SEM for Experimental Designs: An Information Systems Example

Saurabh Gupta
Coggin College of Business, University of North Florida, USA
s.gupta@unf.edu

Abstract: IS research has matured significantly over the last three decades, leading to increasingly complex research designs as well as complex analytical techniques to analyze the data collected. Similar advances have happened in the experimental and quasi-experimental designs. Some key characteristics of these advances are: 1) use of latent variables approaches to operationalize key variables; 2) the need to understand the causal relationship between elements of the study; 3) the need to study the effects of technology as an addition to existing methods of working; and, 4) recognition that some conditions create greater change in outcomes over time. In spite of these advances in data collection and design, researchers are still confirming data collected via experiments to use ANOVA for analysis. This paper outlines an analytical technique that moves Information Systems experimental research beyond ANOVA. By combining and extending three advances in Structural Equation Modeling techniques, namely Mean and Covariance Structure analysis, Stacked Group modeling and Latent Growth modeling, the paper outlines a robust analysis technique that accommodates the above-mentioned advances in experimental design. The technique provides for an in-depth test of all model assumptions, as well as the flexibility to accommodate an increasing variety of experimental designs. A detailed example is provided to illustrate the usage of the technique in an Information Systems context. The example shows now only the accommodations needed in an information systems context, but also how this technique can be used to extract results from existing research methods that was not possible with ANOVA. The arguments presented in the paper as well as the example on how to use should provide future researchers with a guideline on how to use these techniques as well as provide a platform for how they can extend these techniques to accommodate more research method advances.

Keywords: SEM, experiments, stacked group modeling, latent growth modeling, invariance, Information Systems

1. Introduction

The landscape of information systems research is filled with different methods, ranging from surveys, case studies to experiments (Jenkins, 1985). To a large extent, the choice of the research method depends on the ability of the research question to answer the research question. Each of these methods has its own strength and weakness, especially as it relates to the analytical abilities associated with each. While most research methods used in business research have graduated to advanced analytical techniques (especially with the advances in Structural Equation Modeling (SEM)), analytical techniques for experimental data have been stuck in the *"ANOVA mindset syndrome*" (MacCallum, 1998) (e.g. (Sutanto et al., 2013, Gregg and Walczak, 2008)). This is especially true for Information Systems research, which relies heavily on experimental studies. The advances in analytical capabilities in other methods have provided researchers with the ability to answer more questions from a given research method. Consequently, there is a decrease in the number of empirical papers using experiments as a research method.

The use of experiments as a research method in information systems have a long history (Benbasat, 1989, Jarvenpaa et al., 1985). This research design makes intuitive sense for a lot of questions asked by IS researchers as they are generally comparing a context without information technology with a context with advanced information technology or comparing two different technology usages. Such questions have been asked throughout the history of technology use, from the original Minnesota experiments (Dickson et al., 1977), to the contemporary studies (Zhang et al., 2009).

Over the last 35 years IS experimental designs have matured considerably. Some of the complexity is due to the adoption of latent variable approaches to operationalize key variables in the studies. Further, there is recognition that experimental conditions may not only result in mean differences, but also differences in the strength of association between the key variables (i.e., a regression coefficient may grow stronger or weaker in the presence of one condition vs. another). For example, a recent study comparing the effectiveness of two different end-user training methods implied that the underlying cognitive latent variables might influence outcomes (in this case, it implied that paired training might improve over time) (Davis and Yi, 2004, Gupta and Bostrom, 2013). Similar examples exist in other IS domains as well (e.g. (Kearns, 2012, Hosack et al., 2012)).

In spite of these changes, researchers have continued to use ANOVA/MANOVA as their dominant analytical approach; fitting their data to conform to constraints and assumptions of this technique. These traditional

Reference this paper as: Gupta S. "SEM for Experimental Designs: An Information Systems Example" *The Electronic Journal of Business Research Methods Volume 12 Issue 1 2014 (pp 27-40),* available online at www.ejbrm.com

analytical tools cannot simultaneously account for the above mentioned complexities and their application results in inferences that may lack empirical validity. Given the prevalence of MANOVA, we do not have a formal section reviewing MANOVA literature. However, appropriate studies are cited where necessary. The purpose of this paper is to introduce how SEM based analysis can be extended to address some of these design challenges -- latent variables as inputs, compare non-equivalent structural models, as well as special cases like longitudinal designs by simultaneously comparing modeling variance, covariance and mean differences. While SEM does not lend itself directly to experiment design (because of imbalanced groups), in this paper we discuss the variations that makes it suitable.

The rest of the paper is laid out as follows. First we compare the traditional techniques with the proposed technique of using SEM. Next, an illustration is provided as these issues are discussed. The paper ends with a set of guidelines for users and issues that have not yet been addressed.

2. Analytical Approach Comparison

MANOVA/ANOVA techniques are familiar to all experimental researchers and thus, serve as a well-established benchmark. As mentioned earlier, the proposed analytical technique outlined in this paper is based on SEM. To understand the issues, benefits and usages, the two techniques need to be discussed across their purpose, assumptions and modeling process including the power to predict / model validation. This is presented the next section.

2.1 Purpose of Analytical Methods

Most IS experiments are designed as input-output studies i.e. comparing dependent variable means across two or more treatments (as outlined in the reviews of IS areas such as virtual teams (Pinsonneault and Caya, 2005), computer self-efficacy (Compeau et al., 2005), or end-user training (Gupta et al., 2010)). In such cases, i.e. where the interest is in simple means test, MANOVA is a good analytical technique and SEM based technique add nothing of importance or even of value.

The most important reason for using SEM is in cases where predictors beyond the treatment conditions are of interest. Called regression for explanations, in such cases, researchers want to know not only how well the predictors explain the criterion variable, but also which specific predictors are most important. In experimental setting, such techniques allow researchers interested in prediction that go beyond variance accounted for (r-squared) to specific regression weights (Maruyama, 1997). This has led some researchers to use PLS based analysis in an experimental setting, where a proxy variable is used to represent latent variables (e.g. (Cooper and Haines, 2008)). In cases of balanced designs, such use of PLS or covariance based techniques is fine and has been discussed extensively in Qureshi and Compeau (2009).

However, PLS/covariance based techniques do not allow for simultaneous comparison of mean differences between dependent variables – a critical requirement in an experimental study. The technique outlined in this study borrows heavily on work generally known as Means and Covariance analysis (MACS) by Ployhart et al. (2004) which allows for such analysis. Secondly, in cases where a group with advanced information technology is compared with a group without the advanced information technology, the introduction of technology also introduces a new latent variable in the model such as technology usage; this results in an unbalanced SEM design. This is a big issue for IS research because of the nature of the underlying research questions. For example, in a recent study, Gupta and Bostrom (2013) investigated technology based learning vs. traditional learning. In such cases, SEM techniques are not directly applicable since SEM based technique require structural models in all groups being compared to be the same.

Some researchers have run separate PLS analysis for each of the groups (e.g. (Yoo and Alavi, 2001)) to analyze regression paths. Their analysis generally creates a separate independent variable to represent groups. The structural models are similar in this case. However, this creates issues of interpretational confounding due to possible structural invariance i.e. structural model estimates across the groups are assumed to be same. In addition, this analysis assumes a balanced structural model across groups; which might not be always true in experimental studies.

To overcome the constraints outlined above, an SEM technique called stacked group approach can be used. This technique was initially proposed by Hayduk (1996). This technique allows for the existence of a latent variable in one group that is absent in the other group. Overall, the SEM technique proposed in this paper does the following. First, it extends Hayduk's work to latent variables by including item level information. Second, and more importantly, it combines Hayduk's work with MACS analysis technique to provide a comprehensive analysis method for experimental studies. Finally, it is also possible to extent this technique for longitudinal experiments.

2.2 Assumptions among Analytical Methods

As mentioned earlier, the dominant method in analyzing data based on experiment designs is MANOVA. MANOVA has three critical assumptions that need to be examined in the context of social science research – Independence among observations, measurement invariance and normality of observations.

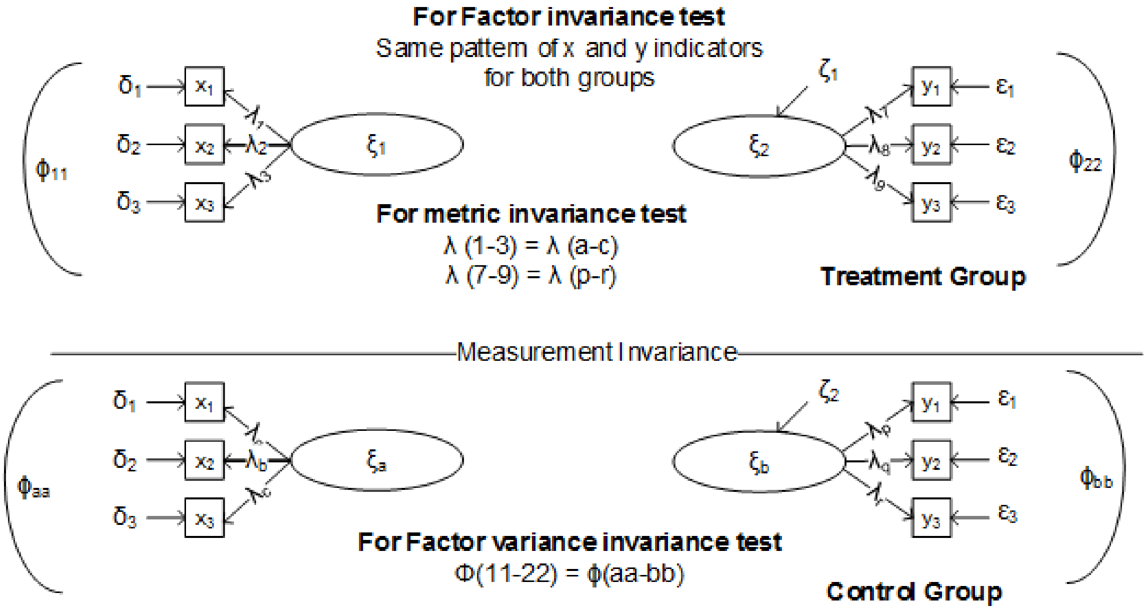

Figure 1: Conceptual Representation of Measurement Invariance for Stacked Group Model

2.3 Independence among observations

In MANOVA, the most basic, yet most serious, violation of an assumption occurs when there is lack of independence among observations. There are three reasons for such a violation: time-ordered effects, gathering information in group settings, and extraneous and unmeasured effect. Most IS experiments are done in group settings with common subject demographics across treatments. While this does increase the internal validity of the result, it could result in a lack of independence among observation. In addition, most experimental techniques assume orthogonality between the predictor variables, i.e., the treatment conditions. However, in behavioral science research, given the psychological nature of the variables, predictors are not completely orthogonal. In addition, there are no tests that detect all forms of dependence with absolute certainty. The advantage of using SEM methods is that they allow for exogenous variables to correlate with each other alleviating some of the effects of dependence. In addition, SEM does not assume orthogonality when analyzing data (Maruyama, 1997).

2.4 Latent Variables and Measurement Invariance

The Information systems discipline sits at the interception of individual behavior and information technology. Thus, researchers in this area increasingly employ latent variables as a part of the measurement approach (Boudreau et al., 2001). In such an approach, latent variables data is collected using various underlying items – which when loaded onto the latent variable provide information regarding the same. MANOVA based approached, however, are not designed to accommodate item level information. In such case, researcher generally start out appropriately by first examining the properties of the latent variables through a confirmatory factor analysis (CFA). Then, however, they fall back to some aggregate form of the measure (e.g., its algebraic mean or average) to test the hypotheses usually in an ANOVA framework. Doing so is inappropriate because the measure is assumed to contain no error. In addition, IS researchers have assumed, in the vast majority of studies, that a variable will have the same conceptual meaning to one group receiving one level of the treatment as it does to the other groups receiving other levels of the treatment; i.e., the measures are invariant across conditions (Biros et al., 2002, Yi and Davis, 2003, Davis and Yi, 2004, Piccoli and Ives, 2003, Yoo and Alavi, 2001). An SEM approach, in contrast, includes measurement error because variables (i.e., latent variables) are operationalized using item-level information rather than an average or summed value.

The related assumption in MANOVA is the equivalence of covariance matrices across the groups (James et al., 1982). A violation of this assumption might lead to overestimation of estimates and thus, resulting in a Type 1 error (although more recently researchers have suggested that a violation of this assumption has minimal impact if the groups are of approximately equal size (Horn and McArdle, 1992)). In well-controlled IS experiments, the groups sizes are generally not that different and this is not a problem in case of discrete measurements. However, in case of latent variables, other elements of invariance need to be tested. When measures are not invariant, any conclusions based on the findings can lead to misleading or even false conclusions (Taylor and Todd, 1995, Williams et al., 2009, Vandenberg and Lance, 2000, Ployhart and Oswald, 2004). SEM methods allow for testing of each of these concerns. A review of the literature points to three critical tests that are relevant to the current scenario: configural invariance, metric invariance and factor variance invariance (Schmitt and Kuljanin, 2008). Configural invariance requires a demonstration that the same factors and pattern of factor loadings explains the variance-covariance matrices associated with the groups' responses (see Figure 1). If configural invariance is not supported, then comparisons between groups on the measures is not possible because this literally means that there is no conceptual equivalence of the constructs (Horn and McArdle, 1992). This is generally not a problem in IS research as most researchers tend to use well-validated existing instruments. However, the concern is valid when using new measures.

The second test is metric invariance, also called a test of strong factorial invariance (Horn et al. 1992). It tests a null hypothesis that factor loadings for like items are invariant across time/group, i.e., a test of tau equivalence across groups. It evaluates whether the groups are calibrating their responses to the underlying latent variable to the same degree (e.g., a response of 3 means neither agree-nor disagree in all groups). At least partial metric invariance must be established in order for subsequent tests to be meaningful. The assumption of homosedacity or parallel equivalence among constructs of different groups is usually not a problem, except in cases where the means of dependent variables are correlated with the variance across groups. This assumption is addressed in the factor variance invariance test. Metric and factor invariance analysis is done by looking at the model fit estimate as well as Δ CHI for the changes in degrees of freedom from configural invariance. The three tests are summarized in Figure 1 and are further explained in the illustrative example later.

2.5 Normality

The last assumption for covariance based analysis relates to normality of variables. In the strictest sense, the assumption is that all variables are multivariate normal. Multivariate normality assumes that joint effect of two variables is normally distributed. Even though this assumption underlies both of the mentioned techniques, there is no direct test for multivariate normality. Violations of this assumption have little impact with larger sample size. For SEM based methods, a departure from normality (as shown in a recent study based on Monte Carlo simulations) does reduce the ability to detect differences between groups (Qureshi and Compeau, 2009).

2.6 Modeling Process

The sequence of steps required to setup a MANOVA based analysis is well known. MANOVA based analysis relies on comparison of mean vectors across two or more groups. Thus, it requires equity of means matrix dimensions.

Hayduck (1996) extended the basic multi-group model approach (the term used to convey the comparison between groups in SEM) to accommodate the experimental design imbalance. While Hayduck's technique did not focus on latent variables, this paper extends the approach to incorporate latent variables. This approach requires the following modifications of the covariance matrix, the means vector, beta matrix, phi mate, and theta-epsilon matrix.

Covariance matrices for all groups should typically consist of an equal number of items. However, in the cases where a condition is not present (and thus, a latent variable does not exist), proxy variables need to be created with dummy codes. This helps equate the number of items across groups. The correlation of these dummy items with observed items should be fixed to 0, and the variance of the items should be fixed to 1. The error of the item is fixed to 0. This prevents the dummy variables from impacting other variables in the model, but creates a proxy latent variable in the overall model to deal with the unbalanced conditions.

The variance of the latent constructs in cases where they do not really exist should be fixed to 1. Their factor loadings (lamdas) should also be fixed to 1. This allows for model identification in an SEM analysis.

Beta values of the causal path for the dummy latent variables are fixed to 0 in groups without these latent variables. This prevents the dummy variable from having any causal effect on the endogenous variables. In groups where the latent variables exist, they should be freely estimated.

Latent means (alpha values in the case of the example) for the dummy variables should be fixed to 0 in groups without these variables. Latent means for the endogenous variables are freely estimated, if group mean differences are expected. Mean differences are usually the case in experimental designs.
All other variables should be constrained to be equal across groups.

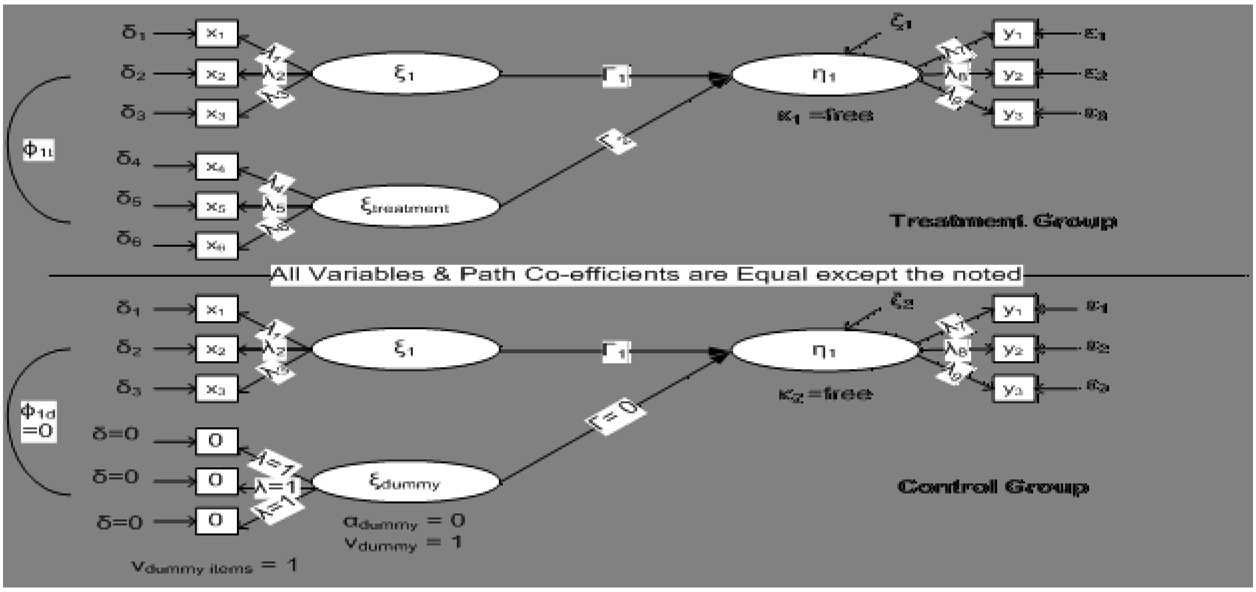

Figure 2: Conceptual Representation of a Latent Variable Stacked Group Model

Figure 2 summarizes these manipulations for a two-group, three latent variable analyses. The conceptual model could be extended for more groups/latent variables. It is important to note that it is not recommended restricting the values of the measurement model based on an invariance test. The model parameters discussed in the above sections can used as starting estimates to help with model identification. Considerable literature exists in this area to support this position of not identifying measurement model and structural model separately (Anderson and Gerbing, 1988).

2.7 Assessing Overall Fit

In MANOVA, the focus is on seeing if there are mean differences between dependent variables across groups. Thus, the criteria assess the differences across dimensions of the dependent variables. There seems to be an agreement that either Pillai's criterion or Wilks' lambda best meets the needs, although evidence suggests that Pillai's criterion is more robust and should be used if sample size decreases, unequal cell sizes appear, or homogeneity of covariance is violated (Hair, 1998).

SEM based analysis, on the other hand, tend to focus on overall model fit. The most basic chi-square goodness of fit test, although valuable, is limited because it is a direct function to sample size. A lot of other fit indexes exist and can be generally divided across two categories: Absolute and Relative. Absolute indexes address the question: Is the residual or unexplained variance remaining after model fitting acceptable? These indexes include those that use the function that is minimized. Chi-squared fit index falls in this category. Other fit indexes that are popularly reported goodness of fit index and the root mean residual. Relative indexes address the question: How well does a particular model do in explaining a set of observed data compared to other possible models? Thus, the model is viewed as falling along a continuum between the worst possible fitting model and a best fit model that perfectly fits the observed data. Overall, two commonly used indexes are the Tucker Lewis Index and the Non-Normed Fit Index. For a summary of fit indexes, see (Bentler, 2007, Cheung and Rensvold, 2002, Hu and Bentler, 1999).

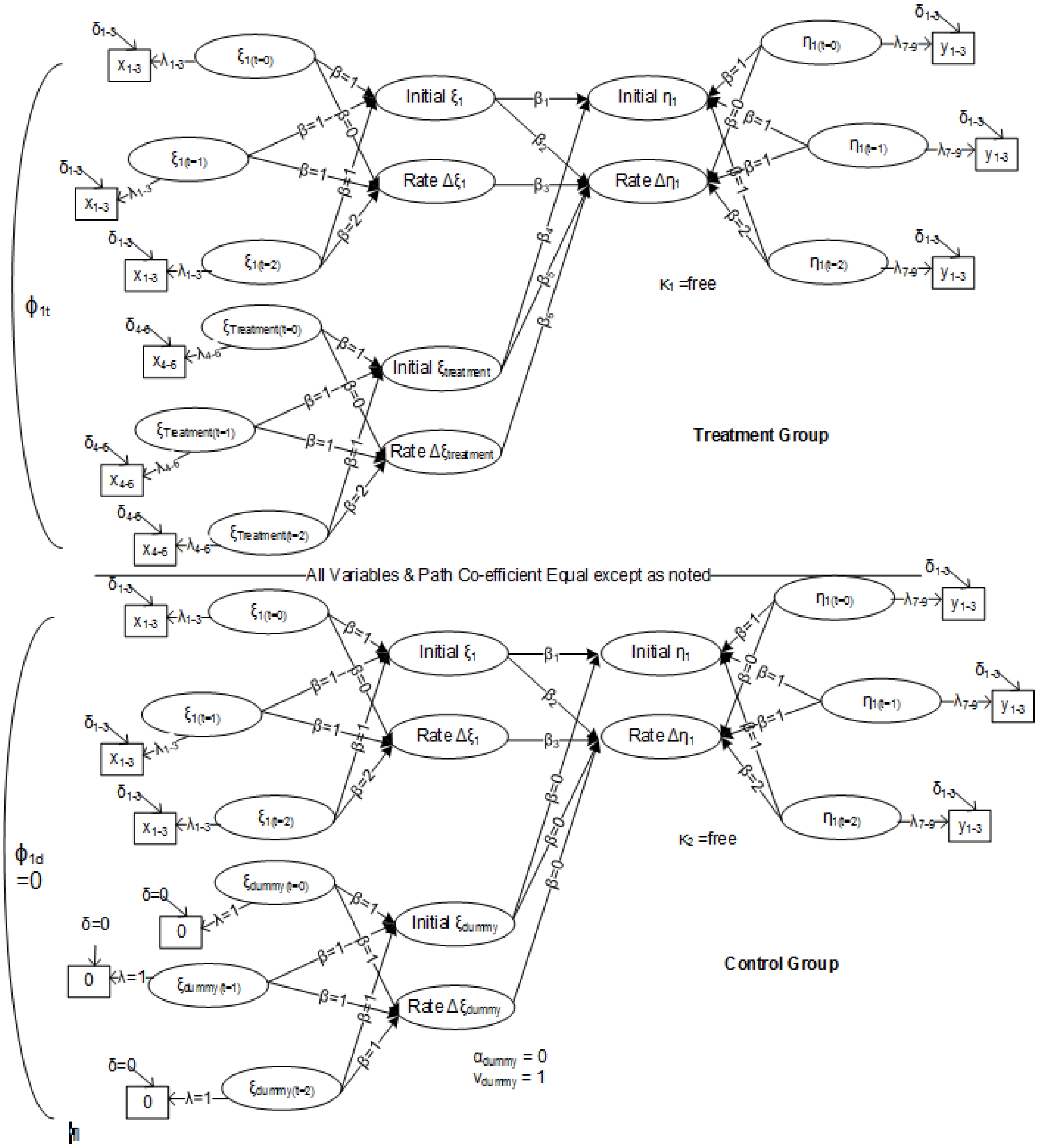

Figure 3: Conceptual Model of Second Order LGM Stacked Group Model

2.8 A Special Case: Repeated Measures

IS researchers have always pointed to ‘time’ as an important element in understanding IS phenomenon. However, to date, experimental researchers have restricted the analysis of data to repeated measures ANOVA (e.g. Sein et al. (1989)). While the latter permits testing of mean differences across time, it is not an analysis of actual change over time (Ployhart and Vandenberg, 2010).

A special situation occurs when the same respondent provides several measures, such as test scores over time and the researchers wish to examine them to see whether any trend emerges. Generally repeated measures MANOVA is used account for this dependence and still ascertain whether any differences occurred across individuals for the set of dependent variables. Such an analysis presents three concerns. First, a common practice in traditional analyses with Time 1 and Time 2 measures of some constructs is to regress the Time 2 onto the Time 1 and to treat the residual as if it is a full representation of the focal construct underlying the measure. It is not the latter at all and indeed, it is “something” of unknown validity at this point. By controlling initial levels (including autoregressive effect), traditional models eliminate all predictors except those that predicts change in rank order of observations over time. This is a disadvantage when studying monotonically stable phenomenon (Meredith and Tisak, 1990) where the rank order is the same although significant change at individual and group level is occurring. Secondly, these traditional approaches fail to provide adequate generalizations to more than two points in time. This also constrains growth curves to linear models. The outlined approach provides the ability to incorporate latent growth models and thus, to accommodate change across multiple time points, as well as different growth curves (Duncan, 1999). Third, traditional approaches assume means to be 0. However, in experiment designs, sample means of outcome variables carries useful statistical information needed to estimate both inter- and intra-group differences across time.

The analysis of actual change requires the use of latent growth modeling (LGM) (Lance et al., 2000). Thus, as explained shortly below, the model in Figure 2 needs to be extended to include LGM. Before performing the actual analysis, measurement invariance across time needs to be established. Similar to the invariance across groups, this is completed by constraining the respective free parameters of the dependent and independent latent variables and their items across time to be equal (Chan, 2002, Fan, 2003, Lance et al., 2000). The end results of the LGM are two latent variables – one representing the initial status (aka intercept) on the focal variable and the other representing the slope or change in the variable across time. To incorporate the LGM into the “unbalanced” design technique explained above, the following modifications (in addition to the stacked group model modifications) are required:

Dummy LGM parameters (slope and initial status) are created for the missing latent variables in the appropriate groups. The means for these LGM parameters are fixed to 1. This allows for model identification.

The beta estimates from the dummy LGM parameters are fixed to 0 to prevent causal effects on outcome/endogenous variables.

The conceptual model is depicted in Figure 3. The model shows only two groups with two exogenous and one endogenous variable. However, the concepts outlined in this paper so far can be extended to multiple groups and multiple exogenous and endogenous variables. The entire model (both measurement and structural model) can be estimated as a single analysis.

2.9 Result Interpretation

MANOVA allows researchers to examine results at two levels: (1) assessing which dependent variable(s) exhibit differences across the groups; and (2) identifying which groups differ on single dependent variable or the entire dependent variate. Additionally, MANOVA also allows for an analysis of covariates. The most direct approach to analyzing the overall impact of covariates is to run the analysis with and without covariates. Effective covariates improve the statistical power of the tests and reduce within-group variance.

The SEM based approach outlined in this paper, on the other hand, provides three key results. First, similar to MANOVA, this approach provides mean estimates (and model variance explained) for each of the dependent variables. Subsequently, the model can be forced to have equal dependent variable latent means. If the model significantly degrades, the means are different. This does not, however, provide an analysis for directional hypothesis. A better approach is to compare the means provided using a standard t-test. Secondly, the approach outlined also estimates regression coefficients in the form of path estimates. These explain the direct and indirect causal effect on dependent variables of independent variables other than treatments. Finally, for longitudinal studies, this method also provides the initial and change values of the latent variables.

In the next section, we illustrate the above mentioned approach using an example from Information Systems discipline

Illustrative Example

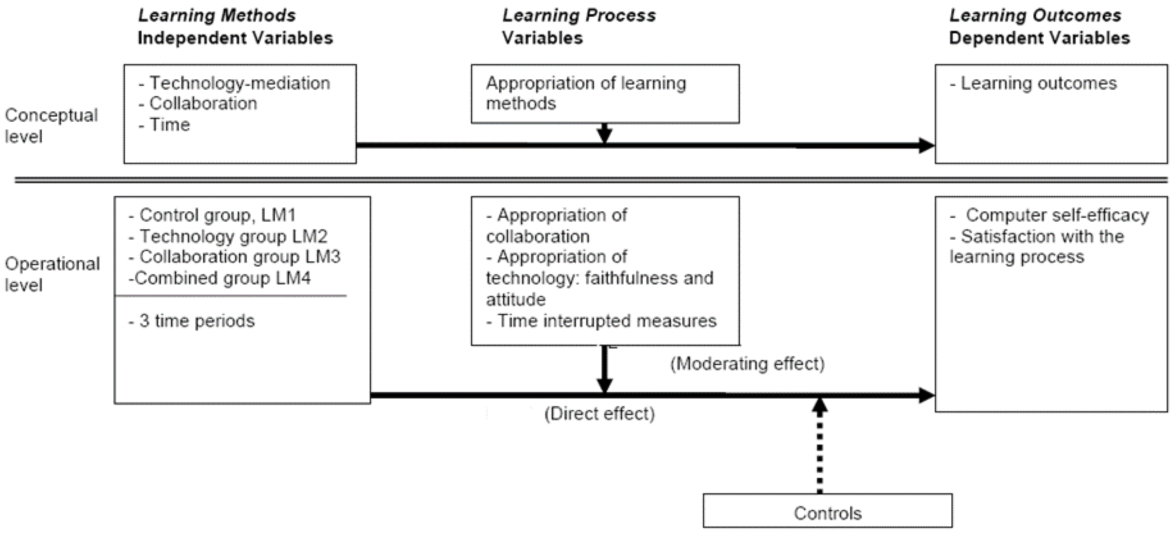

Figure 4: Illustrative Example Overview

The example outlined here is based on the conceptual model outlined in Gupta and Bostrom (2009). Their paper articulates a model based on Adaptive Structuration Theory. In brief, the focus is on technology-mediated learning that explicitly configures elements of the learning process, including team, technology, and learning technique structures. The model argues that the effect of different learning methods on learning outcomes is moderated by the level of appropriation of the learning method (see Figure 4).

Table 1: Measurement Invariance Goodness-of-Fit: Across Groups

Dependent variables – Self-efficacy and satisfaction							
	df	CHI	NNFI	CFI	RMSEA	delta df	Δ CHI
Configural	455	617.01	0.94	0.95	0.06		
Metric	467	661.08	0.93	0.95	0.06	12	44.07
Factor	518	719.59	0.93	0.94	0.06	51	58.51
Independent variable – faithfulness & attitude toward technology use							
Configural	334	412.9	0.97	0.98	0.04		
Metric	339	430.53	0.97	0.97	0.05	5	17.63
Factor	356	472.7	0.97	0.97	0.05	17	42.17
Independent variables – Collaboration							
Configural	252	422.99	0.97	0.97	0.08		
Metric	257	432.77	0.97	0.97	0.08	5	9.78
Factor	261	436.63	0.97	0.97	0.08	4	3.86
Df=degrees of freedom, Goodness-of-fit guidelines - CHI=χ^2 value, delta df = changes in the degrees of freedom compared to the last test, Δ CHI = changes in χ^2 value from the previous test.							

A 2X2 multi-period experimental study was used to operationalize the model in Figure 4. Four training methods were employed: group 1 (LM1) received training only through behavioral modeling; group 2 (LM2) participants received training using both behavioral modeling and enactive learning, using a web-based computer training system (WBCT); and, group 3 (LM3) and group 4 (LM4) participants replicated LM1 & LM2 respectively, but the participants worked in groups of two.

Time was manipulated using a longitudinal design by conducting three training sessions (different content in each session) for each group. The learning process was measured using the faithfulness of use of learning technology (4 items), their attitude toward technology (3 items), and the level of collaboration (6 items).

Learning outcome was operationalized using two measures, change in self-efficacy (3 items) and satisfaction from the learning process (3 items). Sample sizes for each group were: LM1: Control Group (N=113), LM2: Group with technology (N=117), LM3: Group with Collaboration (N=85) and LM4: Group with technology & Collaboration (N=119).

Some important features of the design were: 1) Given the different learning methods across the treatments, there was a possibility that participants going through the training program might perceive / interpret the measurements differently across groups;. 2) The model measured learning outcomes as two latent variables, i.e., self-efficacy and satisfaction from the process; 3)

The model also had three moderating latent variables, i.e., appropriation of technology (faithfulness and attitude) and appropriation of collaboration and 4) While the experiment was balanced in its implementation, it was imbalanced with respect to the structural model. This is seen in the configuration of the treatments where LM2 adds the technology component to LM1 and LM4 adds collaboration to LM3. It also implied that the respective moderating variables only existed when the relevant conditions were present, i.e., appropriation of technology can only exist in conditions where participants are using technology as a part of their learning method. As mentioned earlier, the experiment also had a longitudinal component. Data for the experiment was captured across three sessions. The data from this complex design was analyzed using the SEM analytical approach discussed above. LISREL 8.74 was used for analysis.

Study hypothesis posited a positive effect of use of enactive learning and collaboration on learning outcomes. However, it also predicted that the outcomes would be moderated by the levels of appropriation. Finally, it was hypothesized that levels of appropriation will change positively over time and that this change will have an influence on learning outcomes.

2.10 Invariance Assumption Testing

Table 2: Measurement Invariance across Time for Dependent Variables

LM4	Df	CHI	NNFI	CFI	RMSEA	delta df	Δ CHI
Configural	102	106.98	0.98	0.99	0.02		
Metric	110	128.84	0.97	0.98	0.04	8	21.86
Factor	114	140.32	0.96	0.97	0.04	4	11.48
LM3	Df	CHI	NNFI	CFI	RMSEA	delta df	Δ CHI
Configural	102	137.03	0.90	0.94	0.06		
Metric	110	146.90	0.90	0.93	0.06	8	9.87
Factor	114	150.73	0.90	0.93	0.06	4	3.83
LM2	Df	CHI	NNFI	CFI	RMSEA	delta df	Δ CHI
Configural	102	165.70	0.92	0.95	0.07		
Metric	110	179.23	0.92	0.94	0.07	8	13.53
Factor	114	202.78	0.91	0.93	0.08	4	23.55
Factor SATIS	113	186.36	0.92	0.94	0.07	3	7.13
LM1	Df	CHI	NNFI	CFI	RMSEA	delta df	Δ CHI
Configural	102	120.10	0.96	0.98	0.04		
Metric	110	129.89	0.96	0.97	0.04	8	9.79
Factor	114	140.05	0.91	0.97	0.05	4	10.16
Df=degrees of freedom, Goodness-of-fit guidelines - CHI=$\chi 2$ value, delta df = changes in the degrees of freedom compared to the last test, Δ CHI = changes in $\chi 2$ value from the previous test							

Prior to testing the hypotheses, measurement invariance was evaluated. With respect to the model in our example (see Figure 4), the specific concern was whether the dependent variables (self-efficacy and satisfaction), technology-related moderating variables (faithfulness of technology use, attitude toward technology use), and team-related moderating variables (level of collaboration) are invariant between the four different treatment groups. Table 1 summarizes the results of analysis. Goodness-of-fit indices are reported as a part of the analysis. Similar invariance tests were done across the three time periods for each group for each set of variables. The results for dependent variable invariance across time for each of the four groups is shown in Table 2

Overall, all variables showed good invariance. The exception was for factor variance invariance across time for the satisfaction construct after the second training session in treatment LM2. However, when the factor variance for satisfaction for the second training period was freely estimated, the chi-square difference value was not statistically significant (see Table 2). Thus, in the final analysis, this factor variance was left to be freely estimated in subsequent tests. This is a good example of heterogeneity of variance and how it was incorporated (accounted for) into the analysis.

2.11 Mean Estimate

In this example, we were also interested in knowing if the moderating variables, i.e., level of faithfulness and collaboration, had an impact on outcomes across groups. When performing a comparison across groups, as was the case in this study, we equated the path estimates of groups (Beta/Gamma invariance) with the relevant control group, i.e., Faithfulness of and attitude toward WBCT use were fixed in the WBCT groups (LM2 & LM4), while collaboration paths estimates were equated with the path estimates of groups that had collaboration (LM3 & LM4). Thus, LM1 acted as a control group for training outcomes without WBCT and collaboration, while LM4 acted as a group where both these training methods were used, in addition to the method in LM1. The LISREL code used for analysis utilized the earlier discussed updates to the covariance matrix, the means vector, beta matrix, phi mate, and theta-epsilon matrix. The code also allowed the exogenous constructs to correlate similarly across groups, i.e., fixing the same values of phi across groups. The entire analysis was done in one pass (i.e., although we had measurement model estimates from earlier analysis, these were only as starting values). Table 3 reports the means for the dependent variable as well as what the result meant in the context of the experiment.

Table 3: Means (Variance) of Endogenous / Dependent Variables

Excel Self-efficacy	Video (I)	WBT (J)	Means diff. (J-I)
Individual (X)	2.62 (0.87)	2.88 (0.93)	0.26*
Collaborative (Y)	2.86 (0.87)	3.04 (0.93)	0.18*
Means differences (Y-X)	0.24**	0.16	
Comment	Excel Self-efficacy was significantly higher for individuals in WBT usage.		
Satisfaction (reverse measured)	Video (I)	WBT (J)	Means diff. (J-I)
Individual (X)	2.04 (0.42)	2.95 (0.71)	0.91*
Collaborative (Y)	2.02 (0.44)	1.96 (0.53)	0.06
Means diff. (Y-X)	-0.02	-0.99*	
Comment	Satisfaction from the learning process was significantly higher for individuals in collaboration + WBT training method.		
* = P<0.05, ** = p<0.1			

Table 4: Effect for Initial Value of Exogenous Variables to Initial Values of Endogenous Outcomes

	Self-efficacy	Satisfaction from process
Faithfulness	0.16	0.73*
Comment	The level of faithful use of technology had a positive moderating impact on learner satisfaction.	
Attitude	0.56*	1.32*
Comment	The more students perceived the learning technology to good, the greater was the impact of the on self-efficacy and satisfaction enhancement	
Collaboration	0.01	0.165
Comment	The level of collaboration did not influence the impact of learning method on learning outcomes	
* = P<0.05, ** = p<0.1		

2.12 Regression Estimates

Table 3 shows the means for the two dependent/endogenous variables (self-efficacy and satisfaction from the process). The output also reported standardized beta values. These represent the magnitude and direction of the causal path. Table 4 reports these values as well as what the results meant in the experimental context.

The technique outlined so far is sufficient for most quasi-experiments conducted in IS. It allows for statistical control of all experiment method assumptions, using latent variables as exogenous and endogenous variables, and means analysis as well as providing an explanation for the variance in the endogenous variables.

2.13 Latent Growth Parameters

The next step in the analysis process evaluated whether there were any changes in the latent variables over time. The example was were interested in the rate of change in both the dependent and independent variables. Since the time periods for data collection were well defined and uniform, all models were specified as linear growth models. The fully identified model was the combined treatment LM4. Dummy latent variables for the respective missing growth parameters were created in the rest of the three groups as discussed earlier.

The LGM parameters are reported in Table 5. A comparison with the cut-off values for these estimates showed good model fit. All constructs showed a significant change over time, except for faithfulness. This means that no significant change in faithfulness was detected over time. Thus, in the subsequent analysis, causal paths stemming from changes in faithfulness construct were not considered for interpretation.

Table 5: Latent Growth Parameter Means Estimates (Alpha Estimates)

	LM4	LM3	LM2	LM1
IN Faithfulness	1.69*	NA	2.02*	NA
Ch Faithfulness	0.13	NA	0.08	NA
IN Attitude	5.11*	NA	5.03*	NA
Ch Attitude	-0.27*	NA	-0.27*	NA
IN Collaboration	6.25*	6.18*	NA	NA
Ch Collaboration	-0.08*	-0.12*	NA	NA
IN Self-efficacy	3.07*	2.9*	2.89*	2.6*
Ch Self efficacy	0.27*	0.28*	0.23*	0.2*
IN Satisfaction	2.03	2.08	2.27	2.14*
Ch Satisfaction	0.21*	0.17*	0.37*	0.3*
* Statistically significant (z-value >1.64) IN = Initial Status, Ch= Change Treatments: LM1 = Control, LM2=WBCT, LM3=dyads, LM4=Combined				

The path coefficients or beta estimates are summarized in Table 6. The values for goodness-of-fit indicators were as follows – NNFI =0.99, RMSEA =0.01 and GFI=0.96, all within acceptable range (Hu and Bentler, 1999). The table showed that the initial levels of faithfulness were significant in affecting initial levels of satisfaction. In addition, the initial levels of attitude had a significant effect on initial levels of self-efficacy and satisfaction. No significant effects on initial levels of collaboration were found. Changes in attitude, though, showed a strong correlation to changes in satisfaction.

Changes in appropriation of collaboration structures showed a strong correlation with satisfaction and self-efficacy. Although the discussion of the results is beyond the scope of this paper, the analytical techniques did show that a significant influence of WBCT and collaboration, as well as the combined treatment on specific learning outcomes. In addition, the study found that appropriation as an important determinant of the extent of these effects.

While the laboratory experiment provided a good context to illustrate the application of the analytical technique, it should be noted that the premises of the technique can also be applied to other forms of quasi-experimental designs, as well as field experiments.

Table 6: Effect of Changes in Appropriation to Changes in Learning Outcomes

	Self-efficacy	Satisfaction from process
Change in Faithfulness	0.13	0.0
Comment	The rate of change of faithfulness had no impact on self-efficacy or learner satisfaction. This is not unusual since no change in faithfulness was found.	
Change in Attitude	0.255	0.525*
Comment	The rate of change in attitude with respect to time towards learning technology significantly increases satisfaction from learning process i.e. over time, the impact of attitude on learner satisfaction increased.	
Collaboration	0.075	0.51*
Comment	The rate of change in the level of collaboration has a positive impact on learner satisfaction i.e. It is possible that if the experiment continued over extended period of time, there would be significant impact of collaboration on learner satisfaction.	
* = P<0.05, ** = p<0.1		

3. Limitations, Future Research and Conclusion

The linkage between the method and the analytical techniques is paramount in the selection of the method. Like all maturing disciplines, IS research has become increasingly complex, dwelling in more domains, with an increasing variety of methodological contexts. This has led to increased complexity in IS experimental and quasi-experimental design as well.

This paper outlined an analysis method that allows us to leverage the strengths of an experiment as well as providing causal and mean difference analysis. This technique brings the richness of advanced SEM practices to experimental designs in a structured format. Using the technique outlined, experiments can be designed to test as well as accommodate violations of three critical assumptions of MANOVA i.e. Independence among observations, measurement invariance and normality. In addition, the experiment design now not only provides for means difference analysis, but also has the ability to test for causality. Given the internal validity of the experiment, such a result has the ability to provide great insight into the phenomenon under investigation. This is a significant contribution to the literature on experimental designs.

The outlined technique and the example analysis, though, have their limitations as well. First, the sample size requirements are considerably larger than those needed for traditional techniques. This is because of the number of paths and variables that are needed to be identified in the SEM model. Secondly, the approach and example assumed that all variables are continuous. Extensions of this approach are needed for other types of variables such as discrete variables. These are extensively used in IS studies and, while the technique can be modified to accommodate it, future studies are needed to validate it. Third, it was assumed that the growth curve across all models was linear. This might not be true in certain contexts and thus, curvilinear change of some form will need to be modeled. Fortunately, non-linear change can be easily accommodated. Finally, in spite of the complexity of the SEM design, the fit indicator for the model came close to the suggested cut-off values. Given the complexity of this approach, new cut-off values need to be researched.

Despite the latter limitations, the current approach accommodated three important elements that were outside the scope of traditional approaches to analyze the same data. First, the general use of SEM in analyzing experimental data allows for the inclusion of latent variables. Additionally, the technique also tested for the assumptions of an experiment such as similarity of conditions across the experimental groups. The test of assumptions is especially important in cases where design randomization is not achieved or possible. Further, by testing for invariance, there is some level of confidence that groups in different experimental conditions and across time are interpreting measures in the same manner.

Second, most IS quasi-experiments are restricted to evaluation of group differences, with only theoretical support regarding causes for the difference. The outlined technique allowed regression for path analysis, as well as mean difference analysis. This is a significant advancement for experimental data analysis since it provides data support for the causal arguments. Additionally, it also helps quantify the level of effect for each of the constructs. Finally, this technique allows modeling of the changes in latent variables across groups for both endogenous and exogenous variables, as well as to see their causal effects.

In closing, the paper hopes that the technique outlined above provides the initial stepping-stone for IS researchers (as well as other social science researchers) to move beyond just the ANOVA mindset syndrome to exploit their experiments towards more interpretation. It provides a method that can accommodate a wide array of experimental situations, while providing a richer insight into the data.

Reference

Anderson, J. C. & Gerbing, D. W. 1988. Structural Equation Modeling In Practice: A Review And Recommended Two-Step Approach. *Psychological Bulletin,* 103**,** 411-423.

Benbasat, I. 1989. Laboratory Experiments In Information Systems Studies With A Focus On Individuals: A Critical Appraisal. *In:* Benbasat, I. (Ed.) *The Information Systems Research Challange: Experimental Research Methods.* Boston, Ma: Harvard Business School.

Bentler, P. M. 2007. On Tests And Indices For Evaluating Structural Models. *Personality And Individual Differences,* 42**,** 825-829.

Biros, D. P., George, J. F. & Zmud, R. W. 2002. Inducing Sensitivity To Deception In Order To Improve Decision Making Performance: A Field Study. *Mis Quarterly,* 26**,** 119-144.

Boudreau, M.-C., Gefen, D. & Straub, D. W. 2001. Validation In Information Systems Research: A State Of The Art Assessment. *Mis Quarterly,* 25**,** 1-16.

Chan, D. 2002. Latent Growth Modeling. *In:* Drasgow, F. & Schmitt, N. (Eds.) *Measuring And Analyzing Behavior In Organizations : Advances In Measurement And Data Analysis.* 1st Ed. San Francisco: Jossey-Bass.

Cheung, G. W. & Rensvold, R. B. 2002. Evaluating Goodness-Of-Fit Indexes For Testing Measurement Invariance. *Structural Equation Modeling,* 9**,** 233-255.

Compeau, D., Gravill, J., Haggerty, N. & Kelley, H. 2005. Computer Self-Efficacy: A Review. *In:* Galletta, D. & Zhang, P. (Eds.) *Human-Computer Interaction In Management Information Systems.* M. E. Sharpe, Inc.

Cooper, R. B. & Haines, R. 2008. The Influence Of Workspace Awareness On Group Intellective Decision Effectiveness. *European Journal Of Information Systems,* 17**,** 631-648.

Davis, F. D. & Yi, M. Y. 2004. Improving Computer Skill Training: Behavior Modeling, Symbolic Mental Rehearsal, And The Role Of Knowledge Structures. *Journal Of Applied Psychology,* 89**,** 509-523.

Dickson, G. W., Senn, J. A. & Chervany, N. L. 1977. Research In Management Information Systems: The Minnesota Experiments. *Management Science,* 23**,** 913-923.

Duncan, T. E. 1999. *An Introduction To Latent Variable Growth Curve Modeling Concepts, Issues, And Applications,* Mahwah, N.J., L. Erlbaum Associates.

Fan, X. 2003. Power Of Latent Growth Modeling For Detecting Group Differences In Linear Growth Trajectory Parameters. *Structural Equation Modeling,* 10**,** 380-400.

Gregg, D. G. & Walczak, S. 2008. Dressing Your Online Auction Business For Success: An Experiment Comparing Two Ebay Businesses. *Mis Quarterly,* 32**,** 653-670.

Gupta, S. & Bostrom, R. P. 2009. Technology-Mediated Learning: A Comprehensive Theoretical Model. *Journal Of Association Of Information Systems,* 10**,** *637-660*.

Gupta, S. & Bostrom, R. P. 2013. An Investigation Of The Appropriation Of Technology-Mediated Training Methods Incorporating Enactive And Collaborative Learning. *Information Systems Research,* 24**,** 454-469.

Gupta, S., Bostrom, R. P. & Huber, M. 2010. End-User Training Methods: What We Know, Need To Know. *Sigmis Database For Advances In Information Systems,* 41**,** 9-39.

Hair, J. F. 1998. *Multivariate Data Analysis,* Upper Saddle River, N.J., Prentice Hall.

Hayduk, L. A. 1996. *Lisrel Issues, Debates, And Strategies,* Baltimore, Johns Hopkins University Press.

Horn, J. L. & Mcardle, J. J. 1992. A Practical And Theoretical Guide To Measurement Invariance In Aging Research. *Experimental Aging Research,* 18**,** 117-144.

Hosack, B., Lim, B. & Vogt, W. P. 2012. Increasing Student Performance Through The Use Of Web Services In Introductory Programming Classrooms: Results From A Series Of Quasi-Experiments. *Journal Of Information Systems Education,* 23**,** 373-383.

Hu, L.-T. & Bentler, P. M. 1999. Cutoff Criteria For Fit Indexes In Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling,* 6**,** 1-55.

James, L. R., Mulaik, S. A. & Brett, J. M. 1982. *Causal Analysis : Assumptions, Models, And Data,* Beverly Hills, Sage Publications.

Jarvenpaa, S. L., Dickson, G. W. & Desanctis, G. 1985. Methodological Issues In Experimental Is Reserch: Experiences And Recommendation. *Mis Quarterly,* 9**,** 141-156.

Jenkins, A. M. 1985. Research Methodologies And Mis Research. *In:* Mumford, E. (Ed.) *Research Methods In Information Systems.* Elsevier Science Publishers.

Kearns, M. 2012. Experiments In Social Computation. *Communications Of The Acm,* 55**,** 56-67.

Lance, C. E., Vandenberg, R. J. & Self, R. M. 2000. Latent Growth Models Of Individual Change: The Case Of Newcomer Adjustment. *Organizational Behavior And Human Decision Processes,* 83**,** 107-140.

Maccallum, R. C. 1998. Commentary On Quantitative Methods In I-O Research. *Industrial-Organizational Psychologist,* 35**,** 19-30.

Maruyama, G. 1997. *Basics Of Structural Equation Modeling,* Thousand Oaks, Calif., Sage Publications.

Meredith, M. & Tisak, J. 1990. Latent Curve Analysis. *Pychometrika***,** 107-122.

Piccoli, G. & Ives, B. 2003. Trust And The Unintended Effects Of Behavior Control In Virtual Teams. *Mis Quarterly,* 27**,** 365-395.

Pinsonneault, A. & Caya, O. 2005. Virtual Teams: What We Know, What We Don't Know. *International Journal Of E-Collaboration,* 1**,** 1-16.

Ployhart, R. E. & Oswald, F. L. 2004. Applications Of Mean And Covariance Structure Analysis: Integrating Correlational And Experimental Approaches. *Organizational Research Methods,* 7**,** 27-65.

Ployhart, R. E. & Vandenberg, R. J. 2010. Longitudinal Research: The Theory, Design, And Analysis Of Change. *Journal Of Management,* 36**,** 94-120.

Qureshi, I. & Compeau, D. 2009. Assessing Between-Group Differences In Information Systems Research: A Comparison Of Covariance-And Component-Based Sem. *Mis Quarterly,* 33**,** 197-214.

Schmitt, N. & Kuljanin, G. 2008. Measurement Invariance: Review Of Practice And Implications. *Human Resource Management Review,* 18**,** 210-222.

Sein, M. K. & Bostrom, R. P. 1989. Individual Differences And Conceptual Models In Training Novice Users. *Human Computer Interaction,* 4**,** 197-229.

Sutanto, J., Palme, E., Chuan-Hoo, T. & Chee Wei, P. 2013. Addressing The Personalization-Privacy Paradox: An Empirical Assessment From A Field Experiment On Smartphone Users. *Mis Quarterly,* 37**,** 1141-A5.

Taylor, S. & Todd, P. A. 1995. Understanding Information Technology Usage: A Test Of Competing Models. *Information Systems Research,* 6**,** 144-176.

Vandenberg, R. J. & Lance, C. E. 2000. A Review And Synthesis Of Measurement Invariance Literature: Suggestions, Practices, And Recommendations For Organizational Research. *Organizational Research Methods,* 3**,** 4-70.

Williams, L., Vandenberg, R. & Edwards, J. 2009. 12 Structural Equation Modeling In Management Research: A Guide For Improved Analysis. *The Academy Of Management Annals,* 3**,** 543-604.

Yi, M. Y. & Davis, F. D. 2003. Developing And Validating An Observational Learning Model Of Computer Software Training And Skill Acquisition. *Information Systems Research,* 14**,** 146-170.

Yoo, Y. & Alavi, M. 2001. Media And Group Cohesion: Relative Influences On Social Presence, Task Participation, And Group Consensus. *Mis Quarterly,* 25**,** 371-390.

Zhang, J., Luo, X., Akkaladevi, S. & Ziegelmayer, J. 2009. Improving Multiple-Password Recall: An Empirical Study. *European Journal Of Information Systems,* 18**,** 165-176.