Communications of the Association for Information Systems

Volume 34 Article 30

1-2014

Improving Between-group Comparisons in IS Research Through the Use of SEM and Latent Variables: An Introduction and Some Examples

Miguel Aguirre-Urreta
DePaul University, maguirr6@depaul.edu

Recommended Citation
Aguirre-Urreta, Miguel (2014) "Improving Between-group Comparisons in IS Research Through the Use of SEM and Latent Variables: An Introduction and Some Examples," *Communications of the Association for Information Systems*: Vol. 34 , Article 30.
DOI: 10.17705/1CAIS.03430
Available at: https://aisel.aisnet.org/cais/vol34/iss1/30

Communications of the Association for Information Systems

CAIS

Improving Between-group Comparisons in IS Research Through the Use of SEM and Latent Variables: An Introduction and Some Examples

Miguel I. Aguirre-Urreta

School of Accountancy and Management Information Systems, Richard H. Driehaus College of Business, DePaul University

maguirr6@depaul.edu

Abstract:

Information Systems researchers are often interested in comparing outcomes across different groups of interest, as in the case of experimental and quasi-experimental studies. These designs have traditionally been modeled, as we show through a review of our literature, by using analysis of variance techniques on observed scores, typically the sum or average of all items measuring the dependent variable of interest. These designs, however, can be analyzed with structural equation modeling and latent variables (SEM-LV) techniques, which can better accommodate measurement error and more complex models than would otherwise be possible using the traditional techniques. This research introduces the foundations of the SEM-LV approach for these research designs and highlights these advantages, and provides several examples that underscore the flexibility of the latent variable techniques discussed here. It also compares the two main alternatives for implementing this approach and discusses the advantages and disadvantages of each.

Keywords: latent variables, experimental design, analysis of variance, structural equation modeling

Editor's Note: The article was handled by the Department Editors for Information Technology and Systems

Volume 34, Article 30, pp. 595-618, January 2014

I. INTRODUCTION

Information Systems (IS) researchers are often interested in comparing outcomes across different groups of interest, as in the case of experimental and quasi-experimental research designs, the latter lacking random assignment of participants to treatments or experimental conditions. Analysis of these designs has traditionally been performed using the family of techniques under the general heading of analysis of variance (ANOVA), such as one-way, two-way with interaction, factorial, repeated measures, ANCOVA when employing a covariate, MANOVA when there are multiple dependent variables, MANCOVA when adding covariates to that design, etc. These techniques are well established and have long been used; indeed, ANOVA was used for the first time more than ninety years ago. It is important to recognize, however, that these techniques are special cases of the general linear model and can be subsumed under it. Moreover, the general framework is itself much more flexible and can accommodate more varied designs and tests of underlying assumptions, thereby strengthening experimental research. Therefore, this article has two main objectives. First, to introduce to the IS community the analysis of these research designs using latent variable techniques and to discuss the advantages of using these approaches. To this extent, we seek to provide an introduction into the basic mechanisms by which these techniques work and point to the relevant literature for those interested in applying these models to collected data. Second, we show how these techniques can be applied to research designs commonly used by IS researchers and how to examine those designs using the approaches proposed here.

The main goal of this research is to introduce the use of structural equation modeling with latent variables (hereafter SEM-LV for short) for the analysis of research designs that include between-group comparisons with either dependent variables measured with error or intervening variables between the between-group indicators and the dependent variables in the model (or any combination of the two) to the IS research community. Particular emphasis is placed on the integration and comprehensive modeling of manipulation checks as part of the overall research model. Though many of the discussions included here are known in the methodological literature, they have still to find their way into the work of the majority of IS researchers conducting this kind of studies. In order to achieve these objectives, the rest of this article is organized as follows. First, we present a framework that outlines in which scenarios either traditional techniques or the ones proposed here are most appropriate and discuss why. Then the framework is used to categorize recent IS research on conducting between-group comparisons, which helps underscore the existing mismatch between research designs and analytical techniques in the discipline. Third, the two alternative approaches to the specification of between-group designs in SEM-LV are discussed, and compared to traditional techniques with regards to their expected outcomes. Fourth, more complex designs, drawn from contemporary IS research, are used to better show how complexity can be accommodated by the proposed techniques. Finally, limitations of those are discussed and a summary of this research, with a list of more advanced readings, is provided.

II. RESEARCH DESIGNS AND ANALYTICAL TECHNIQUES

Throughout this article we compare SEM-LV techniques to what we call the *traditional ANOVA approach,* and so a definition of the latter is in order here. The family of analyses under the ANOVA heading are quite well established in IS research. As we discuss later—and support through a review of IS between-group research—the traditional ANOVA approach is commonly accepted as the appropriate way of analyzing research designs, including between-group comparisons. Some of these studies include Stewart [2006] on the effects of links between websites of different organizations on trust on these organizations; Kuechler and Vaishnavi [2006] on the effects the explicit inclusion of goal information has on comprehension, decision confidence, and recall; and Jiang and Benbasat [2007] on the effects of alternative presentation formats and task complexity on product knowledge and perceived website diagnosticity.

In the traditional two-group design, subjects are assigned to either a treatment or a control group (randomly in an experiment and otherwise if in a quasi-experiment), and the independent variable is manipulated only for the treatment condition. In a between-group design, the independent variable is either categorical or nominal, representing group assignment, and the dependent variables can be directly observable, a multi-item measure representing a latent construct, or any combination of the two. When the latter are employed, researchers determine whether reliability is appropriate, through the use of Cronbach's alpha or a similar statistic, and then proceed to test for equality of between-group means, through an analysis of variance on the sum of the items collected. Often, some form of factor or component analysis is performed prior to statistical testing, and items may be dropped if they fail to

load adequately; however, the analysis itself is typically conducted on a sumscore of all the retained items. In more sophisticated designs, additional questions serve as "manipulation checks" and verify whether the intended manipulation was indeed effective; however, these generally are not modeled in the statistical analysis itself and are used only to present evidence as to the adequacy of the manipulation or treatment. Alternatively, ANOVA models can be formulated under the multiple regression framework, in which the different treatments are implemented through the inclusion of dummy variables that represent the presence or absence of a treatment condition (or sets of these variables if there are multiple levels of a treatment). It should be noted, however, that both representations are equivalent and produce identical results. Researchers generally prefer the ANOVA formulation for the presentation of these analyses. Even if formulated as a regression, however, these models are still focused on the comparison of averages between groups of interest. As discussed below, the preceding is an accurate picture of the majority of between-group comparisons conducted in IS research, as shown through a review of recent research in the field.

In order to better illustrate how the proposed modeling approach, based on SEM-LV techniques, differs from current practice in the field, we compare between-group research designs across two important dimensions, namely whether the dependent variable or variables of interest are deemed to be latent in nature (and thus subject to imperfect measurement by one or more manifest indicators) and whether there are one or more intervening variables between the manipulation or between-group indicator (for those scenarios where no manipulation has been effected) and the dependent variables of interest or any combination of the two scenarios. This framework, depicted in Figure 1 below, allows us to make a clear distinction between those scenarios where ANOVA techniques (which include all related techniques, such as ANCOVA, MANOVA, or MANCOVA) and SEM-LV techniques, are most appropriate. These two dimensions are discussed in more detail below.

		Intervening variables between manipulation or between-groups indicator and ultimate dependent variable	
		N	**Y**
Measurement error in any dependent variable	**N**	Quadrant 1: ANOVA Techniques	Quadrant 3: SEM-LV
	Y	Quadrant 2: SEM-LV	Quadrant 4: SEM-LV

Figure 1. Categorization of Between-group Research Designs and Most Appropriate Statistical Technique

Measurement Error in Any Dependent Variable

This dimension of the framework in Figure 1 distinguishes between research designs where the dependent variables can be directly measured without error in an objective manner from those in which the manifest variables involved in the comparison are taken to be indicators of an unobservable latent variable. Both cases are common in IS research. Examples of the former include number of passwords successfully recalled [Zhang, Luo, Akkaladevi, and Ziegelmayer, 2009], decision correctness [Heninger, Dennis, and Hilmer, 2006], test and task performance [Santhanam, Sasidharan, and Webster, 2008], and decision time and quality [Tan, Teo, and Benbasat, 2010]. In other cases, however, dependent variables are measured with one or more imperfect indicators representing an underlying latent variable, which is itself the main focus of interest in the comparison. Examples here include cognitive and affective involvement [Jiang, Chan, Tan, and Chua, 2010], ease of understanding [Burton-Jones and Meso, 2008], trusting beliefs [Kim and Benbasat, 2006], and perceptions of organizational information sharing [Arnold, Benford, Hampton, and Sutton, 2010].

Whether a research design includes one or more dependent variables that are unobservable in nature and measured with imperfect indicators is an important determinant of the most appropriate statistical technique for data analysis. When this is the case, SEM-LV techniques are more suitable, largely due to their ability to model each individual item—both the portion representing the latent variable of interest and any residual variance due to measurement error—separately. Doing so allows researchers to obtain more accurate and unbiased estimates of the experimental or between-group effects of interest. In addition to more accurate estimates, SEM-LV results will exhibit more statistical power due to the removal of error variance from the denominator of the effect size, thus resulting in a larger statistic which, for any given sample size, will result in enhanced power to detect significant differences.

On the other hand, when researchers employ ANOVA to analyze differences across constructs measured with multiple items, each of which contains measurement error, a number of tradeoffs are necessary. Most notably,

researchers are limited to use sum or average scores over all items as proxies for the latent variable of interest, thus implicitly assuming that all indicators represent the construct equally well (which is equivalent to assuming essential tau-equivalence as the underlying measurement model [Millsap and Everson, 1991]). In contrast to SEM-LV techniques, as just noted, traditional ANOVA techniques have no mechanism for modeling measurement error and researchers, therefore, must operate as if all variables were error-free. While it is recognized that this is not the case and hence the use of reliability statistics to ascertain the degree to which high levels of measurement error are present in the composite, these do not typically figure prominently in the discussion of obtained effects after some reliability threshold has been achieved. Finally, because traditional ANOVA techniques cannot take measurement error into account, estimates of between-group effects are underestimated, which further limits the statistical power of these analyses [Ree and Carretta, 2006].

A related issue is more conceptual in nature and has to do with the level at which causal inferences are established. In general, theories, particularly those used in IS research, are posited, developed, and tested in terms of constructs that are unobservable and that are represented with multiple items which are individually susceptible to error, that is, in terms of latent variables. In experimental research, the manipulation is a means of effecting change in the predictor construct only in some groups of subjects so that causality in observing changes in the dependent construct is strengthened. Under the traditional ANOVA-based approach, on the other hand, theoretical relationships are established between observable variables instead of between latent ones that are, in turn, operationalized by manifest indicators. In terms of the framework presented in Figure 1, commonly used ANOVA-based techniques are appropriate solely when the research designs include only dependent variables measured without error and will produce biased results otherwise, due to not taking into account measurement error and differential item reliabilities, as discussed earlier.

Intervening Variables

This second dimension of the framework shown in Figure 1 considers whether the research design under consideration includes intervening variables between the manipulation or between-group indicator and the ultimate dependent variables of interest. These may arise due to the inclusion of a manipulated variable or state, represented by manipulation checks, or due to a research design that includes more than one stage in a causal chain (that is, a system of mediating and dependent variables), or any combination of the two. In both cases, and even if the ultimate dependent variable is not latent in nature, SEM-LV techniques are most appropriate for analysis. In addition, there are a number of benefits associated with employing these techniques in the more sophisticated tests that can be conducted and in better distinguishing between the effects of the manipulation on the intervening variable and the effect of this variable on others in the research model.

In the first scenario above, a researcher introduces a manipulation, by means of assignment to a treatment group, to a group of participants in which said manipulation is intended to affect some underlying conceptual variable that is not directly observable—the hypothesis under consideration would be that the manipulation would influence the dependent variable of interest by means of impacting this unobservable intervening variable. In this case, manipulation checks are employed to verify that the manipulations have indeed been successful in achieving the desired effects on the intervening variable or psychological state. The rationale for their value lies in providing evidence that a change was effected in the subject such that observed effects in the dependent variable can be attributed to the manipulation or treatment, thus allowing for a stronger inference of causality.

When traditional techniques are employed researchers can only analyze whether the manipulation checks are themselves reliable and the extent to which they differ by treatment or manipulation. In contrast, SEM-LV allows for the inclusion and modeling of manipulation checks together with the manipulation and ultimate dependent variable of interest. The key issue here is how those manipulation checks should be included in the model, which requires researchers to consider what those checks are actually representing. To the extent that the value of including manipulation checks in a research design lies in showing that the manipulation successfully affected the intended target, it seems reasonable to include manipulation checks as indicators of the unobservable state that is the object of the manipulation. A sample design is shown in Figure 2 and discussed next. Note, however, that while the ultimate dependent variable of interest here is depicted as latent, the same logic would apply if it were directly observable.

In this context, a manipulation or intervention is said to have an effect on some psychological state or understanding in the subject, which, in turn, influences some dependent variable of interest. SEM-LV makes it possible to include manipulation checks in a comprehensive causal model, as well as require the researcher to make explicit (a) the assumptions underlying the research model and (b) how the effects are causally transmitted from the manipulation to the dependent variable of interest. In Figure 2 above, the independent dummy variable X represents the assignment to either the treatment or control group—the manipulation. In turn, this is expected to cause an effect

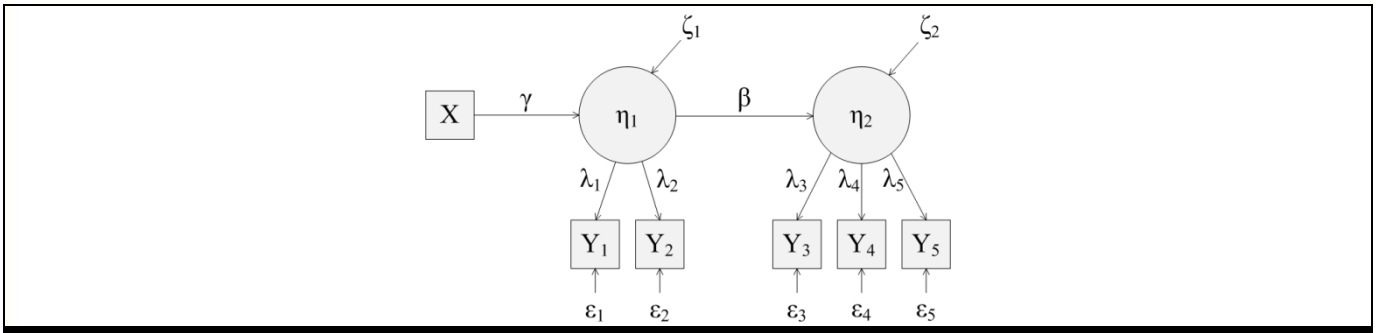

Figure 2. Causal Model Relating an Independent Variable and Manipulation Checks to a Single Dependent Variable

on an intervening and unobservable variable η_1, represented by two manifest indicators, which are what would traditionally be considered manipulation checks. The effect of X on η_1 is measured by the regression coefficient γ. There are a number of advantages in this approach.

First, SEM-LV takes measurement error into account not only for the dependent variable of interest but also for the manipulation checks—the advantages of this have been discussed in the preceding section. Second, researchers are thus able to gauge the effectiveness of the manipulation by means of the γ coefficient, which represents the impact of the manipulation or treatment on the intervening variable of interest. The second regression coefficient, β, depicts the effect of the manipulated state η_1 on η_2. Thus, the overall impact of the manipulation or group assignment is a function of two separate effects: the effectiveness of the manipulation in causing the desired psychological state or understanding (γ) and the causal relationship between that state and the dependent variable (β). It should be noted that it is the second of these relationships that is of theoretical interest.

The second scenario where intervening variables may arise, which may occur in conjunction with the one just discussed, occurs when researchers propose a research model that includes both between-group comparisons as well as extended systems of related variables. Whereas ANOVA and regression techniques can model only one stage in a causal chain at a time, SEM-LV allows researchers to model a system of dependent variables in a single statistical analysis. Doing so allows researchers to move away from a piecemeal analysis—where the between-group effects are first tested with ANOVA and then the path model separately with a different technique—and conduct an integrated test of the research model, one that can also include tests of relationships between the manipulation and constructs other than those hypothesized to be directly affected by the group assignment. Testing for the existence of these relationships is important because of the potential presence of bias due to omitted variables [Judd and Kenny, 1981; Mauro, 1990], which occurs when a variable that is both a cause of the ultimate dependent variable and of the mediating variable is omitted from a regression containing only the mediating and outcome variables. In this case, the resulting regression coefficient of the mediator on the ultimate outcome variable will be biased, whether upwards or downwards would be a function of the signs of the regression coefficients and correlations involved. When conducting separate analyses for each of these relationships, however, this cannot be accomplished. Both of these issues can be addressed by modeling these relationships using SEM-LV, which allows for all these relationships to be tested in a single, integrated analysis, comprising both measurement and structural aspects of the research model under examination.

Current State of Between-groups IS Research

Through an application of this framework to extant IS research, we seek to highlight the mismatch between research designs and analytical techniques that results from the widespread usage of traditional ANOVA-based techniques across all four quadrants of our framework, while we argue that those are most appropriate in only one of them—when neither measurement error nor intervening variables are part of the research design. In order to ascertain the degree to which this mismatch exists, a review of published research in the 2006–2010 period in a group of five premier journals (*MIS Quarterly*, *Information Systems Research*, the *European Journal of Information Systems*, the *Journal of the Association for Information Systems*, and the *Journal of Management Information Systems*) was conducted. A keyword search using relevant terms[1] was first executed, and the results were manually reviewed to identify any examples of between-group comparisons in these journals and time period.

[1] Any article containing one of more of the following terms was manually reviewed to determine its relevance to this research: *experiment, experimental, manipulation, ANOVA, ANCOVA, MANOVA, MANCOVA, random assignment, randomly assigned, manipulation check,* and *control group*.

Each article was coded along the two dimensions of the framework previously discussed: (1) whether the dependent variables were measured with error and (2) whether there were intervening variables in the research design. In this last case, a research design where manipulation checks representing an unobservable state or understanding were employed was considered to implicitly include an intervening variable—hence the need for manipulation checks to verify that the intervening state had been successfully manipulated. There were also cases where a manipulation was effected but no checks were deemed necessary, and those instances were coded as not including an intervening variable. This most commonly occurred in those research designs were the tasks to be completed by the participants were themselves the manipulation, for example, requiring participants to write queries against data structures varying in their level of ontological expressiveness to understand the degree to which query correctness, time taken to complete the exercise and confidence in the solution are affected as a result [Bowen, O'Farrell, and Rhode, 2009]. In this case, the manipulation was not intended to effect a change in an unobservable intervening variable, but rather to change the nature of the task itself. Designs where at least one dependent variable was measured with error were coded as including measurement error; even if both directly observable and latent variables were included in the research design, the presence of at least one of the latter would make SEM-LV techniques necessary.

After obtaining an accurate count of studies in the reviewed period and journals in each cell of the framework, the next step involved comparing the statistical techniques employed in the reviewed research against those that the framework in Figure 1 would dictate. As traditional ANOVA-based techniques are by far the most commonly used in all scenarios, this helps highlight areas where IS research practice can be improved by considering the use of SEM-LV methods when conducting between-group comparisons.

		Intervening variables between manipulation or between-groups indicator and ultimate dependent variable	
		N	**Y**
Measurement error in any dependent variable	**N**	Quadrant 1: ANOVA Techniques – Analyzed with ANOVA: 29 – Analyzed with SEM-LV: 0	Quadrant 3: SEM-LV – Analyzed with ANOVA: 2 – Analyzed with SEM-LV: 0
	Y	Quadrant 2: SEM-LV – Analyzed with ANOVA: 12 – Analyzed with SEM-LV: 0	Quadrant 4: SEM-LV – Analyzed with ANOVA: 41 – Analyzed with SEM-LV: 4

Figure 3. Results of 2006–2010 Review of IS Research

A total of eighty-eight studies were identified and coded according to the procedure described above. As the results included in Figure 3 show, the majority of reviewed studies (fifty-five out of a total of eighty-eight) employed analytical techniques that were not the most appropriate, given the characteristics of their research designs. In particular, all these used first-generation statistical techniques, with ANOVA, MANOVA, and ANCOVA being the most common. As discussed elsewhere, these techniques operate on observable (manifest) variables that are presumed to be measured without error. Given that in most of these cases (fifty-three out of fifty-five) the research design contained one or more dependent variables that were imperfectly measured, results obtained from these analyses would exhibit bias compared to the underlying population values (with the magnitude of this bias being dependent on the reliability of the composites employed as dependent variables). In addition, a number of these studies (forty-three out of fifty-five) included intervening variables between the manipulation and the dependent variables of interest. Single stage techniques, such as ANOVA or regression, cannot model chains of causality in that manner. As well, a number of these explicitly modeled a path system of variables that was tested in a separate step, where the SEM-LV discussed here can accommodate those in a single, integrated statistical analysis.

The goal of this review has been to highlight the mismatch that currently exists in the IS literature that employs between-group comparisons in terms of research designs and analytical techniques. Though the reviewed journals and time period surely represent but a small fraction of all relevant research, these journals are well known for their emphasis on methodological rigor. Thus, though we can only speculate about any other non-reviewed literature, the

situation is likely to be at least as problematic elsewhere. We believe these results help underscore the importance of attending to the choice of most appropriate statistical technique based on the research design proposed.

In what follows, the proposed techniques are described in more detail in order to foster their usage with IS researchers who are interested in between-group comparisons for research designs where these techniques would be appropriate—which we would argue encompass the majority of between-group research designs in the IS literature. The two basic approaches to specifying these designs within the SEM-LV framework—group code and structured means—are first discussed and compared to the traditional approach based on the ANOVA family of techniques. In their simplest form, these two alternatives would be appropriate for those research designs where dependent variables are measured with error, but no intervening variables (or systems of causal linkages) are involved. Then the use of manipulation checks and how to incorporate them into each alternative is covered. At each step, we also discuss how results based on ANOVA and SEM-LV analyses would differ, which underscores the additional value that can be obtained from the application of SEM-LV to IS research.

III. BASIC APPROACHES TO SEM-LV ANALYSIS OF BETWEEN-GROUP DESIGNS

We introduce the two existing approaches to the estimation of mean differences in latent variables using a simple two group example. Standard SEM notation will be used throughout this article, and basic knowledge of confirmatory factor analysis is assumed. In this scenario, a researcher is interested in making an inference as to whether there is a difference in mean levels of a construct η between two populations (or treatments, or experimental manipulations) of interest, with this construct being represented by three observed indicators Y_1, Y_2, and Y_3, and the first indicator being used to set the scale of the construct by fixing its loading to one. In what we called the traditional approach, the researcher would sum the observed scores for all three indicators, calculate a reliability coefficient and, if satisfactory, subject the sum of indicators for each group to an analysis of variance. If significant, that would provide evidence supporting the existence of mean differences between subjects in the two groups. We describe below how this analysis would be conducted using two alternative approaches within the latent variable framework. The developments presented below draw heavily from previous work by Hancock [1997, 2003]; for a comprehensive treatment, including examples and software code, see also Thompson and Green [2006].

Group Code Approach

The group code approach is a special case of the more general class of MIMIC (Multiple Indicator Multiple Cause) models [Muthén, 1989] and is analogous to the representation of ANOVA models as special cases of multiple regression. Recall that a one-way ANOVA with two groups can be expressed as a regression model through the use of a dummy variable (coded, for instance, 1 for one group and 0 for the other). More generally, several variations of analysis of variance models can be expressed as regressions with the creative use of dummy variables to represent the presence or absence of multiple treatments, including interactions. For instance, a two-way ANOVA with two levels in each factor and an interaction can be expressed as:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 AB + \varepsilon$$

In the equation above, A and B are dummy variables taking values of 1 and 0, depending on subject membership on the different levels of the two factors involved, and their multiplication represents the interaction. In the group code approach introduced here and for the two-group scenario described above, a dummy variable, X, is introduced to the latent variable model, which takes values of either 1 or 0 depending on the presence or absence of treatments or membership in experimental and control groups. In order to estimate the relationship between group membership and the construct of interest, the data from both groups are combined and the construct is regressed on the dummy variable, as shown in Figure 4 below.

Assuming adequate fit to the data, the parameter of interest in the model depicted in Figure 4 is the regression coefficient γ, which represents the effect of group membership on mean levels of the construct η. The interpretation of this parameter also relies on the expression of between-group models in regression form. In particular, the relationship here can be expressed as $\eta = \gamma X + \zeta$ such that for those subjects with group code 1 it equals $\eta = \gamma + \zeta$ and for those with group code 0, $\eta = \zeta$. Since the expected mean of the residual term is 0, the expected means for the two groups are $\hat{\gamma}$ and 0, respectively. Therefore, $\hat{\gamma}$ represents the estimated difference between the two construct means. This parameter can be tested for significance using the ratio to its standard error, and the sign would indicate the direction of the difference (e.g. which group exhibits higher mean levels of the construct). Formally defined effect sizes to assess the magnitude of the estimate are discussed in more detail by Hancock [2001, 2003]. For two different approaches to sample size determination and statistical power in SEM-LV, in addition to the work just referenced, see Muthén and Muthén [2002] and Saris and Satorra [1993].

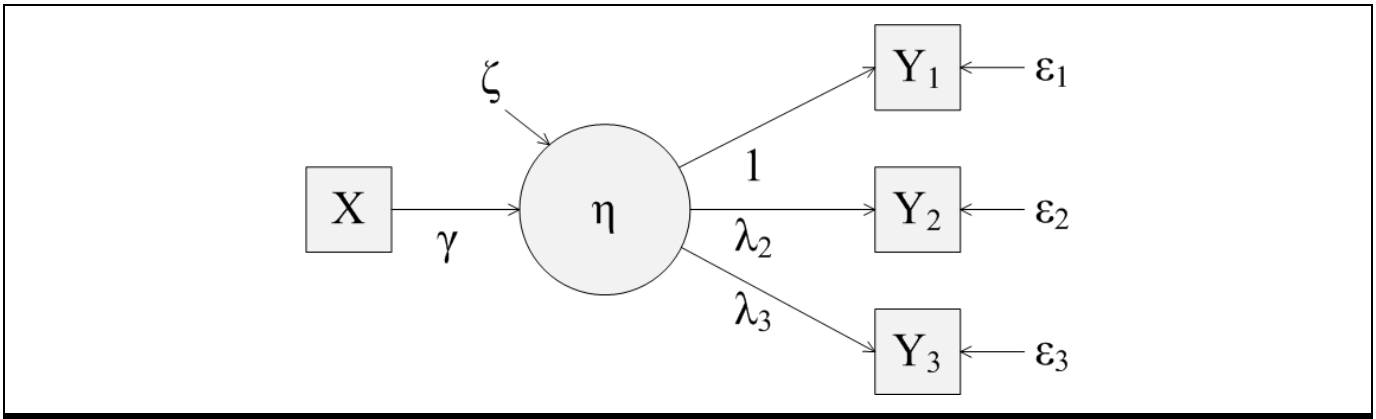

Figure 4. Group Code Analysis

Structured Means Approach

The majority of work conducted using SEM-LV, including the approach just discussed, is focused on the relationships among variables. In these cases, for both measurement and structural models, the means and intercepts of these variables, either latent or observed, are irrelevant, since the objective is to reproduce the matrix of observed variances and covariances. On the other hand, in the family of statistical procedures based on partitioning the variance of observed variables (e.g., ANOVA, ANCOVA, MANOVA, MANCOVA), differences among means are of prime importance. As we show here, however, SEM-LV can also be used to address questions about means by introducing intercepts as parameters to be modeled, in what has been called *structured means models* [Sörbom, 1978]. These models represent a special case within the more general framework of measurement invariance [Vandenberg and Lance, 2000].

Unlike the group code approach, where data from both samples are combined, when using structured means analysis, data from the different groups are kept separate, thus eliminating the need for a set of coding variables to differentiate between groups. Rather, equations involving construct means and indicator intercepts are employed. We illustrate this approach in the context of a simple two-group scenario, as was done before. Figure 5 depicts this alternative, together with the relevant equations involving construct means and indicator intercepts. Note the absence of a coding variable, as well as the equality constraints imposed: all three loadings are equivalent across both groups (1 for the first indicator, which sets the scale of the latent variable, and λ_2 and λ_3 for the second and third indicators, respectively), as are the intercepts for the indicators (τ_1, τ_2 and τ_3).

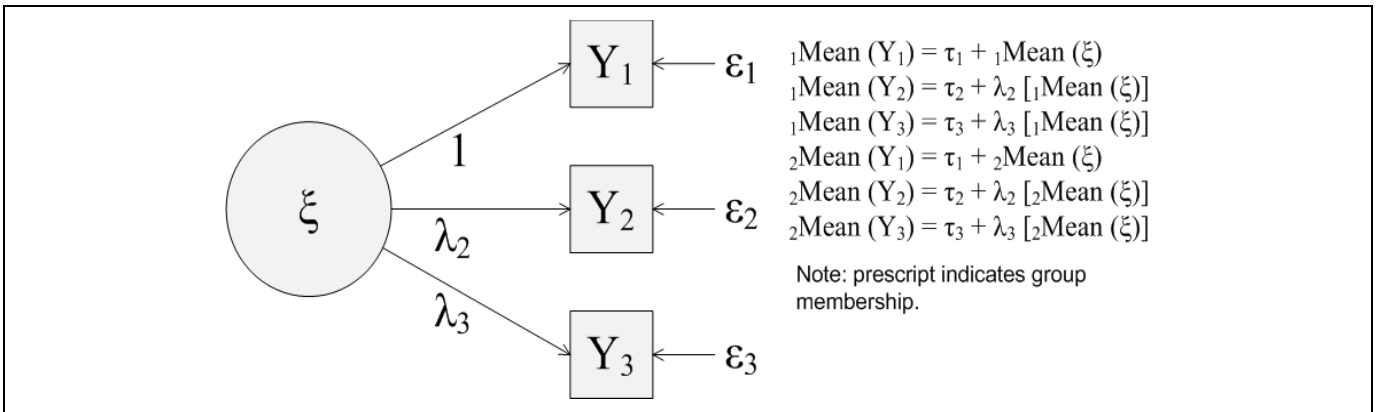

Figure 5. Structured Means Analysis

Structured means analysis makes inferences about groups by using information provided by means of observed variables. The underlying assumption is that if mean differences exist in the observed variables, these would be caused by mean differences in the latent ones. For this to be the case, however, a number of constraints must be imposed on the model in order to ensure that the structural relationship between the latent and observed variables is comparable across groups. Specifically, the loading and intercept terms should be constrained to be equivalent. It should be noted that these are the same constraints recommended by Vandenberg and Lance [2000] as steps in the assessment of measurement invariance. The adequacy of these constrains (i.e., whether they are supported by the data), is assessed by means of the significance in the difference between the chi-square statistics for the unconstrained and nested models.

If fit of the constrained model is not significantly worse (and assuming the overall model fits acceptably well, based on commonly used cut-offs, which is a requirement of all SEM-LV analyses), one more constraint is still necessary for the identification of the latent means and, consequently, the estimation of differences between them, if any. Since the goal is the estimation of the difference between latent means and not the means themselves, one of the latent means can be fixed to a set value, which does not affect the magnitude of the difference between them but allows for the identification of the mean structure of the model. Customarily, one latent mean is fixed to zero, and a test of the difference between the free latent mean and the fixed one is accomplished by the significance of the ratio between the estimate of the free mean and its standard error which, if standard statistical assumptions are met, is normally distributed. Alternatively, researchers could compare the significance of constraining both latent means to be equal compared to a model with a free latent mean.

Comparison Between SEM-LV Approaches and with ANOVA Techniques

Each of the two approaches has advantages and disadvantages and associated underlying assumptions [Hancock, 1997]. In general, structured means analysis is more complex, requiring a larger sample and the estimation of more parameters than the group code approach. The apparent simplicity of the latter, however, comes at the cost of making a number of simplifying assumptions and using constraints. On the other hand, these assumptions are made explicit when using structured means and their adequacy can be empirically tested.

The primary assumption underlying the group code approach is that the same measurement model applies to all groups included in the analysis. This is necessary since data are combined and a single model is tested, where dummy variables are used to code the differential effects of belonging to one group or another (note that researchers using ANOVA are implicitly making this assumption as well). In the case of structured means analysis, each group is modeled separately, and the assumption that measurement invariance holds can be tested. The invariance requirement necessary for the group code approach includes all sources of covariation (loadings, construct variance, error variances), in effect amounting to the assumption of equal variance/covariance matrices across groups. Some of these restrictions can be relaxed in the structured means model, although the sequence of increasingly more constrained tests of invariance proposed by Vandenberg and Lance [2000] does require full configural, metric (e.g., loadings), and scalar (e.g., intercepts) invariance before latent means can be compared. Given a sufficient sample size, there is no reason to prefer one alternative to the other. When constraints on the research setting do not make this possible, the group code approach allows researchers to limit the number of subjects required to obtain estimates, albeit at the expense of not being able to explicitly test the adequacy of the invariance assumptions underlying these techniques (on the other hand, if a model using the group code approach shows no significant misfit, this can be interpreted as validation of these assumptions).

In addition, given the lack of a direct path from the group code variable to the manifest indicators, the group code approach does not allow for differences in observed means due to group membership, which in effect implies that any individual item variation in means is due to a difference in the underlying latent variable—that is, an assumption of intercept invariance [Vandenberg and Lance, 2000]. In structured means analysis, this assumption is made explicit through the use of equality constrains on the intercepts of individual items. The assumption would also be subject to empirical testing.

Therefore, both approaches make assumptions about invariance, such that subjects with the same levels of the latent value would display the same values in the observed indicators. They differ, however, in the extent to which those assumptions are made explicit and subject to testing. Both approaches, on the other hand, are capable of accommodating specific violations of invariance. For example, a direct path from the group code variable to a specific indicator amounts to releasing intercept invariance for that specific item, which can also be accomplished through the removal of the equality of means constraint for that item in structured means analysis. The group code approach, however, cannot accommodate differential loadings. It should be noted that (a) which invariance assumptions are required for making inferences in different scenarios and (b) whether those can be made when invariance needs to be relaxed, are somewhat contested issues. The interested reader is referred to the work of Vandenberg [2002] and Vandenberg and Lance [2000] for a thorough treatment review, and also Byrne, Shavelson, and Muthén [1989] and Cole, Maxwell, Arvey, and Salas [1993] for discussions of partial measurement invariance.

To the extent that research designs call for latent dependent variables that are imperfectly measured by one or more manifest indicators, analysis of those designs using the traditional ANOVA-based techniques—which is largely the case in contemporary IS research practice, as previously discussed—will lead to estimates that are biased when compared to the population values of the corresponding parameters. This occurs because first-generation statistical techniques such as ANOVA assume the dependent variables are measured without error. We show this by means of two simple examples. Consider a two-group research design with a single dependent variable which is latent and measured with four manifest indicators—structurally similar to the models shown in Figures 4 and 5. The latent mean is 0 in one group and 0.5 in the other, with the variance equal to 1 in both groups—a medium effect size as *d*

= 0.5 [Cohen, 1988]. In this first example, we assume equal loadings across all indicators and between groups of 0.7, which leads to indicator means of 0 (latent mean of 0 multiplied by a loading of 0.7) for one group and 0.35 for the other (latent mean of 0.5 multiplied by a loading of 0.7). Error variances for the manifest indicators are set at 0.51, which gives each indicator unit variance.

Through these examples we work with population values and assume equal group sizes to leave aside the effects of sampling variability and focus on the effects that analysis with one technique or the other has on the results of interest. In the case of SEM-LV specifying the scenario above correctly will yield unbiased estimates, and thus researchers will obtain an accurate result that reflects the underlying medium-sized between-groups effect of d = 0.5. When analyzed using the traditional approach, the four indicators would be summed to obtain a single estimate of the dependent variable. The composite reliability for the sumscore would be 0.79, which would be considered acceptable in light of the commonly used cutoff of 0.70. The variance of the sumscores in both groups would be 9.88 (the sum of all the elements in the covariance matrix of the four indicators, which has 1 in the diagonals and 0.72 or 0.49 in the off-diagonal elements), which results in a standard deviation of 3.143. The sumscore means would be 0 in one group and 1.4 (0.35 $\times$ 4 indicators) in the other.

In this case, the estimate of d for the differences between sumscores would be 0.445 (or 1.4/3.143), which is different from the underlying population difference of 0.5 by more than 10 percent. The difference between the population difference and estimated difference by traditional means varies with the reliability of the indicators (for example, had the indicators been less reliable at 0.50 loadings, the estimated d would have been 0.378, almost 25 percent off the population value). The simplifying assumption of equal loadings across all indicators in each group does not affect our results. Following the same calculations as just performed, a population with loadings of 0.8, 0.7, 0.6, and 0.5 for the four indicators would have resulted in an estimate of d of 0.433 following the traditional analysis, which is biased against the population value of 0.5 as well. Thus, in research designs where no intervening variables are involved, but where the dependent variables are measured with error, the use of traditional analytical approaches based on the ANOVA family of techniques will result in estimates of between-group differences that are biased downward due to the negative effects of measurement error. This is a well-known difference between first generation techniques and those based on latent variables; however, current practice in IS research employs ANOVA analyses throughout.

IV. INCORPORATING MANIPULATION CHECKS AND INTERVENING VARIABLES

This section builds on the two basic approaches to the specification of between-group research designs using SEM-LV techniques by incorporating intervening variables between the manipulation or between-group indicator and the ultimate dependent variables of interest. In what follows, the case of incorporating manipulation checks into a comprehensive research model will be discussed, but the same underlying rationale applies to models where a system of causal linkages follows the manipulation. The advantages of doing so—allowing for measurement error in manipulation checks, representing those in a manner that is more consistent with the notion of a manipulated intervening variable or state, and better distinguishing between manipulation strength and other theoretical relationships—have already been discussed. In this section we focus on the specification of these in either of the two approaches just presented and on their interpretation, and highlight with examples how these improve upon traditional ANOVA techniques.

Incorporating manipulation checks into the basic designs previously discussed—group code and structured means approaches—is straightforward and essentially entails adding a new latent variable that stands between the manipulation and the dependent variable of interest, whether latent or observed. Alternative ways of specifying the analytical model are shown in Figure 6, where the indicators Y_1 and Y_2 would be what are commonly known as manipulation checks.

In the models shown in Figure 6a,c, which have been specified using the group code approach, the gamma parameter estimates the strength of the manipulation to effect a change on the intervening variable, while the beta parameter estimates the relationship of interest. A significant estimate for the gamma parameter would be indicative of a successful manipulation (and the magnitude of the estimate an indication of the effect size), whereas a significant beta parameter is an indication that there is a relationship between the two latent variables of interest. In the group code approach, researchers are making the assumption that the beta parameter is the same across groups, which would be theoretically expected: while the manipulation should affect the intervening variable, it should not affect the underlying relationship between the latent variables.

In the structured means approach, as in Figure 6b,d, the effects of the manipulation would be observed by comparing the latent means of the independent variable across groups, subject to all the required measurement invariance tests necessary to conduct such a test [Vandenberg and Lance, 2000]. Then, the gamma parameter

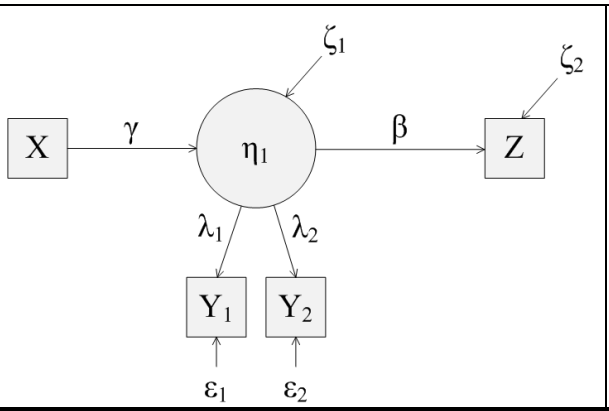

Figure 6a. Manifest Dependent Variable and Manipulation (Group Code)

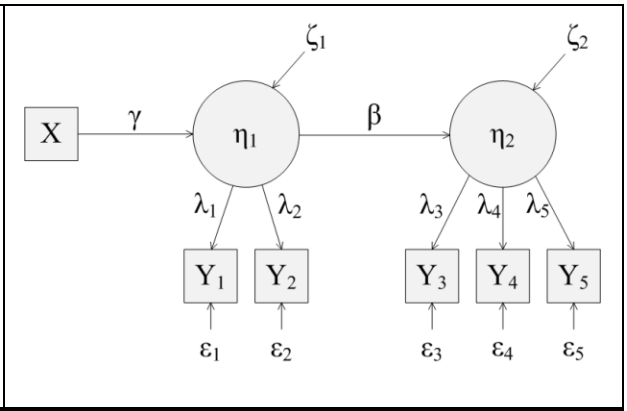

Figure 6c. Latent Dependent Variable and Manipulation (Group Code)

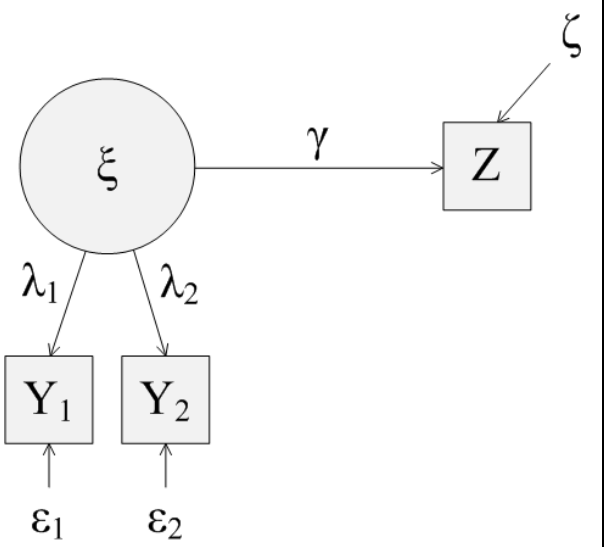

Figure 6b. Manifest Dependent Variable and Manipulation (Structured Means)

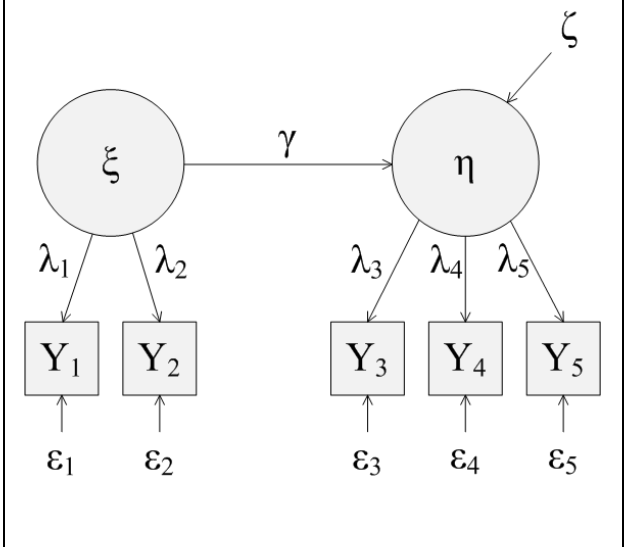

Figure 6d. Latent Dependent Variable and Manipulation (Structured Means)

relating the independent and dependent variables would estimate the theoretical relationship of interest between them. By using structured means to conduct the analyses, researchers can verify that the relationship between the two latent variables is indeed the same across groups. It should be noted that both the group code and structured means approaches work similarly, regardless of the nature of the dependent variable—observed or latent.

We now show how the traditional ANOVA analyses confound the strength of the manipulation and the theoretical relationship of interest between latent variables into a single estimate. We do so by means of two examples, the first with a directly observable dependent variable, which helps highlights the limitations of ANOVA with regards to distinguishing between manipulation effects and the causal relationship between variables without involving error in the dependent variable. In the second example, both occurrences are shown, thus directly comparing the proposed SEM-LV techniques with ANOVA-based ones in what is the most common type of research design in contemporary IS research.

Consider a two-group experimental design with one intervening latent variable, which is the object of the manipulation, and a single dependent variable that is directly observable—that is, measured without error. The effectiveness of the manipulation is assessed by means of two items. This case is similar in structure to the ones shown in Figure 6a,b. A medium effect size (e.g., Cohen's d = 0.5) would lead to the means of the independent latent variable subject to manipulation to be 0 and 0.5 in each group, with an assumed standard deviation of 1. The loadings for the two manipulation checks would be 0.8, which would lead to their intercepts being 0 in one group and 0.40 (0.5 latent mean in the manipulated group multiplied by a loading of 0.80). Assume that the path coefficient between the independent latent variable and the observable dependent variable is 0.4, and that the dependent variable has a standard deviation of 1 as well. This would lead to the mean of the dependent variable to be 0 in the control group and 0.2 (mean of the independent variable of 0.5 in the manipulated group multiplied by the path coefficient of 0.4 relating the two variables) in the experimental group. This example works using population values

Communications of the Association for Information Systems

to further filter out the effects of sampling variability in the results. When correctly specified, an SEM-LV analysis will accurately recover the population parameters. The analysis based on the traditional ANOVA approach would work as follows.

First, the effectiveness of the manipulation would be assessed by comparing the means of the manipulation checks across the two groups. Even with relatively reliable items, the analysis based on the observed individual scores would indicate a *d* equal to 0.40, which underestimates the strength of the manipulation by 20 percent when compared to a population *d* of 0.50. Most of the time, however, the focus is on whether they are significantly different across groups and not on the magnitude of the difference. Next, researchers would compare the means of the dependent variable between the two groups, which, being perfectly measured, would lead to an unbiased estimate of d = 0.20. When working within the SEM-LV framework, the manipulation effect of d = 0.50 is distinguished from the path coefficient of 0.40 relating the two variables of interest; multiplying the *d* of 0.50 times the path coefficient of 0.40 leads to the observable difference of *d* = 0.20 in the dependent variable. Researchers working under the ANOVA approach would have observed a manipulation effect of *d* = 0.40 and an overall effect on the dependent variable of 0.20, which would have led them to conclude that the path coefficient relating the two variables should equal 0.5 (or 0.2/0.4), which would overestimate the magnitude of the relationship between the two variables, which is typically the main estimate of interest (in this case by 25 percent).

The situation becomes even more dire when the dependent variable is latent in nature, e.g., similar to the models shown in Figure 6c,d. Assume the same effects as before, but a latent dependent variable represented by three indicators, all with loadings of 0.8 and unit variance (Y_3, Y_4, and Y_5 in Figure 6). The traditional analysis on the individual observed manipulation checks would lead researchers to an effect size of *d* = 0.40 for the manipulation. The sumscore of the three manifest indicators of the dependent latent variable would have means of 0 in one group and 0.48 (0.2 latent mean in the manipulated group multiplied by a loading of 0.80 and summed over the three indicators), with a variance in both groups of 6.84 (standard deviation of 2.615). A comparison of these sumscores would lead to a *d* of 0.1835 (or a difference in means of 0.48 divided by a pooled standard deviation of 2.615), which would underestimate the population effect size—as discussed before when considering the effects of measurement error when the traditional approach is employed. When estimating the path relating the two latent variables, researchers would arrive at an estimate of 0.459 (or 0.1835 divided by 0.4), which again overestimates the population coefficient of 0.4. Furthermore, in this scenario—which is the most common one in contemporary IS research—researchers working under the traditional approach would be left with estimates for all three coefficients of interest (manipulation strength, relationship between variables, and overall effect) that are biased when compared to their population values.

V. MORE COMPLEX EXAMPLES

The description of the two general approaches to SEM-LV modeling of between-group comparisons above was couched in terms of the simplest model possible, the two-level one-way ANOVA, for ease of exposition. This should not be taken to mean, however, that the techniques described here can only be applied to simple models. Quite to the contrary, many other more sophisticated designs can be analyzed, some of which have been described by Mackenzie [2001]. This section shows how to specify more advanced models using the SEM-LV approach. All examples shown in this section are structurally similar to recently published IS research in major journals. The goal of this section is to show how sample cases from contemporary research practice can be specified using the SEM-LV approach proposed here in order to further foster its applicability to future studies.

Two Groups, Multiple Dependent Constructs

The first scenario shows the case of multiple, correlated dependent constructs, each measured with multiple items, compared across two groups of interest. Examples of IS research employing a similar design include Wakefield and Whitten [2006], Tiwana and Keil [2009], or Shanks, Tansley, Nuredini, Tobin, and Weber [2008]. This example also helps highlight the many advantages of the latent variable approach over the commonly used MANOVA alternative for cases with more than one dependent variable. The general approach of running a MANOVA when there is more than one dependent variable and following with univariate ANOVA analyses if the test of the MANOVA omnibus hypothesis is significant, is quite common in published research. There are a number of issues with this approach, however, that makes it less than appropriate in practice. First, when running a MANOVA analysis, each of the dependent variables would be represented by the sum of their observed indicators, therefore not accounting for the effects of measurement error. Second, the MANOVA procedure works on a linear composite of the observed variables, and not on the observed variables themselves; MANOVA creates a composite using the observed scores and thus statistical tests are not performed on the same constructs as would be conceptualized in SEM [Hancock, 2003; Hancock, Lawrence, and Nevitt, 2000]. A significant result in a MANOVA analysis means that a weighted combination of the composite variables included in the analysis exhibits mean differences across groups. Third, a MANOVA analysis of scenarios when intervening latent variables, such as those affected by a manipulation, need to

be included in the research design will result in biased estimates of both the strength of the manipulation as well as of the relationship between the manipulated perception and the dependent variables of interest. Finally, MANOVA would be more appropriate for evaluating differences in means in an emergent (e.g., formative, where causality flows from indicators to constructs, but without including a disturbance term, therefore not dealing with a latent variable) rather than a latent (e.g., reflective) variable system [Thompson and Green, 2006]. This is an important but not widely known distinction.

Using the SEM-LV approach, and ignoring for a moment the inclusion of manipulation checks, we can conceptualize this research model in a way that is consistent with the use of multiple-indicator measures. We are including here three correlated constructs (A, B, and C), each measured with multiple indicators (4, 6, and 3, respectively), but the model would be similar for any number of constructs. Following Kano [2000], Figure 7 shows the latent variable specification using the group code approach, whereas Figure 8 shows the same research model using the structured means alternative (the constraint equations shown in Figure 5 are not included in this section for ease of exposition, but can be similarly derived). As discussed before, when all sources of variation are equivalent across groups, both specifications are identical and will yield the same results.

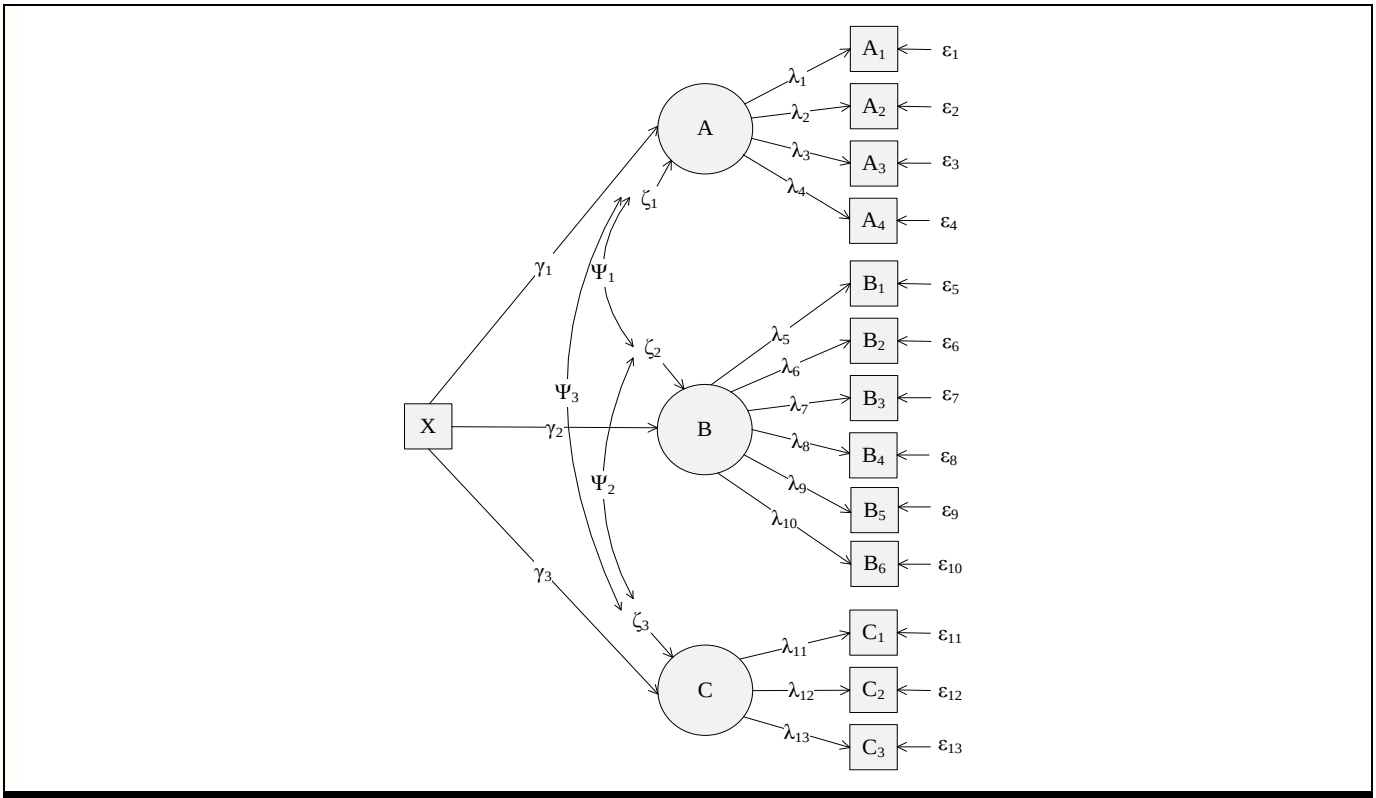

Figure 7. Two Groups, Multiple Dependent Constructs Group Code Approach

If researchers wanted to include—as they likely should—items to verify that the manipulation had indeed successfully effected the intended change in the perceived context in which the participants answered questions about the three dependent variables of interest, those could be added to the models shown in Figures 8 and 9 by means of incorporating a new latent variable, represented by the three manipulation checks, as an intervening variable between the group indicator X and A, B, and C in Figure 7 or as an antecedent to A, B, and C in Figure 8. An assessment of the manipulation could then be conducted by assessing the significance of the path from group indicator to intervening variable, if using the group code approach, or by comparing the latent means of the manipulated independent variable, if using the structured means approach. In both cases, the relationship between this variable and the dependent ones of interest (here A, B, and C) would be assessed by analyzing the appropriate paths linking pairs of latent variables.

Two Factors, Interaction, Observed and Latent Dependent Variables

The next example shows an experiment in which two different design factors are manipulated, with two levels each, and participants are randomly assigned to each of the resulting four conditions. The goal of this design is to examine both main effects and their interactions on the dependent variables of interest. Of note here is the fact that, while

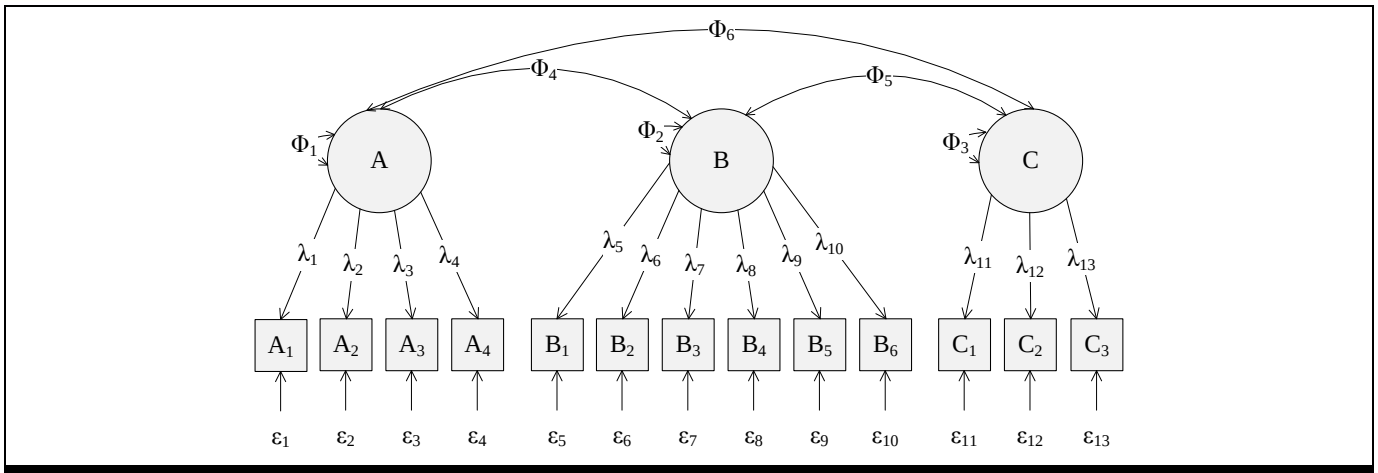

Figure 8. Two Groups, Multiple Dependent Constructs Structured Means Approach

some of those dependent variables may be multi-item constructs, others can be observed variables, with no assumption of measurement error (for example, time taken to complete a task). We take advantage of this research design to highlight the flexibility of the SEM-LV approaches described here to accommodate, in a single model, both observed and latent dependent variables and the analysis of both main effects and interactions in an experimental design. For ease of exposition, only one of each type—latent and observed—are discussed here, but the design can be extended to the case of more than two dependent variables in a straightforward manner. Also of note in this research design is the fact that the manipulations, while expected to affect the dependent variables of interest, are not posited to influence an unobservable state or perception in the respondents. In the framework shown at the beginning, this study would be classified in Quadrant 2 (measurement error in one or more dependent variables, but no intervening or manipulated variable). Recent studies with a similar design include Hong, Thong, and Tam [2005], Qiu and Benbasat [2009], or Balijepally, Mahapatra, Nerur, and Price [2009].

Figures 9 and 10 show how to specify this research model using the group code and structured means approaches. Two dependent variables are included, one modeled as a latent variable measured by six items (X) and the other modeled as a single observed variable (Z). Note that, for clarity, the covariance between disturbance terms is not shown in Figure 9, but would be included in the model to account for sources of covariation other than the experimental manipulation. The two experimental factors are labeled A and B.

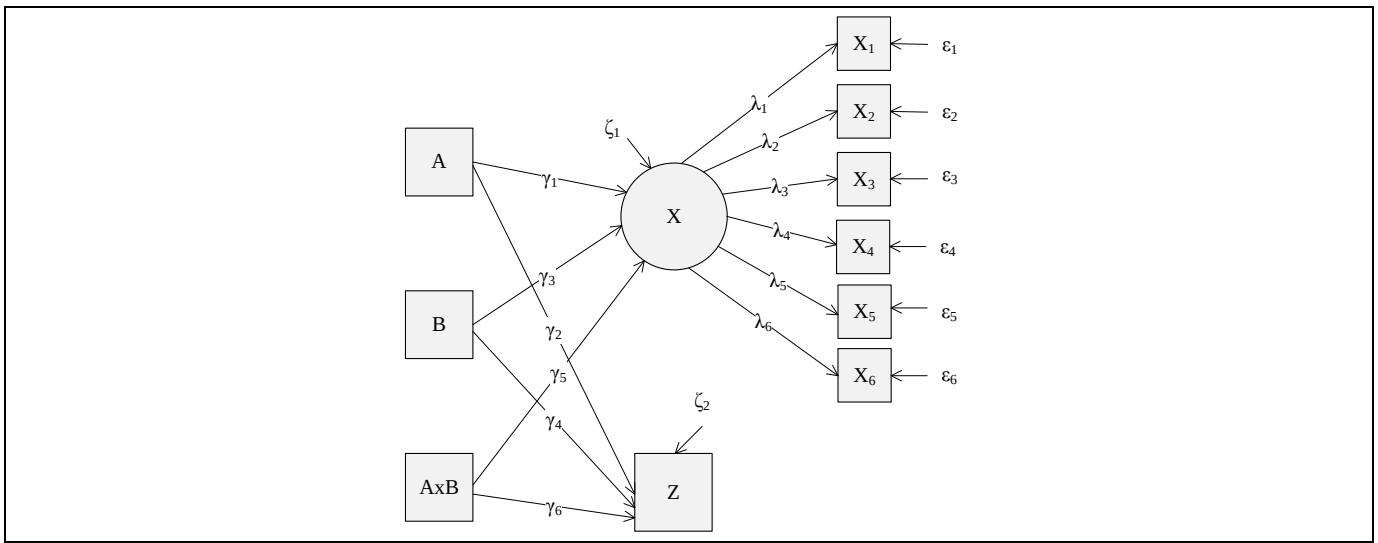

Figure 9. Two Factors, Interaction, Observed, and Latent Dependent Variables

In the group code approach, shown in Figure 9, the main and interaction effects of the experimental treatments are depicted by the paths going from the dummy coded variables representing group assignment to both dependent variables. As discussed before, the residual terms of the two dependent variables are allowed to correlate to account for any relationship between them that is not the product of experimental assignment. If strict independence were expected (e.g., the dependent variables should not correlate conditional on group assignment), this could be

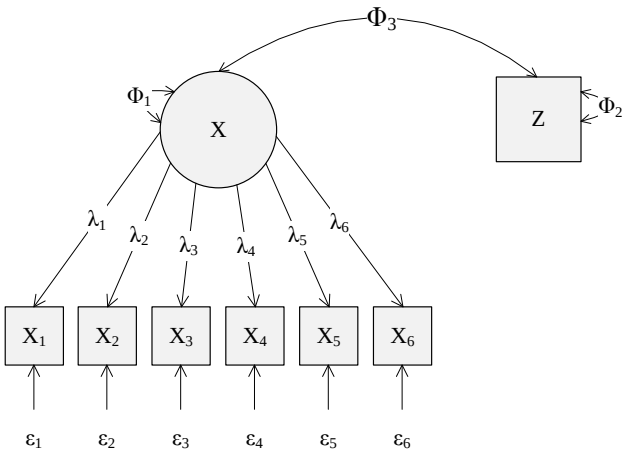

Figure 10. Two Factors, Interaction, Observed, and Latent Dependent Variables

tested by estimating the difference in fit from a model with a freely estimated relationship and one where the correlation between residuals has been fixed at zero. The same testing logic applies to any of the paths of interest in this model. For example, an omnibus test of no overall effects would compare fit between a model with freely estimated regression paths and one where all have been fixed to zero. Specific paths of interest (for example, those showing the effect of the interaction between the two conditions on each dependent variable) can be tested in a similar manner.[2]

Researchers familiar with the regression formulation of ANOVA models described above will find the testing and interpretation of these paths similar in nature. The estimation and testing of interaction effects using the structured means approach is more involved and requires the use of more complex cross-group constraints. In order to best understand the rationale behind these constraints, a brief review of the workings of a 2×2 research design is necessary. This scenario is treated here in some detail, as testing of interactions is quite common in multi-factor experimental designs. Figure 11 shows such an example.

Factor A	Factor B: Low (L)	Factor B: High (H)
Low (L)	LALB	LAHB
High (H)	HALB	HAHB

Figure 11. 2×2 Experimental Design

When employing a 2×2 design researchers are interested in comparing the means of certain variables of interest for each separate group, and whether these are different across groups. In a 2×2 design there are eight possible scenarios for results, ranging from no effects of any kind to interaction effects accompanied by simple main effects of each factor. The common practice is to assess the presence of an interaction effect first, and only then the presence of simple effects within each factor (if there is an interaction) or main effects of each factor (if there is no interaction). In this type of design, there is support for an interaction effect when the difference between the group means within each level of a factor varies depending on which level is considered. In terms of the example shown in Figure 11, if LALB – LAHB is different from HALB – HAHB, then there is an interaction between the two factors. When using the structured means approach (e.g., Jaccard and Wan, 1996) this is accomplished by conducting a test

[2] Researchers often evaluate the significance of a finding by examining the z value associated with a particular path of interest. Recent research by Gonzalez and Griffin [2001], however, raises the issue of the sensitivity of standard errors to the choice of indicator fixed for identification purposes. The likelihood ratio test, on the other hand, appears immune to this problem and is thus recommended for significance testing.

Communications of the Association for Information Systems

of the significance of constraining the means across these two groups to be identical, after all other necessary invariance constraints have been specified. Depending on the significance of this test, either simple or main effects would be assessed next. Figure 12 shows the sequence of these tests in terms of the example design shown in Figure 11 above. It should be noted that these tables can be easily constructed from the results of the unconstrained model where all latent means are allowed to be freely estimated, save for the one that is fixed to zero for identification purposes. In this case, all other latent means are expressed as deviations from the one used as a reference, but that does not impact the magnitude of the differences between groups themselves. The sequence of tests shown in Figure 12 is analogous to that commonly used when analyzing 2×2 ANOVA models. All these tests are performed by including constraints in the model to force the means of the desired groups (or their average, in the case of main effects) to be equal, and comparing the resulting fit with that of an unconstrained model (e.g., a chi-square difference test).

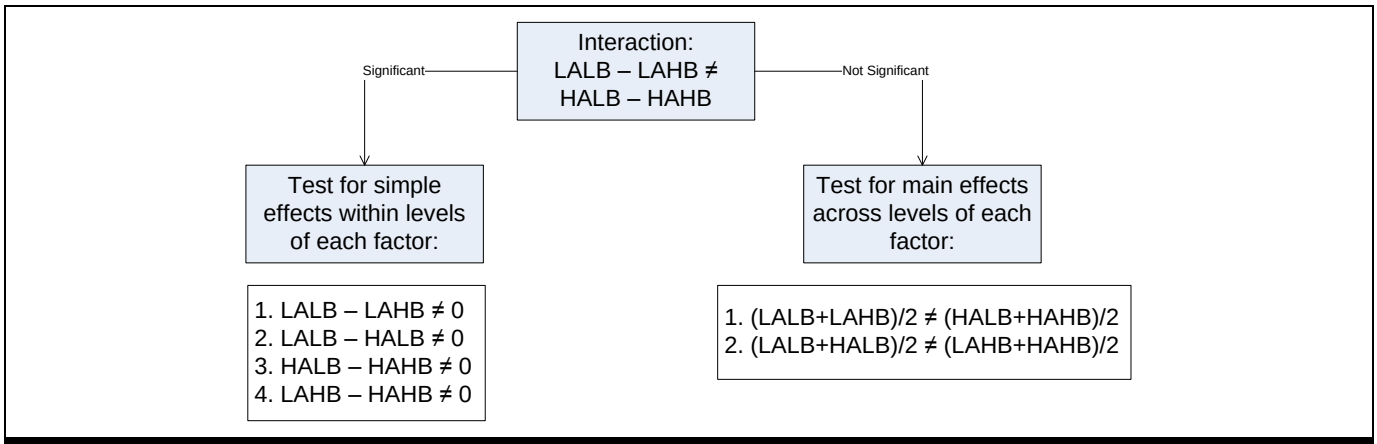

Figure 12. Sequence of Tests for a 2×2 Design

The latent variable approaches described here can also accommodate extensions of the basic 2×2 model just discussed. Consider, for example, a replication and extension of this scenario in which a researcher is interested in examining these effects after accounting for (or controlling) a covariate, measured in this example as a latent variable (items not shown for clarity)—this is conceptually similar to ANCOVA or MANCOVA analyses. Figures 13 and 14 show this extension using the group code and structured means approaches. Two sets of parameters are of particular interest in Figure 13, in addition to those already discussed in the more basic model above. First, those paths from the latent covariate to the two dependent variables serve to control for the effects of this covariate so that the effects of the experimental treatments on the two dependent variables can be more precisely assessed. Second, the relationship between the latent covariate (an observed covariate could be easily accommodated as well) and the dummy variables representing group assignment provides researchers with a testable assessment of the effectiveness of random assignment to experimental groups in controlling for other individual differences. A significant correlation between the latent covariate and the dummy variables would indicate there is a significant difference on the levels of the covariate between some of the experimental groups, which should be taken into account when interpreting the results.

Because the group code approach assumes complete equivalence between the two groups, the effects of the latent covariate on the dependent variables are assumed to be equivalent across groups as well. The structured means formulation of this research model allows the testing of this assumption before proceeding with testing for the equivalence across means of the dependent variables in the different groups by assessing the effects on fit of a cross-groups equivalence constraint on the path from the latent covariate to each dependent variable. The adequacy assumption of equivalent variances that is required before this test can be performed (cf. Vandenberg, 2002) can also be assessed in a similar manner. Finally, the effectiveness of random assignment can be tested by comparing the latent means of the latent covariate across the four experimental groups. As just discussed, finding a significantly different mean indicates the presumed control of extraneous variables by random assignment to experimental conditions has been compromised. As with any other statistical test, findings of non-significance (which would indicate appropriate random assignment) need to be evaluated in light of the statistical power of the test.

Pre- and Post-test Designs

The last example shows how to model and analyze the commonly used pre- and post-test experimental design. In this scenario, one is interested in understanding the effects of an experimental intervention on a latent variable of

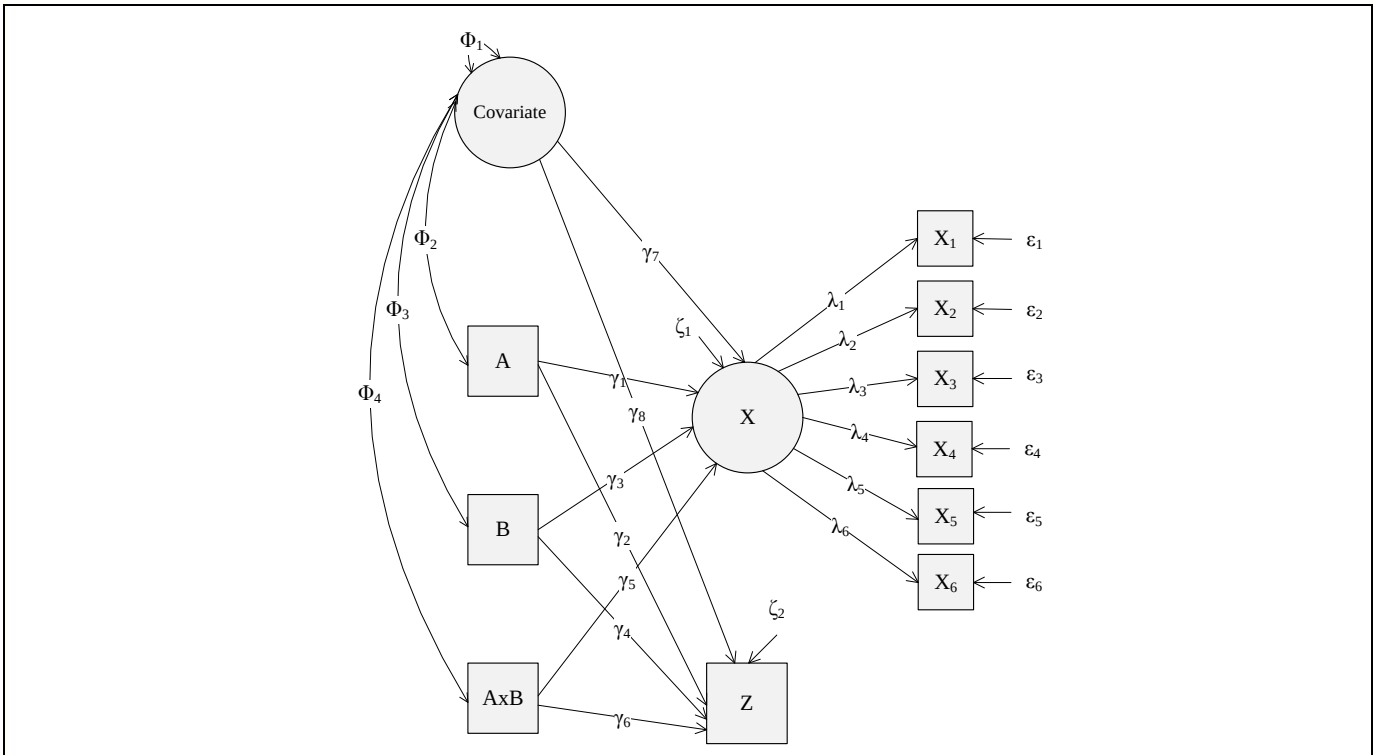

Figure 13. Addition of a Latent Covariate Group Code Approach

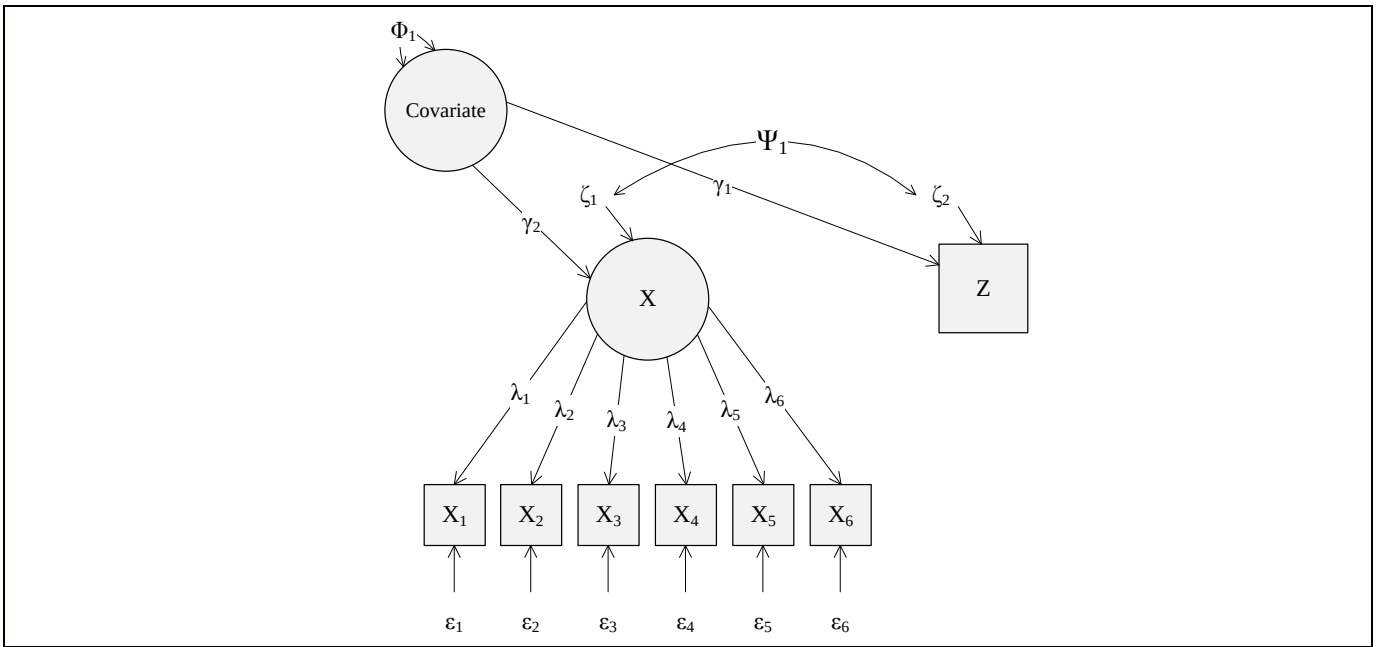

Figure 14. Addition of a Latent Covariate Structured Means Approach

interest while taking into account the pre-manipulation levels of the same variable in the participant, where the focal construct is measured with the same items on both occasions, as would commonly be the case. For ease of exposition, only three indicators are included. See Piccoli and Ives [2003] for empirical research employing a similar design.

Figure 15 shows the specification of the research described here through the use of the group-code alternative. The figure shows the latent construct of interest represented as a latent variable in both measurement occasions. The parameter γ_1 represents the effects of the intervention, after controlling for pre-intervention scores, whose effect is captured by the coefficient γ_2; the parameter ψ captures the relationship, if any, between pre-intervention scores and the experimental treatment, allowing researchers to assess whether subjects differed significantly on the pre-

Communications of the Association for Information Systems

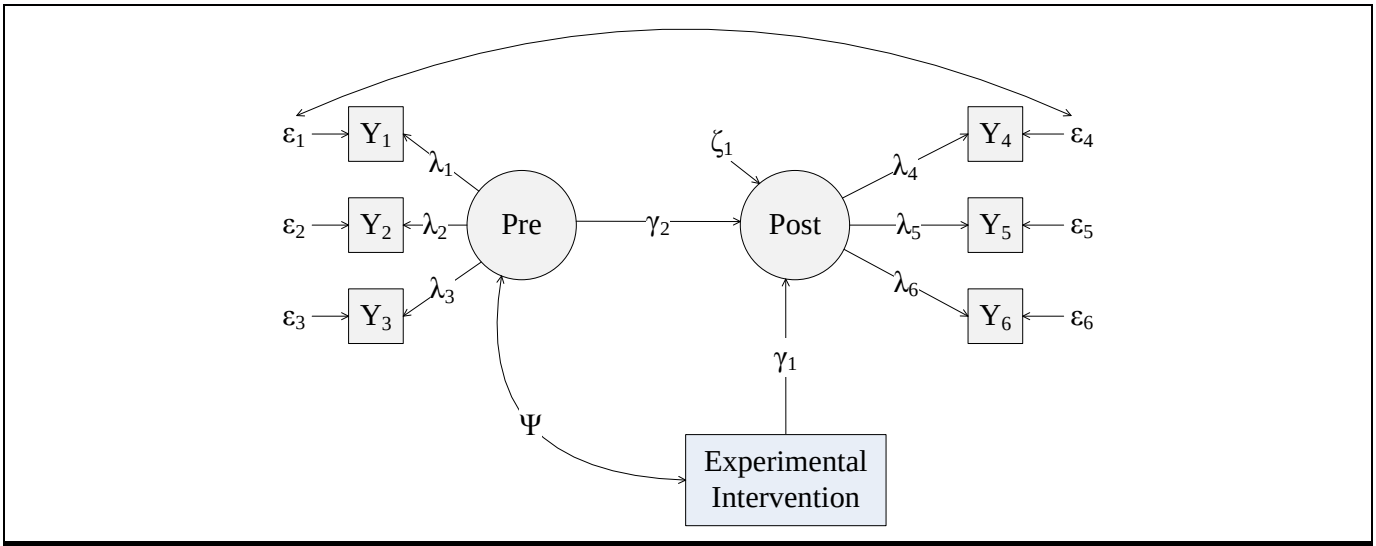

Figure 15. Pre- and Post-test Design Group Code Approach

intervention scores and thus the degree of effectiveness of random assignment to experimental conditions. When analyzing this research design, researchers assess the significance of the experimental effects through the significance of γ_1, and the adequacy of random assignment through the significance of ψ. This is equivalent to conducting a t-test on the composite pre-test scores across both groups, which is sometimes done to provide evidence of adequate random assignment. The latent variable formulation, however, allows for the consideration of measurement error in the indicators and thus enhances the power of this test, as already noted.

The research model depicted in Figure 15 also includes a freed relationship between the residuals of each pair of identical items used to measure the construct on both occasions (for ease of exposition, only one relationship is shown, between ε_1 and ε_4). As is well known, the total variance of an indicator can be decomposed into three separate elements: (1) common variance, which is shared with other related measures and modeled through the inclusion of a path from a latent variable to the indicator, (2) specific variance, which is unique to that particular indicator, and (3) random error. The last two components are usually grouped together and modeled as unexplained residual variance in common applications of factor analysis. Particularly, cross-sectional designs do not allow for each component to be modeled separately [Raffalovich and Bohrnstedt, 1987]. The rationale for allowing the correlation between specific pairs of residuals in Figure 15 can be stated as follows.

Consider the case of indicators Y_1 and Y_4, which represent the same survey question in the pre- and post-test measures. As just noted, the variance of each of these two indicators is a combination of common, specific, and random error components. In this particular example, both indicators share a common antecedent in the pre-test latent variable. In the case of Y_1 this effect is direct, through the loading λ_1. In the case of Y_4, this effect operates indirectly through the regression coefficient relating both latent variables, γ_2, and the loading relating the post-test measure to Y_4, shown as λ_4 in Figure 14. The two indicators also share the specific component of their individual variance, since they are identical copies of each other, albeit measured on different occasions. If this other shared variance is not explicitly modeled, as would be done by allowing the residuals to correlate, the statistical algorithm will attempt to fit a model that best accounts for both sources of variance through the only common path between the two indicators, just described.

Since there are several pairs of identical indicators in this research model, but all of them related through the same channel, doing so would not only confound the true relationship between the pre- and post-test constructs, but also force the shared uniqueness between pairs of indicators to be equal across each pair. This would likely lead to a decrease in the fit of the model and may lead researchers to inaccurately conclude that the proposed research model does not adequately account for the relationship between the observed indicators. Allowing residuals to correlate across each pair of indicators takes into consideration the presence of this relationship between them that is not accounted by the structural model, and removes this confounding from the model estimates for both parameters of interest and overall fit.

Figure 16 shows the same research model specified using the structured means alternative (only one group is shown for clarity). The preceding discussion on the need to allow for the correlation between pairs of residuals equally applies here. In this Figure, γ_1 represents the relationship between the pre- and post-test representations of the construct. Since the treatment and control groups are modeled separately, other parameters of interest are

estimated by comparing pairs of constrained and unconstrained models. In this particular case, testing for the equality of latent means between the pre-test measure in both groups serves to assess the effectiveness of random assignment, as there should not be any significant differences between the two groups. While the group code approach limited the effect of the pre-test construct on its post-test counterpart to be equal across the two groups, the structured means approach can empirically test this assumption by analyzing the effects of constraining γ_1 to be equal across both groups.

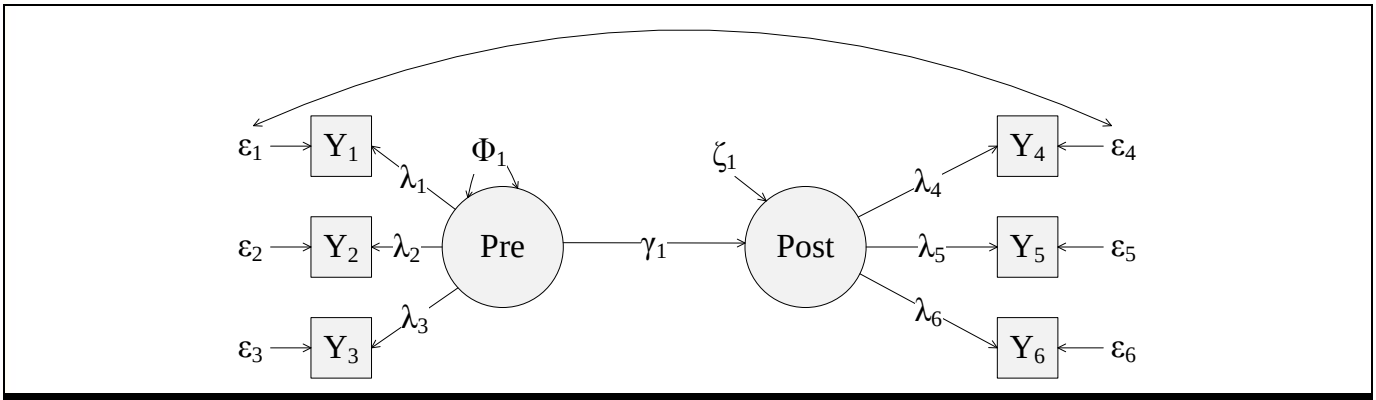

Figure 16. Pre- and Post-test Design Structured Means Approach

V. LIMITATIONS OF THE SEM-LV APPROACH

This section discusses some of the limitations inherent in the techniques discussed thus far. As has been shown above, the techniques proposed here are quite flexible and can accommodate the variety of research designs employed by IS researchers, and other more complex ones that have yet to become commonplace in the discipline (such as, for example, latent growth models). On the other hand, the application of these techniques to actual research places sometimes stringent demands on data collection and analysis. Some of those are discussed next.

Sample Size

Because of their ability to take measurement error into consideration—which, when not modeled appropriately, results in inflated standard errors and reduced power for approaches based on observed variables [Ree and Carretta, 2006]—latent variable techniques exhibit more power to detect existing effects at the same sample size than the ANOVA family of techniques does. On the other hand, researchers employing latent variable approaches must be wary of convergence and stability issues arising from the use of small samples in conjunction with these techniques. At the same time, any results based on small samples are more subject to sampling variability and less likely to accurately estimate the underlying experimental effect, if any. In addition, smaller samples and observed variable approaches compound the limitations of the latter with regards to statistical power. Therefore, at the smaller end of the spectrum of sample size, the ANOVA family of techniques can provide a researcher with results, whereas latent variable approaches would be hard pressed to do so (unless the research model under examination is very simple and only a few parameters are estimated). The downside of choosing this alternative is working with more variable samples that are less representative of the underlying population of interest and reduced statistical power that is less likely to result in a significant finding.

Partial Least Squares

For the purpose of this discussion, there are two important limitations when using PLS to model the type of research designs described here. First, PLS is not a latent variable technique [Marcoulides, Chin, and Saunders, 2009]. Rather than working with the theoretical constructs of interest that are presumed to underlie observed indicators, PLS substitutes those with weighted combinations of the latter. Given that each observed indicator contains both common and random error variance, the weighted composites employed by PLS will also contain common and error variance, in proportions reflecting the quality of each individual indicator and the weight given to each by the algorithm. As a result, PLS does not represent a significant improvement in this regard compared to traditional ANOVA techniques, which work on unweighted sums of indicators. Second, it is well-established in the measurement invariance literature that comparisons between groups with regards to construct means, such as those conducted when employing the structured means alternative discussed before, require the equivalence of certain parameters to be tenable before meaningful conclusions can be reached. Doing so, however, is beyond current approaches to between-groups comparisons using PLS [Qureshi and Compeau, 2009], which assume the necessary invariance requirements to hold. The issue is not minor and speaks directly to the ability of PLS to provide meaningful comparisons between groups. Researchers using PLS to estimate the research designs discussed here should be mindful of these issues and how these would impact their results.

Construct Specification: Reflective and Formative

All the discussions and examples in this research are limited to reflective specifications of latent variables. Although discussions of methodological issues associated with formative specifications of latent variables are becoming more common, by and large these assume that the only antecedents to a formatively specified construct are its cause indicators—in this sense, formatively specified latent variables can never be an independent variable, by definition. An important conceptual issue that needs to be addressed is the nature of the relationship between a formatively specified construct and its indicators. As noted by Bollen and Lennox [1991], the set of indicators employed for a formatively specified construct should represent a census, and not a sample; all indicators that form the latent variable should be included. When a formatively specified latent variable is the subject of an experimental manipulation, an additional cause of that latent variable, the experimental manipulation itself, is added to the research model. If using the group code approach, is the dummy variable representing group assignment an additional cause indicator of the latent variable, and should it be treated as such? The answer, and its implications, is not entirely clear. Much still remains to be understood about formative specifications of latent variables. While not discarding the possibility of employing formatively specified latent variables in research designs such as the ones discussed here, more research is needed into alternative approaches for doing so, and their theoretical and statistical implications.

VI. CONCLUSION AND FURTHER READING

The goal of this article was to present, in an integrated manner, a comprehensive treatment of the foundations of SEM-LV analysis of research designs that include between-group comparisons, as well as to discuss its advantages over traditional ANOVA on the sum of observed variables, and to illustrate the potential of latent variables analysis to study more complex scenarios. Some technical and analytical details were omitted, where necessary, to maintain ease of exposition, but interested readers can refer to the original cited sources.

The discussion presented here is by no means complete, and there is a large body of methodological literature bearing on the issue. Confidence intervals for the difference between means, for example, were not included in this introduction. In addition to the expository treatments by Hancock [1997, 2003, 2004], Bagozzi [1977], Bagozzi and Yi [1989], and Mackenzie [2001], more technical developments can be found in Hancock [2001], Hancock et al. [2000], Cole et al. [1993], Kano [2000], Kühnel [1988], Millsap and Everson [1991], Muthén [1989], Russell, Kahn, Spoth, and Altmaier [1998], Sörbom [1978], Ployhart and Oswald [2004], Thompson and Green [2006], Choi, Fan, and Hancock [2009], and Mcdonald, Seifert, Lorenzet, Givens, and Jaccard [2002], among others.

There are a number of advantages associated with SEM-LV that lead to recommending the use of the latent variable approach outlined here instead of the traditionally used analysis of observed variables. These are summarized in Table 1 below. The only disadvantage associated with SEM has to do with the relatively higher sample size requirements that are needed for implementing the latent variable approach, due to the estimation of a higher number of parameters, such as loadings and residual variances. This will likely result in an increase in the cost of conducting research. Indeed, this type of analysis may not be feasible for populations where participants are either difficult or expensive to recruit. On the other hand, given that the analysis of observed variables is biased by

Table 1: Advantages of the SEM-LV Approach over Traditional Methods

Ability to include manipulation checks in a comprehensive model, instead of being analyzed separately and ad-hoc
Explicit representation in the research model of intervening unobservable variables or constructs that effectively become the target of manipulation or group assignments
Accounts for the biasing effects of measurement errors in the dependent variable or variables, manipulation checks, and latent covariates, if any
The effect of the manipulation on the dependent variable can be decomposed into two distinct ones: the effectiveness of the manipulation in varying the intervening variable, which also allows for assessing the effectiveness of the treatment, and the theoretically interesting effect of the intervening variable on the ultimate dependent construct.
Allows for the differential reliability of the items used to measure the constructs involved; whereas working on a sum score assumes equal importance for all items measuring a construct, latent variable modeling allows for the possibility of heterogeneous loadings.
Relationships are modeled and tested between the constructs of a theory and not between their observable operationalizations.
Systems of dependent variables, as well as latent covariates, can be accommodated.
Assessments of construct validity can be integrated with the main analysis instead of conducted separately through exploratory factor analysis.
Flexibility to accommodate complex models, such as those discussed here.

measurement error, it is not very clear to what extent the need for larger samples (to achieve desired power with traditional techniques) compensates this downside of the latent variable approach. This is one of many areas that remain open for future research to address.

There are also some instances in which the use of latent variable techniques is not necessary, and these involve studies that focus on observed dependent variables that are of interest in their own right, and not because they are indicators representing an underlying construct which is the main focus of the study. Examples of these scenarios, from published IS research, include the work of Tsai, Egelman, Cranor, and Acquisti [forthcoming], who examined the effects of online privacy information on purchasing behavior, comparing mean prices paid (an observed variable of interest) by subjects in the different conditions; Shanks et al. [2008], who examined the effects of ontological clarity on the time taken to solve a problem under each separate condition; or Keith, Shao, and Steinbart [2009], who compared the number of login failures across different approaches to the creation of passphrases.

The use of latent variables to analyze data from experimental and quasi-experimental research designs, as well as more complex models involving the assignment of subjects to groups, random or otherwise, is well-grounded in methodological literature and has been extensively studied. Indeed, much of the foundation for this type of analysis was developed more than thirty years ago. It has only been recently, however, that this has begun to be applied in the organizational and social sciences, and then mostly by researchers who are themselves methodological experts. Moreover, this is largely absent in contemporary Information Systems research. It is our hope that this brief, non-technical, exposition will help with its diffusion into the discipline and improve our research practice.

ACKNOWLEDGMENTS

I would like to thank the editors and reviewers for the many thoughtful comments and suggestions offered during the review process. Any errors or omissions remain solely my own.

REFERENCES

Arnold, V., T. Benford, C. Hampton and S. Sutton (2010) “Competing Pressures of Risk and Absorptive Capacity Potential on Commitment and Information Sharing in Global Supply Chains”, *European Journal of Information Systems,* (19)2, pp. 134–152.

Bagozzi, R. (1977) “Structural Equation Models in Experimental Research”, *Journal of Marketing Research,* (14), pp. 209–226.

Bagozzi, R. and Y. Yi (1989) “On the Use of Structural Equation Models in Experimental Design”, *Journal of Marketing Research,* (26), pp. 271–284.

Balijepally, V., R. Mahapatra, S. Nerur and K. Price (2009) “Are Two Heads Better Than One for Software Development? The Productivity Paradox of Pair Programming”, *MIS Quarterly,* (33)1, pp. 91–118.

Bollen, K. and R. Lennox (1991) “Conventional Wisdom on Measurement: A Structural Equation Perspective”, *Psychological Bulletin,* (110)2, pp. 305–314.

Bowen, P., R. O’Farrell and F. Rhode (2009) “An Empirical Investigation of End-user Query Development: The Effects of Improved Model Expressiveness vs. Complexity”, *Information Systems Research,* (20)4, pp. 565–584.

Burton-Jones, A. and P. Meso (2008) “The Effects of Decomposition Quality and Multiple Forms of Information on Novices’ Understanding of a Domain from a Conceptual Model”, *Journal of the Association for Information Systems,* (9)12, pp. 148–802.

Byrne, B., R. Shavelson and B. Muthén (1989) “Testing for Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance”, *Psychological Bulletin,* (105), pp. 456–466.

Choi, J., W. Fan and G. Hancock (2009) “A Note on Confidence Intervals for Two-Group Latent Mean Effect Size Measures”, *Marketing Letters,* (44)3, pp. 396–406.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences, 2nd edition,* New York, NY: Lawrence Erlbaum.

Cole, D., S. Maxwell, R. Arvey and E. Salas (1993) “Multivariate Group Comparisons of Variable Systems: MANOVA and Structural Equation Modeling”, *Psychological Bulletin,* (114)1, pp. 174–184.

Gonzalez, R. and D. Griffin (2001) “Testing Parameters in Structural Equation Modeling: Every ‘One’ Matters”, *Psychological Methods,* (6)3, pp. 258–269.

Hancock, G. (1997) “Structural Equation Modeling Methods of Hypothesis Testing of Latent Variable Means”, *Measurement and Evaluation in Counseling and Development,* (30)2, pp. 91–105.

Hancock, G. (2001) “Effect Size, Power, and Sample Size Determination for Structured Means Modeling and MIMIC Approaches to Between-groups Hypothesis Testing of Means on a Single Latent Construct”, *Psychometrika,* (66)3, pp. 373–388.

Hancock, G. (2003) “Fortune Cookies, Measurement Error, and Experimental Design”, *Journal of Modern Applied Statistical Methods,* (2)2, pp. 293–305.

Hancock, G. (2004) “Experimental, Quasi-experimental, and Nonexperimental Design and Analysis with Latent Variables”, in Kaplan, D. (ed.) *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, Thousand Oaks, CA: SAGE Publications.

Hancock, G., F. Lawrence and J. Nevitt (2000) “Type I Error and Power of Latent Mean Methods and MANOVA in Factorially Invariant and Noninvariant Latent Variable Systems”, *Structural Equation Modeling,* (7)4, pp. 534–556.

Heninger, W., A. Dennis and K. Hilmer (2006) “Individual Cognition and Dual-task Interference in Group Support Systems”, *Information Systems Research,* (17)4, pp. 415–424.

Hong, W., J. Thong and K. Tam (2005) “The Effects of Information Format and Shopping Task on Consumers’ Online Shopping Behavior: A Cognitive Fit Perspective”, *Journal of Management Information Systems,* (21)3, pp. 149–184.

Jaccard, J. and C. Wan (2006) *LISREL Approaches to Interaction Effects in Multiple Regression*, Thousand Oaks, CA: Sage Publications.

Jiang, Z. and I. Benbasat (2007) “The Effects of Presentation Formats and Task Complexity on Online Consumers’ Product Understanding”, *MIS Quarterly.* (31)3, pp. 475–500.

Jiang, Z., J. Chan, B. Tan and W. Chua (2010) “Effects of Interactivity on Website Involvement and Purchase Intention”, *Journal of the Association for Information Systems,* (11)1, pp. 34–59.

Judd, C. and D. Kenny (1981) “Process Analysis: Estimating Mediation in Treatment Evaluations”, *Evaluation Review,* (5)5, pp. 602–619.

Kano, Y. (2000) “Structural Equation Modeling for Experimental Data”, in DuToit, S., R. Cudeck, and D. Sörbom (eds.) *Structural Equation Modeling: Present and Future*, Chicago, IL: Scientific Software International.

Keith, M., B. Shao and P. Steinbart (2009) “A Behavioral Analysis of Passphrase Design and Effectiveness”, *Journal of the Association for Information Systems,* (10)2, pp. 64–88.

Kim, D. and I. Benbasat (2006) “The Effects of Trust-assuring Arguments on Consumer Trust in Internet Stores: Application of Toulmin’s Model of Argumentation”, *Information Systems Research,* (17)3, pp. 286–300.

Kuechler, W. and V. Vaishnavi (2006) “So, Talk to Me: The Effect of Explicit Goals on the Comprehension of Business Process Narratives”, *MIS Quarterly,* (30)4, pp. 961–979.

Kühnel, S. (1988) “Testing MANOVA Designs with LISREL”, *Sociological Methods & Research,* (16)4, pp. 504–523.

MacKenzie, S. (2001) “Opportunities for Improving Consumer Research Through Latent Variable Structural Equation Modeling”, *Journal of Consumer Research,* (28), pp. 159–166.

Mcdonald, R., C. Seifert, S. Lorenzet, S. Givens and J. Jaccard (2002) “The Effectiveness of Methods for Analyzing Multivariate Factorial Data”, *Organizational Research Methods,* (5)3, pp. 255–274.

Marcoulides, G., W. Chin and C. Saunders (2009) “A Critical Look at Partial Least Squares Modeling”, *MIS Quarterly,* (33)1, pp. 171–175.

Mauro, R. (1990) “Understanding L.O.V.E. (left out variables error): A Method for Estimating the Effects of Omitted Variables”, *Psychological Bulletin,* (108)2, pp. 314–329.

Millsap, R. and H. Everson (1991) “Confirmatory Measurement Model Comparisons Using Latent Means”, *Multivariate Behavioral Research,* (26), pp. 479–497.

Muthén, B. (1989) “Latent Variable Modeling in Heterogeneous Populations”, *Psychometrika,* (54), pp. 557–585.

Muthén, L. and B. Muthén (2002) “How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power”, *Structural Equation Modeling,* (9)4, pp. 599–620.

Piccoli, G. and B. Ives (2003) “Trust and the Unintended Effects of Behavior Control in Virtual Teams”, *MIS Quarterly,* (27)3, pp. 365–395.

Ployhart, R. and F. Oswald (2004) “Applications of Mean and Covariance Structure Analysis: Integrating Correlational and Experimental Approaches”, *Organizational Research Methods,* (7)1, pp. 27–65.

Qiu, L. and I. Benbasat (2009) “Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems”, *Journal of Management Information Systems,* (25)4, pp. 145–182.

Qureshi, I. and D. Compeau (2009) “Assessing Between-group Differences in Information Systems Research: A Comparison of Covariance- and Component-based SEM”, *MIS Quarterly,* (33)1, pp. 197–214.

Raffalovich, L. and G. Bohrnstedt (1987) “Common, Specific, and Error Variance Components of Factor Models”, *Sociological Methods & Research,* (15)4, pp. 385–405.

Ree, M. and T. Carretta (2006) “The Role of Measurement Error in Familiar Statistics”, *Organizational Research Methods,* (9)1, pp. 99–112.

Russell, D., J. Kahn, R. Spoth and E. Altmaier (1998) “Analyzing Data from Experimental Studies: A Latent Variable Structural Equation Modeling Approach”, *Journal of Counseling Psychology,* (45)1, pp. 18–29.

Santhanam, R., S. Sasidharan and J. Webster (2008) “Using Self-regulatory Learning to Enhance e-Learning-based Information Technology Training”, *Information Systems Research,* (19)1, pp. 26–47.

Saris, W. and A. Satorra (1993) “Power Evaluations in Structural Equation Models”, in Bollen, K., and J. Long (eds.) *Testing Structural Equation Models*, Newbury Park, CA: Sage.

Shanks, G., E. Tansley, J. Nuredini, D. Tobin and R. Weber (2008) “Representing Part–Whole Relationships in Conceptual Modeling: An Empirical Evaluation”, *MIS Quarterly,* (32)3, pp. 553–573.

Sörbom, D. (1978) “An Alternative to the Methodology for Analysis of Covariance”, *Psychometrika,* (43), pp. 381–396.

Stewart, K. (2006) “How Hypertext Links Influence Consumer Perceptions to Build and Degrade Trust Online”, *Journal of Management Information Systems,* (23)1, pp. 183–210.

Tan, C., H. Teo and I. Benbasat (2010) “Assessing Screening and Evaluation Decision Support Systems: A Resource-matching Approach”, *Information Systems Research,* (21)2, pp. 305–326.

Thompson, M. and S. Green (2006) “Evaluating Between-group Differences in Latent Variable Means”, in Hancock, G., and R. Mueller (eds.) *Structural Equation Modeling: A Second Course*, Greenwich, CT: Information Age Publishing.

Tiwana, A. and M. Keil (2009) “Control in Internal and Outsourced Software Projects”, *Journal of Management Information Systems,* (26)3, pp. 9–44.

Tsai, J., S. Egelman, L. Cranor and A. Acquisti (forthcoming) “The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study”, *Information Systems Research,* (22)2, pp. 254–268.

Vandenberg, R. (2002) “Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures”, *Organizational Research Methods,* (5)2, pp. 139–158.

Vandenberg, R. and C. Lance (2000) “A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research”, *Organizational Research Methods,* (3)1, pp. 4–70.

Wakefield, R. and D. Whitten (2006) “Mobile Computing: A User Study on Hedonic/Utilitarian Mobile Device Usage”, *European Journal of Information Systems,* (15), pp. 292–300.

Zhang, J., X. Luo, S. Akkaladevi and J. Ziegelmayer (2009) “Improving Multiple-password Recall: An Empirical Study”, *European Journal of Information Systems,* (18)2, pp. 165–176.

ABOUT THE AUTHOR

Miguel I. Aguirre-Urreta is an Assistant Professor in the School of Accountancy and Management Information Systems, Richard H. Driehaus College of Business, at DePaul University in Chicago, IL. His work has been published or is forthcoming in *MIS Quarterly, Information Systems Research*, *The DATA BASE for Advances in Information Systems*, the *Journal of Organizational and End User Computing*, *Research Synthesis Methods*, *Behaviour & Information Technology,* the *Review of Business Information Systems*, and *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*. His research interests are in the areas of individual performance with the use of technology, technology acceptance and diffusion, systems analysis and design, and research methods.

Communications of the Association for Information Systems

ISSN: 1529-3181

Communications of the Association for Information Systems