnberg. She is a LEO award winner for lifetime accomplishments to the IS discipline and an Association of Information Systems Fellow. She has served on a number of editorial boards, including a three-year term as editor-in-chief of *MIS Quarterly*. She served as general conference chair of ICIS '99, as well as Telecommuting '96. She was a Distinguished

Fulbright Scholar at the Wirtschafts Universität – Wien (WU) in Austria, and earlier held a Professional Fulbright with the Malaysian Agricultural Research and Development Institute. She has held research chairs in New Zealand, Singapore, and the Netherlands. Her current research interests include overload, back-sourcing, virtual teams, virtual worlds, and time. Carol has published in top-ranked Information Systems, Management, Computer Science, and Communication journals.

Copyright of MIS Quarterly is the property of MIS Quarterly & The Society for Information Management and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.