

A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study

Wynne W. Chin • Barbara L. Marcolin • Peter R. Newsted

C. T. Bauer College of Business, University of Houston, Houston, Texas 77204

*Haskayne School of Business, University of Calgary, 2500 University Drive NW,
Calgary, Alberta, Canada, T2N 1N4*

*Centre for Innovative Management, Athabasca University, 22 Sir Winston Churchill Avenue,
St. Alberta, Alberta, Canada, T8N 1B4*

wchin@uh.edu • marcolin@ucalgary.ca • pnewsted@mba.athabascau.ca

The ability to detect and accurately estimate the strength of interaction effects are critical issues that are fundamental to social science research in general and IS research in particular. Within the IS discipline, a significant percentage of research has been devoted to examining the conditions and contexts under which relationships may vary, often under the general umbrella of contingency theory (cf. McKeen et al. 1994, Weill and Olson 1989). In our survey of such studies, the majority failed to either detect or provide an estimate of the effect size. In cases where effect sizes are estimated, the numbers are generally small. These results have led some researchers to question both the usefulness of contingency theory and the need to detect interaction effects (e.g., Weill and Olson 1989). This paper addresses this issue by providing a new latent variable modeling approach that can give more accurate estimates of interaction effects by accounting for the measurement error that attenuates the estimated relationships. The capacity of this approach at recovering true effects in comparison to summated regression is demonstrated in a Monte Carlo study that creates a simulated data set in which the underlying true effects are known. Analysis of a second, empirical data set is included to demonstrate the technique's use within IS theory. In this second analysis, substantial direct and interaction effects of enjoyment on electronic-mail adoption are shown to exist.
(PLS; Moderators; Interaction Effects; Structural Equation Modeling; Measurement Error)

Introduction

The ability to detect and accurately estimate interaction effects¹ between quantitative variables can

be difficult. Within the IS discipline, both empirical and theoretical models presenting such relationships can easily be found going back several decades

*Tables, figures, and appendices omitted in the printed version of the article are available online at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

¹ Interaction term within this paper can be thought of in terms of a multiplicative term ($X * Z$) in regression or ANOVA-based

techniques, a moderating arrow in a causal diagram, or high and low lines on a two-dimensional graph. Such an effect implies the existence of a moderator that affects the direction and/or strength of the relation between an independent (i.e., predictor) variable and a dependent or criterion variable.

(e.g., Powers and Dickson 1973, Ginzberg 1979, Franz 1979). In fact, it can be argued that a significant percentage of IS research has been devoted to examining those moderating variables that create interaction effects (i.e., the conditions and contexts under which theoretical relationships may vary) often under the general umbrella of contingency theory (McKeen et al. 1994, Weill and Olson 1989), and within emerging theories (Goodhue and Thompson 1995, Chan et al. 1997, Taylor and Todd 1995, Davis et al. 1992).

Yet, as to be presented in this paper, a majority of past IS studies have either failed to detect a moderating influence or have failed to provide an estimate of the size of the interaction effect. If presented, the effect size values are generally small. Collectively, these results suggest that moderators have a small influence on our developing theories, which in turn has led some researchers to question the usefulness of contingency theory and the need to detect interaction effects (e.g., Weill and Olson 1989). Essentially, the argument is: "Why develop contingency theories or attempt to measure such interaction effects if the extant research indicates either no or minimal effects?"

The position of this paper is that our current lack of understanding and development of contingent effects may be a byproduct of the analytic method as opposed to a failing of theoretical development. Problems in measuring interaction effects are especially pronounced in field research and observational studies (e.g., McClelland and Judd 1993), where much of the difficulties can be attributed to measurement error. Measurement error has been cited as the cause of both lowering the ability to detect as well as underestimating the true effects (Busemeyer and Jones 1983; Aiken and West 1991, pp. 160–165).

The most common techniques employed in IS research for moderator analysis are regression and ANOVA.² Yet, these techniques assume the single item measures being used are perfectly reliable (i.e., error free). To compensate for this, additional item measures are often created, combined into a summated scale, and then used in moderated regression

or ANOVA analysis. The reliability of the scale is assessed using internal consistency measures such as Cronbach's alpha, which (consistent with the summation process) assumes equal weighting of items.

While an improvement over single item measures, it is important to recognize at least two assumptions held in creating summated scales. The first assumption, by nature of the equal weighting process, treats all items as equal in their reliabilities. In turn, this implies all items are equal in their contribution towards estimating the interaction effect. The second assumption is the reliability of the summated scale will remain the same when applied later within a theoretical model. This assumption may not be true given that the reliability estimate is made in isolation as opposed to within the context for which it is subsequently used. Chin and Marcolin (1995), for example, have shown that the reliability of individual items, and therefore the internal consistency of scales, can differ when used in different models. Unfortunately, by virtue of the summation process, we have no opportunity to assess the validity of these two assumptions—that being equal item reliability and unchanging scale reliability.

Thus, moderator analysis should not only account for measurement error during the initial scale construction, but also during the statistical analysis that estimates the interaction effect. A more realistic approach would be to weight the individual items for each scale based on their individual contributions in the context of the interaction model being tested. By weighting items that are more predictive and reliable in estimating an interaction effect, a more accurate scale can be created. It should also provide information when the analysis is completed on the variability of each measure such that only more reliable measures are used for future research.

As a possible solution for assessing and accounting for measurement error when analyzing interaction effects, Kenny and Judd (1984) proposed using a product-indicator approach, in which measures of latent constructs are crossmultiplied to form interaction terms that are used to estimate the underlying latent interaction construct within the LISREL algorithm. While this general approach has been employed in several studies (Kenny and Judd 1984,

² See Tables 1, 2, and 3 at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

Jöreskog and Yang 1996), growing evidence from the literature suggests that the LISREL product-indicator approach can be problematic for several reasons. These reasons include the fact that the LISREL product-indicator approach is technically demanding, often resulting in analytical errors and, even if modeled correctly, is not a complete solution (Bollen and Paxton 1998, p. 267; Li et al. 1998; Ping 1996). These problems are discussed in the Literature Review section below.

Overall, if IS theory is to develop in appropriate ways, our analytical techniques must help uncover the “true” underlying interaction effects we seek. It should neither hinder this search nor propagate the inherent weaknesses of older techniques. Hence, any IS researcher considering moderating influences within a theoretical model should be aware of the problems discussed in this paper so as to produce valid and accurate results.

Towards this end, we argue that a moderator’s measurement error should be considered not only in initial reliability assessment, but also in subsequent analyses of the theoretical model. A new latent variable modeling approach within partial least squares (PLS) is shown to provide this subsequent assessment of measurement, thus overcoming problems within both traditional analytical techniques that can mask measurement error (e.g., aggregated or single indicators) and current problems associated with LISREL-based approaches.

Empirical data is presented to demonstrate how these analytical techniques impact theoretical development through the presentation of two studies: A simulated data set where the underlying true effects are known and an electronic-mail adoption data set where the emotion enjoyment is shown to have not only a substantial direct effect on adoption intention, but also an interaction effect that is stronger than those typically found in IS research. The simulated data set was developed using rigorous Monte Carlo techniques. The analysis of this data enables the comparison of factors—reliability, effect size, number of indicators, and sample size—that can influence measurement error within the analytical techniques.

It will be argued that the new latent variable modeling approach provides more accurate estimates of

interaction effects by accounting for the measurement error within measures that, when ignored, can attenuate the estimated relationships. These more accurate estimates should improve the validation of theories, ensuring that fruitful avenues are maintained and new ones are detected. To the best of our knowledge, the use of product indicators to model interaction effects in a PLS analysis has never been published and by providing this new procedure, we believe researchers will be better able to more accurately detect and estimate contingent relationships—an ability that has been argued in the past to be a useful approach for advancing social science research (Greenwald et al. 1986).

The remainder of this paper is structured as follows. First, the growing importance of interaction terms within IS research is established, and the continuing tendency to employ analysis techniques that accounts for or assesses the fallibility of measures in this work is noted. A summary of the cumulative knowledge is presented, providing an overall picture of theoretical advancements surrounding interaction terms (i.e., moderators). These theoretical advancements include a discussion of measurement error, reliability, LISREL-based approaches, and summated scales to illustrate their role within the analysis of interaction terms. Then a new latent variable PLS modeling approach is introduced in an effort to improve on the previous shortcomings. Following that, two studies are presented to provide empirical evidence. Discussion and conclusions are then offered, summarizing how best to use this new approach.

Literature Review

Interaction terms have been used in the IS field for some time; moderators, multiplicative terms, contingency terms or contingency theory, and interaction effects are all expressions used to refer to interaction terms. We begin by examining the growing importance of interaction terms within the IS literature and the extent to which previous IS contingency research has primarily applied analytical techniques, such as ANOVA³ and multiple regression, that assume

³Note that what is concluded regarding effect size estimations using regression throughout this paper is equally so for ANOVA because the latter represents a special case of the former.

infallible measures. Measurement error is the primary problem that is exposed within these analytical practices, which raises other problems within this literature, such as the use of summated scales. The primary problem with these analytical practices is their inability to handle or present information relating to impact of measurement error. These problems are highlighted and discussed.

Importance to Emerging Theory

Moderators are important to theories being advanced within the IS field, as can be seen by their long history in the literature and their increasing use in dominant theories. Two of these emerging IS theories that employ moderators are task-technology fit (Goodhue and Thompson 1995) and business-IT strategic fit (Chan et al. 1997), which are discussed after a brief review of the history of moderator use in IS research.

Through an exhaustive review of the information systems literature back to 1980,⁴ moderators were found to be present from the start and their importance is evident. For instance, Schonberger (1980) considered how information systems design approaches led to good systems but only when the contingent factors were appropriately matched, such as information needs for MIS supporting functions matching the decision-making type. More recently, McKeen et al. (1994) summarized the contingency studies that existed for the relationship between user participation and satisfaction, and suggested that two out of four moderating variables, namely task and system complexity, changed the relationship. However, with a change in R^2 of 0.012, these two moderators were considered unsubstantial. The authors concluded that these moderating variables were probably not the only two important ones and suggested five other promising moderators for future investigation. From the authors' experience, the unmoderated relationship

was not seen as adequate for an investigation into user participation and satisfaction. Throughout the years, similar studies employing moderators can be found in information systems research providing a long and important history.

In addition to this long history, moderators are increasingly used in dominant theories within the field. Two well-known theories in which the concept of moderation is important are Goodhue and Thompson's (1995) task-technology fit model and Chan et al.'s (1997) business-IT strategic fit model. Goodhue and Thompson (1995) suggest that technology characteristics, moderate relationships between task and individual characteristics, and system use. And, although the authors theorize the impact of technology characteristics on these relationships as moderating, they go on in this paper to develop these constructs in mediating,⁵ as opposed to moderating, relationships. Chan et al. (1997) carried the moderation concept from theory to analysis. They found that alignment between business strategy and IT strategy was a better predictor of business performance and IS effectiveness than either strategy on its own. In their final analysis, they modeled alignment as a moderating influence on the relationship between IS strategy and business strategy. Several other dominant IS theories (Davis' 1989 TAM model, Doll and Torkzedah's 1991 user participation/involvement model) and the streams of research that extend these models and others (e.g., Hartwick and Barki 1994, p. 461) suggest that moderators are an important avenue of future development, and many calls are noted in the literature. McKeen et al. (1994) noted these calls as far back as 1984 (Ives and Olson 1984) and heralded the investigation of more contingency factors and the expansion of the theoretical complexity of moderated models. These calls are constantly repeated within the field (e.g., Anderson 1985, Vanderslice 1987, Tait and Vessy 1988, Doll and Torkzedah 1989, Sambamurthy and Zmud 1999).

In addition to calls from within the field, there is evidence from other fields that the relatively young

⁴ All articles in *MIS Quarterly*, *Information Systems Research*, *Management Science*, *the Journal of MIS*, *Decision Sciences*, *the Communications of the ACM*, *Information & Management*, and proceedings of ICIS and HICSS up to 1995 and as far back as the journal goes or 1980 whichever is more recent were reviewed by two research assistants and the authors. The choice for the start date of 1980 coincides with the first ICIS conference and has been used by Culnan (1987) as the starting point for mapping the intellectual structure of the IS field.

⁵ A mediating relationship is an intervening construct that sits between two other constructs and passes/mediates the influence of the one variable through to the other.

IS field can expect moderators to grow in prominence. If the IS field follows a similar evolutionary path to that of other fields, such as organizational behavior and psychology (Stone 1988), moderators will likely be used more and assume a prominent role within IS research investigations. This trend is evidenced in the literature⁶ with the number of articles that investigate moderators steadily rising since 1980 and the number of moderators within these articles continually expanding. As the IS field applies more complex theories, its methodologies and analytical approaches, such as the use of moderators, will become more important. Dominant theories are, increasingly, employing moderators in a central role, one that is likely to remain for some time.

Problems Emerging

Although many laud the importance of moderators within theoretical development, others within the IS field have questioned the overall usefulness of contingency research (which is based on moderators). Weill and Olson (1989) did exactly this when they stated that the “highly mixed empirical results” (p. 79) is evidence of a lack of contribution to knowledge (p. 67). Cavaye (1995) suggests that the “literature regularly proclaims” (p. 319) the importance of moderating influences for user participation and IS development project success, but “empirical research continues to provide inconsistent results” (p. 319). These inconsistencies raise doubt in researcher’s minds, bringing into question the value of this work (p. 317) and creating a mistaken impression that moderators are not important.

In fact, this negative conclusion is exactly the opposite of what Cavaye (1995) was arguing, although never stated explicitly. She argued that contingency factors might explain many of these “inconsistent” or “inconclusive” results, listing numerous moderating influences that might better capture this theoretical relationship. Raising the importance of moderators within theoretical development is exactly her point and an underlying message of this paper. To accomplish this, however, we need to be aware of the shortcomings of our current techniques.

⁶ See Figure 1 at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

Current Shortcomings in the Moderator Literature

The difficulties, as expressed within the moderator literature,⁷ are presented next. In particular, we focus on eight, emergent, key problem areas for assessing moderators,⁸ including measurement error and how reliability, summated scales, number of indicators, effect size, sample size, power, and incomplete reporting influence the subsequent empirical analyses and results.

Measurement Error and Reliability. Measurement error is most often assessed first through reliability analysis and then ignored or constrained in the subsequent analysis of the theoretical relationships. The sequence goes as follows: Measurement error assessments typically begin with a Cronbach’s reliability check where reliability levels are determined and poorly convergent items are excluded, then variables are aggregated into a single score (e.g., summed or averaged).⁹ Once this aggregation or reduction to a single item takes place, further assessment of measurement error in any subsequent analysis is impossible because the measurement error has been fixed by the scale construction method. Fixing in this way causes inaccuracies in the subsequent estimates of the theoretical relationships, as will be shown below.

Compounded with this fixed-scale construction is the use of subsequent analytical techniques that cannot assess measurement error. Regression and ANOVA are the preferred methods used to test moderators (see Tables 1 and 2^{7,8}) in IS research, and they, by definition, focus on single measures (scaled or otherwise). Unfortunately, additional tests to assess the discriminant and nomological validity of the individual items within the theoretical model (i.e., main and interaction effects) can never be made. Measurement error as a result of a multi-item scale cannot be assessed within these techniques, and, as a result, issues of multidimensionality, poor items and construct interpretation are hampered.

⁷ See the Literature Review summary at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

⁸ See the Moderators Employed summary at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

⁹ Single-indicator scores also can be used; however, their measurement error cannot be assessed because there is only one item.

Summated Scales. Special emphasis on summated scales is noted here as the practice is often proposed as the solution to the measurement error problem. To overcome the issue of single indicators, many researchers have employed the practice of summing items to create a single indicator for use within regression or ANOVA,¹⁰ with the belief that the resulting summated scale better accounts for the underlying measurement error (Gelderman 1998, same issue for averaged scales; Igbaria and Baroudi 1993). Although this summation practice performs better than single indicators, it can mask measurement error through the two-step process described above in which item aggregation is performed outside of the theoretical context in which the aggregated score is subsequently used. Thus, summated scales, while found to be the standard practice in the IS literature, can be suboptimal in two ways: Equal item reliability and unchanging scale reliability.

In summing items into a single measure, the assumption is made that all items are equally reliable. However, this summing approach, while reducing measurement error, is suboptimal relative to the PLS algorithm. PLS treats each indicator separately, allowing each item to differ in the amount of influence on the construct estimate. Therefore, indicators with weaker relationships to related indicators and the latent construct¹¹ are given lower weightings (Lohmöller 1989; Wold 1982, 1985, 1989), resulting in higher reliability for the construct estimate and thus stronger theoretical development.

Summated scales can also be suboptimal because the reliability estimate of the construct is made in isolation from the theoretical model in which it is to be used. The practice of aggregating items may give a false sense that measurement error has been handled when, in fact, measurement error should be rechecked in the context of the final theoretical model before assurances of minimal measurement error are made. As will be shown in the empirical data section, the two-step approach leads to suboptimal estimates and

yet, as seen from our review, represents the current dominant approach used in IS research.

Number of Items. The next logical extension is to begin employing multiple measures so as to capture and analyze measurement error. Reliability theory tells us that the greater the number of indicators used to measure a construct, the more reliable and accurate the subsequent analyses will be (Carmines and Zeller 1979). Furthermore, reliability increases even faster when higher quality (i.e., more reliable) individual items are used, and hence, higher quality items should always be the goal. However if high quality items cannot be guaranteed, extra items should be proposed and analyzed.

Few moderator studies are employing these high quality or multiple item scales. Although reliability is not always reported, our best guess from a review of the reported data is that the IS field is achieving about 0.70 reliability. From a review of the regression and path analysis articles, which are closest to accounting for measurement error, the average number of indicators observed was three.¹² Most structural equation modeling literature suggests that three indicators should be a minimum and not an average (Bollen 1989). Consequently, when reliability is lower or the number of indicators is small, measurement error is more problematic, which further complicates theoretical model development.

LISREL Solution for Summated Scales. Covariance-based techniques, such as LISREL, can accommodate measurement error but are not necessarily the best alternative for overcoming summated scale problems. These techniques have been shown to be less than ideal under many data conditions for analyzing interaction terms using the product-indicator approach (Bollen and Paxton 1998, p. 267; Li et al. 1998, p. 240; Ping 1996). Comparisons between LISREL and PLS are highlighted in Table 4,¹³ discussed below, and compared later to the PLS product-indicator approach.

¹⁰ See Tables 1, 2, and 3 for list of studies at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

¹¹ Latent constructs are unobserved factors that are reflected by or formed from its measures.

¹² See Table 2 at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

¹³ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

Various LISREL specifications have been conducted in the past (Kenny and Judd 1984, Jöreskog and Yang 1996); however, Ping (1996) recently noted that these covariance-based procedures “may produce specification tedium, errors, and estimation difficulties in larger structural equations models” (p. 166). Part of the difficulty involves the need to calculate and specify in the software the required set of *nonlinear constraints*, which increases exponentially with the number of indicators. In agreement, Bollen and Paxton (1998, p. 267, emphasis added) stated that “the best known procedures for models with interactions of latent variables are *technically demanding*. Not only does the potential user need to be familiar with structural equation modeling (SEM), but the researcher must be familiar with programming nonlinear and linear constraints and must be comfortable with fairly large and complicated models.” In agreement, Li et al. (1998, p. 26) stressed “the need for care in the specification of nonlinear constraints in models” and go on to state that “mistakes can easily be made, the consequences of which may be worse than ignoring the interaction effect in the first place.” Again, these constraints grow exponentially with the number of interaction terms.

To overcome the need to calculate such constraints, Ping (1995, 1996) recommended an *alternative two-step approach* that advocates conducting analysis in two steps, separating measurement and structural model assessments. However, the ability to assess the reliability and validity of individual items using such a two-step approach has been questioned (Fornell and Yi 1992). The two-step approach makes more demanding analytical assumptions concerning both uncorrelated errors and multivariate normality of observed variables, and fails to adjust standard errors for constrained parameters that are necessary so as to anchor scales for interpretation (Bollen and Paxton 1998, p. 280). Thus, the standard errors of estimates for parameters, such as loadings, error terms, and variance of latent product variables, are unknown and *no significance test statistics are available*. Further, the two-step method provides only an *approximation of the true results* and only when somewhat demanding assumptions are met. Those assumptions are that (a) the model is correct and (b) the data are multivariate normal (Li et al. 1998, p. 24), which are two conditions that are often not met.

Even if the preceding conditions are met, there are additional operational issues to consider. In general, when the number of indicators in a model goes beyond 40 or 50, computation tends to *not even converge*. Because of the full information algorithm, the computational demands grow exponentially as one increases the number of items used. In addition, the sample size required to yield stable parameter and model fit estimates must also increase (more on this issue later).

Finally, in contrast to both summated regression and our PLS approach, the LISREL algorithm does not explicitly calculate construct scores. Thus, if a goal of the researcher is to obtain a single best approximation of each construct, LISREL will not provide this as part of the process in estimating the interaction effect. In all, LISREL moderator specifications are technically demanding and not necessarily the complete solution.

Effect Size, Power, Sample Size, and Reporting

Other problems noted in the literature included effect sizes, sample size, power, and incomplete reporting, which are often intertwined with each other and, hence, are handled here collectively so as to highlight the types of problems encountered. Only 21% of the moderators tested in IS studies (Table 1) were found to be significant, which begs the question as to whether the other effects did not exist or were just not detected?

Effect size¹⁴ was not reported for 71% of the significant moderators in the nine regression-based articles

¹⁴ Effect size is the strength of the theoretical relationship found in an analysis and provides an estimation of the degree to which a phenomenon exists in a population. The standard approach for determining the strength of an interaction effect involves contrasting the difference between the squared multiple correlation for the baseline or “main effects” model consisting of the measures that are eventually used to estimate the interaction effects and the interaction model containing the same main effects variables plus the interaction terms. Thus, the interaction model always represents a superset of the baseline model (i.e., the main effects model is hierarchically nested within the interaction model). An explicit calculation of the change in R^2 between these two models provides an estimate of the effect size of the interaction term. Cohen (1988) recommends using the difference in R^2 to calculate an f^2 effect size (this is discussed later in the paper). Caution should be made in using the beta estimate for the interaction term as an estimate of effect size. Only under certain conditions would this regression estimate approximate the effect size. Specifically, you’d need to be

(Table 3). The three remaining moderator estimates, based on Cohen's (1988) guidelines, yielded two small effects (0.036 and 0.050) and a medium effect (0.123). Several things accounted for our inability to report the effect sizes of the interaction terms in the reviewed work, including the use of techniques that restrict an effect size calculation such as split samples, the lack of sufficient information (e.g., standard error) to standardize unstandardized estimates, or omission of effects in the manuscripts for other unspecified reasons. Inconsistent reporting, the latter finding, hinders any attempt to aggregate the field of knowledge (Weill and Olson 1989).

Sample sizes averaged 81.5 for articles using regression or path analysis, and 148 for articles using analysis of variance techniques (Table 1). Statistical power was reported in only four articles with just one of these indicating the recommended 0.80 (Cohen 1988). In general, power is impacted by small effect sizes, small sample sizes, random measurement error, larger standard deviation, and nonnormal data distributions. Given the fact that many of these power-limiting conditions existed within the literature reviewed, one wonders whether an explanation for the null results is that they were just not detected.

In conclusion, moderators have a long history within the IS research field and are gaining prominence in many emerging theories. Despite the ongoing calls to increase investigation into moderating effects, problems have emerged around measurement error, reliability, single item measures, summated scales, inconsistent results, and small overall impact. Few studies have used analytic techniques that can calculate effect size or can present a complete solution (e.g., LISREL). Power levels tend to be low. Meta-analyses are either theoretical (Trice and Treacy 1986, Goodhue 1986) or report inconsistent results (Cavaye 1995). These inconsistent results are

using standardized measures to produce an interaction term, where the calculated interaction term tends to yield low correlations with the individual component parts of the term. In the case of zero correlation, the square of the unstandardized estimate of the beta estimate for the interaction term can represent an approximation of the change in R^2 . This is the situation for our Monte Carlo simulation and thus allows for the convenience of examining the regression estimate as an approximation of the effect size.

often blamed on different operationalizations of the constructs, uncontrolled research factors, inappropriate quantitative studies with little richness in the understanding of the influences, and poorly validated instruments (Cavaye 1995). We believe these problems are opportunities for improvement.

To such an end, a new latent variable modeling approach for analyzing interaction effects is now introduced and its ability to address these problems is discussed. Empirical evidence is then presented from two studies: A Monte Carlo study and an IT-adoption data set. The Monte Carlo study allows us to assess the approach's improvement over previous techniques, while the IT adoption study demonstrates the use of the new approach within an actual IS data set and theoretical framework.

PLS Product-Indicator Approach for Measuring Interaction

To account for the effects of measurement error, a product-indicator approach in conjunction with PLS is proposed. To the best of our knowledge, this represents the first time such a technique has been applied using PLS for assessing interaction effects. Both Chan et al. (1997) and Bergeron et al. (2001), for example, discuss the theoretical implications of moderators as a fit variable but never fully demonstrate the empirical properties of such a product-indicator technique. Chan et al.'s (1997) work comes closest to a moderated model but was missing two crucial main effects paths in the final model analyzed. Paths between each exogenous¹⁵ construct and each endogenous construct, in this model, must be analyzed. When the main effect variables are missing in the analysis, interaction path coefficients are not true interaction effects (Jaccard et al. 1990). Furthermore, the individual items were not centered or standardized, the moderation scores were averaged and the

¹⁵ Exogenous constructs in this situation can be thought of as independent variables, which have no antecedents that are currently being evaluated in the statistical analysis. More simply, if there are no causal arrows leading into the independent variable, then it is an exogenous construct, otherwise it is called an endogenous construct. The distinction is important for issues surrounding estimation.

scores were then used as a formative measure,¹⁶ a sequence that results in uninterpretable coefficients.¹⁷ Consequently, an empirical demonstration of the PLS technique is crucial for understanding an appropriate analytical process in which such errors in execution are avoided.

PLS Appropriateness

The use of PLS has been gaining interest and use among IS researchers in recent years (Compeau and Higgins 1995, Aubert et al. 1994, Chin and Gopal 1995) because of its ability to model latent constructs under conditions of nonnormality and with small to medium sample sizes. It is important to recognize that the operational act of creating product terms by multiplying measures together (to be discussed next) in and of itself is not new. These product terms are used in traditional multiple regression and, as described earlier, have been used for covariance-based solutions using software such as LISREL. Rather, it is the coupling and conceptualization of these product indicators within an easy-to-use PLS context that is new.

Being a components-based structural equation modeling technique, PLS is similar to regression, but simultaneously models the structural paths (i.e., theoretical relationships among latent variables) and measurement paths (i.e., relationships between a latent variable and its indicators). Rather than assume equal weights for all indicators of a scale, the PLS algorithm allows each indicator to vary in how much it contributes to the composite score of the latent variable. Thus, indicators with weaker relationships to related indicators and to the latent construct are given lower weightings, and those varied weightings are carried through to an assessment of the theoretical estimates. In this sense, PLS is preferable to techniques such as single-item regression that assumes error-free measurement, summated regression that assumes equal-weighted measurement and factor score-based regression that assumes constrained measurement

error within the estimates of the theoretical variables (Lohmöller 1989; Wold 1982, 1985, 1989).¹⁸

Our PLS product-indicator approach represents a one-step technique that requires no additional specification of parameter constraints or assumptions of multivariate normality, can be used to estimate large complex models (even when embedded in the middle of a nomological network), and estimates standard errors via resampling procedures (see Chin 1998b for a discussion of resampling procedures in conjunction with PLS). Furthermore, *sample size* (as to be demonstrated by the Monte Carlo study) is not constrained by the number of product indicators as would be the case in LISREL estimations, which require increasingly larger sample sizes as the number of indicators grows. A simple PLS heuristic¹⁹ for main-effects-only models indicates that a sample size of 30 would be a reasonable starting point for the three-construct model discussed here and is independent of the number of indicators used. As shown in Appendix A, the comparable LISREL sample would be 200 and increases quickly as the number of indicators increase (up to 1,820 sample size for 12 indicators). Practically, as the model complexity increases beyond 40 or 50 indicators, the LISREL software may *not even converge*. The PLS approach, in contrast, has been shown to yield computational results for a model with 672 indicators, 21 latent variables, and 200 cases in approximately 1.5 minutes on a 166-MHz Pentium computer (Chin and Newsted 1999, p. 335).

The final point of comparison to be clarified is that the underlying assumption of *uncorrelated error terms* among indicators cannot, by definition, hold true for any moderator analysis (see Kenny and Judd 1984 for derivation). Because they are created through multiplication, the error terms for the product indicators are partially correlated with the error terms for the indicators of the other exogenous constructs.²⁰ While

¹⁶ For a discussion of formative measures, see Appendix D at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

¹⁷ Chan et al. tested the other model but have not published them for public review. Their claim is that the results are similar.

¹⁸ Additional information about the PLS method and how it compares to covariance-based procedures is provided in Appendix A at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

¹⁹ See Appendix A at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

²⁰ This will be discussed in greater detail in the next subsection.

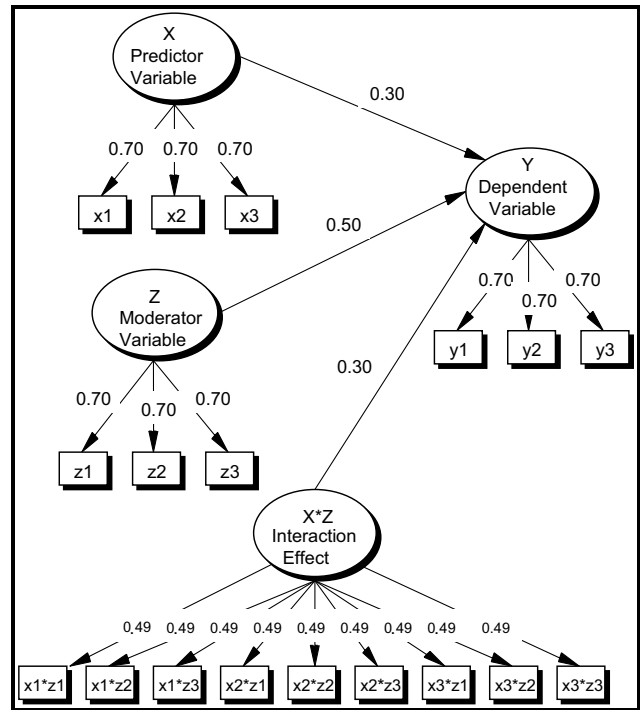
problematic if not accounted for within covariance-based modeling software such as LISREL, these correlations may actually help provide a more accurate estimation of the interaction effect when using PLS. The reason is that there is a known bias in PLS that underestimates the structural effects.²¹ While there are no known ways to estimate the amount of bias or inaccurate estimates in a complex model, we do have formulae to account for this bias in single- and two-construct models (see Chin 1998b, p. 330 for details). In these cases, bias is reduced with more indicators, and hence we can expect the same tempering effects within our Monte Carlo analysis. Whether or not reduced bias continues to be true in the multi-construct case will be assessed in the next section. Overall, we believe that the data conditions within the IS field are likely more aligned with the requirements of the PLS approach than the requirements of the LISREL approach.

PLS Setup

Predictor, moderator, and dependent variables under this PLS approach are viewed as latent variables or constructs, which are ideas that cannot be measured directly. Instead, multiple indicators, or measures, for these latent variables must be obtained. Although it is possible to gather measures in many ways, one example of a measure is a survey question in a data collection instrument. For this analytic technique, each indicator is modeled as being influenced by both the underlying latent variable (i.e., reflective indicators) and error. Product indicators reflecting the latent interaction variables are then created by multiplying the indicators from the predictor and the moderator variables (see Figure 2). This analytical model is consistent with the theoretical model shown in Figure 3. Each set of indicators reflecting their underlying construct or latent variable are then submitted to PLS for estimation resulting in a more accurate assessment of the underlying latent variables and their relationships.

²¹ It will be demonstrated in the next section that this PLS bias is less problematic than regression's underestimation.

Figure 2 Model with Three Indicators per Main Construct and Nine Produce Indicators for the Interaction Construct

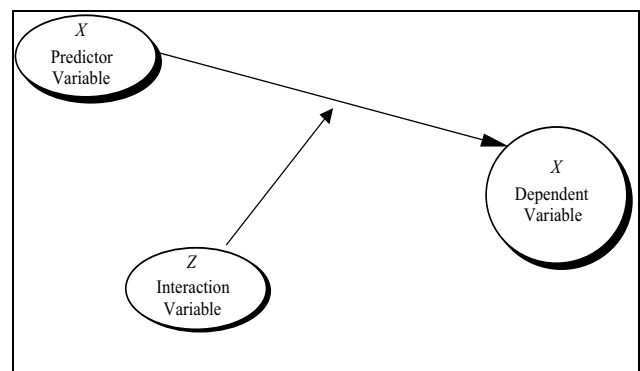


Note. Path coefficients at the levels specified were created through Monte Carlo simulations.

Standardizing or Centering Measures

An important step in undertaking the PLS product-indicator approach is to determine whether indicators must be standardized or centered. Standardizing or centering indicators helps avoid computational errors by lowering the correlation between the product indi-

Figure 3 Comparable Theoretical Model for Analytic Model in Figure 2



cators and their individual components (Smith and Sasaki 1979); consequently, *one* of the techniques must always be used. Furthermore, without such a process, each product term would likely have a different interpretation, limiting the ability of the PLS procedure to accurately estimate the underlying interaction construct. Standardizing or centering the indicators also allows an easier interpretation of the resulting regression beta for the predictor variable. This beta represents the effect expected at the mean value of the moderator variable, which is set to zero.

Standardization is used for reflective measures if it is decided that they can be conceived of as approximately parallel indicators (i.e., equivalent in their measurement of the underlying construct) and no a priori emphasis is given to a particular indicator in the set. Under this situation, all indicators reflecting the predictor and moderator constructs are standardized²² to a mean of zero and variance of one (Jaccard et al. 1990, Aiken and West 1991). This approach can be done for ordinal- and interval-level items, such as Likert-scaled attitudinal items, and must be calculated before submitting the data to PLS. Many statistical packages can save standardized Z scores to a file to facilitate this calculation.

Alternatively, centering can be used to maintain the scale metric (or units of measurement), which might be necessary for theoretical interpretation. If it is decided that some indicators are theoretically more important than others, indicators would only be centered—to achieve a mean of zero—by subtracting the mean from every score. This centering technique is only used if it is felt that the original metric of the items or their variances should be maintained, and usually must be calculated explicitly within a statistical package before submitting the data to PLS. For ratio-level items, it is important to have all items transformed to the same metric in addition to being centered. For example, if you measured temperatures in both Celsius and Fahrenheit, you must

convert them all to the same scale. After that, you need to center each temperature indicator by subtracting the respective means of the converted scales from their respective data values. Otherwise, the estimated latent variable score produced by PLS would be indeterminable and hence uninterpretable.

Calculating Interaction Term Measures

Once the standardized or centered indicators of the predictor variable X and the moderator variable Z are calculated, product indicators are developed by creating all possible products from the two sets of indicators, usually through an explicit multiplication. These product indicators are used to reflect the latent interaction variable. For example, if there are three measures reflecting the main predictor X and three measures for the moderator variable Z , there would be nine measures for representing the interaction term $X * Z$. Graphically, this is depicted in Figure 2. Because any indicator reflecting the predictor X or moderator Z is viewed as interchangeable with another from the same set, any product indicator $x_i * z_j$ would represent a parallel measure of the underlying latent interaction variable $X * Z$.

PLS Estimation

The PLS procedure is then used to estimate the latent variables as an exact linear combination of its indicators with the goal of maximizing the explained variance for the indicators and latent variables. Following a series of ordinary least squares analyses, PLS optimally weights the indicators such that a resulting latent variable estimate can be obtained.²³ The weights provide an exact linear combination of the indicators for forming the latent variable score that is not only maximally correlated with its own set of indicators, as in components analysis, but also correlated with other latent variables according to the structural, or theoretical, model.

Empirical data illustrating the PLS approach is now presented. Because our purpose is to improve

²² Standardizing is accomplished by calculating the mean and standard deviation for each indicator. Then for each indicator score, the corresponding overall mean is subtracted and the result is divided by the respective standard deviation. Centering refers only to subtracting the mean.

²³ For standardized indicators, the default estimation procedure of unit variance/no location should yield similar results to the original scale/with locations option. For centered data, the original scale/with locations should be used.

on the dominant technique found in the literature (i.e., regression and ANOVA), comparisons to single-indicator regression and summated regression are included. The first study presents a simulated Monte Carlo data set where true effects are known and the ability of each technique to estimate these coefficients is shown. The second study presents an IS data set around e-mail adoption with the moderating influence of emotion.

Study 1: Monte Carlo Simulation

To test the efficacy of the PLS product-indicator approach for detecting and estimating interaction effects where measurement error exists, a Monte Carlo simulation study was first executed. PLS, single-indicator regression, and summated regression analytical techniques are compared under varying conditions of effect size, sample size, number of indicators, and measurement error—the influences most prevalent in the literature and central to the interaction term analysis. We begin our presentation of this study by explaining Monte Carlo simulation and follow this with a description of the predetermined population parameters used to assess our PLS product-indicator approach. Because accounting for measurement error is at the heart of the difference between the PLS and regression techniques, loading patterns are first held at a level of 0.70 to provide a baseline comparison, and later varied to extend our understanding. The baseline case and subsequent variations are briefly outlined before the simulation results are shown.

Monte Carlo Simulation

Monte Carlo simulations are typically applied by latent variable/structural equation modeling researchers to ascertain the robustness of statistical estimators (e.g., Chin and Newsted 1999, Chou et al. 1991, Sharma et al. 1989). Monte Carlo simulation refers to a procedure of generating artificial data, based on a specific statistical model that is defined in terms of a stochastic generating mechanism (Noreen 1989). In other words, we create data that conforms to specifically stated model parameters, such as structural paths (e.g., main and interaction effects), factor loadings, and error terms. Two

different approaches have been used to generate data to assess the robustness of latent variable methods. In the Monte Carlo approach, the implied covariance matrix of the observed variables is computed for given values on the parameters in the model and then data are generated on the observed variables from a multivariate distribution having this covariance matrix. Thus, data are only generated for the observed variables and not for the construct level. In the alternative approach, data are first generated for the latent variables according to the relationships specified in the model and then data are generated for the observed variables from the latent variables in the model. This latter approach is better suited to generate data with the distributional characteristics imposed by the model. For this study, we apply the second approach, consistent with the examples and functionality available in the software package PRELIS 2.14 (Jöreskog and Sorbom 1993).

Monte Carlo Population Parameters

Using PRELIS 2.14 (Jöreskog and Sorbom 1993), data were generated to conform to an underlying population model where the standardized beta of X on Y was 0.30, the beta of Z on Y was 0.50, and the interaction effect ($X * Z$) was 0.30. The model is shown in Figure 2. Indicators for all primary constructs, or latent variables, were modeled as having factor loadings of 0.70 because this is a minimum standard of the IS literature.²⁴ Thus, the true scores for the main and interaction effects are known. The goal here is to determine how well the PLS product-indicator approach detects and recovers (i.e., estimates) the true effects under conditions of measurement error.

Monte Carlo Design—The Baseline Case and Subsequent Cases of Comparison

The Monte Carlo design is intended to achieve several objectives. The first objective, as our baseline case, is to assess how well the new PLS product-indicator method performs at retrieving the true population parameter. As such, sample sizes are varied at 20, 50,

²⁴ As an initial test near the end of the paper, we do examine the PLS product-indicator performance with different loadings (and, hence, different reliability).

100, 150, 200, and 500 cases, and the number of indicators per primary construct of X , Z , and Y are varied at 1, 2, 4, 6, 8, 10, and 12. For each cell in this completely crossed design, 500 simulations were performed. For example, in the cell representing sample size 50 and 4 indicators per construct, 500 data sets were generated consisting of sample sizes of 50 where each case had 4 indicators for each of the 3 constructs X , Z , and Y (factor loadings of 0.70), and 16 product indicators for $X * Z$ (factor loadings of 0.49).

A second objective, after examining the absolute performance of the new approach, is a comparison to regression-based estimates. A comparison between the PLS product-indicator approach and single-indicator regression is made, which represents a simple initial point of departure to establish the relative effectiveness of the product-indicator approach in accommodating measures with error. In addition, the comparison between the PLS product-indicator approach and the common practice of summing scales within a regression analysis (the often suggested solution for addressing measurement error) is also undertaken. Using the data generated from the Monte Carlo simulation, summated scales are created and employed in additional regression analyses. The path estimates for the summated regression analyses are then compared to those obtained via the PLS product-indicator approach.

As a final objective, unequal loadings (a situation more typical among research studies) are compared with the Monte Carlo data for the PLS product-indicator and summated regression approaches. While the next analysis is not exhaustive, a brief simulation is provided to see how both techniques perform under conditions of heterogeneous loading patterns (i.e., not all 0.70).

Having described the Monte Carlo objectives, the results are now presented. We begin with a discussion of significant levels and the influences of sample size and number of indicators followed by a discussion of the ability to estimate the known/true underlying path estimates.

Significance

Tables 5, 6, and 7 provide the results and significance tests for the 6×7 crossed design in which PLS runs

were made using PLS-Graph version 3.0 (Chin 2001). As noted, a single indicator multiple regression forms the initial baseline of comparison for how well the PLS product-indicator approach performs. Results of using single-indicator regression with varying sample sizes are provided in the column of Table 7 labeled "one item per construct." When using PLS, the case of one indicator per construct is identical to performing a multiple regression with a single-indicator measure.

Accurate estimates are invaluable to IS research and only achieved if analytical techniques can both detect and estimate the true scores. Advocating that detection by itself is sufficient (i.e., significance) is incomplete because the goal is to accurately detect the true score if it exists. Hence, we will first examine the pattern of significance and then, with this pattern in mind, consider the accuracy of the estimations. Only then can conclusions be drawn.

Significance levels achieved within each cell from the single-indicator regression and PLS analyses are shown in Table 5. In general, small sample sizes or few indicators produced few significant estimations at $p \leq 0.05$. For example, multiple regression using a single indicator, as represented by the first column, did not detect a significant effect. The predictor X term and the moderator Z term achieved significance sooner at sample sizes of 100 and 50, respectively. Consequently, PLS's ability to estimate the noninteraction terms that are either more reliable or have larger effect sizes ($z = 0.50$) is reaffirmed. The single-indicator approach appears inadequate for estimating interaction terms because no results were significant.

Let us pause for a moment to understand three possible patterns of significance. These ideal patterns of results, as shown in Table 6, might be formed if (i) only the number of indicators has an influence, (ii) only sample size has an influence, or (iii) both have an equal influence. A solid line is drawn to distinguish where the results change from nonsignificant to significant.

If only the number of indicators had an influence, then a solid vertical line would be observed as shown in Table 6(i), because no matter how large the sample size grew, the significance within a column would not change. If only sample size had an influence, then a solid horizontal line would be observed as shown in

Table 5 Tests of Significance for the Mean Estimates Shown in Table 7 (500 Runs in Each Cell)

Predictor: X (.30)
(dashed line)

Sample Size	1	2	4	6	8	10	12
20							
50			**	**	*	*	*
100		**	**	**	**	**	**
150		*	**	**	**	**	**
200		*	**	**	**	**	**
500		*	**	**	**	**	**

Moderator: Z (.50)
(solid line)

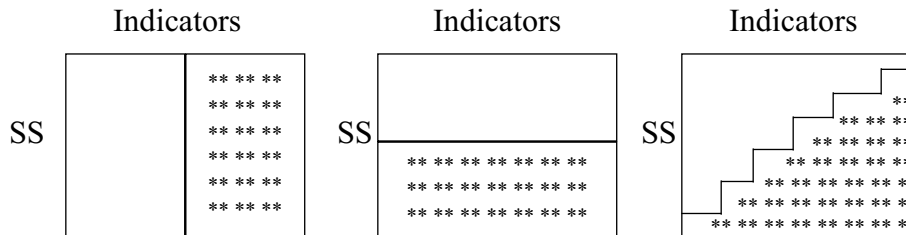
Sample Size	1	2	4	6	8	10	12
20			**	**	**	**	**
50		*	**	**	**	**	**
100		**	**	**	**	**	**
150		*	**	**	**	**	**
200		*	**	**	**	**	**
500		*	**	**	**	**	**

Interaction: X*Z (.30)
(double line)

Sample Size	1	2	4	6	8	10	12
20							
50							
100			*	**	**	**	**
150		*	**	**	**	**	**
200			**	**	**	**	**
500			**	**	**	**	**

Note. * $p < 0.05$ (one-tailed t value: 1.66, $df = 499$).
 ** $p < 0.01$ (one-tailed t value: 2.36, $df = 499$) (outlined by the dark line).

Table 6 Possible Patterns of Significance



- (i) Only number of indicators has an effect
- (ii) Only sample size has an effect
- (iii) Both have an equal influence

Note. ** = significance.

Table 6(ii), because no matter how many indicators were used the significance within a row would not change. Finally, if both had an equal influence, then a tiered, diagonal line would be observed as shown in Table 6(iii), because significance would change incrementally by both rows (sample size impact) and columns (indicator impact). These patterns will be compared throughout the results.

Sample Size. As a general observation regarding Table 5, it can be seen that smaller sample sizes did not produce significant results for many of the combinations across each of the predictor, moderator and interaction terms. The sample size of 20 failed to detect the true effect in all combinations. Interaction term results for the sample size of 50, also, were not significant. Small sample sizes clearly should be avoided when analyzing moderator variables.

For larger sample sizes, there appears to be a threshold after which an increase in sample size does not change the significance level.²⁵ The dashed, solid, and double lines designate the division between non-significant and significant at 0.01, a conservative level that balances both detection and estimation results as will be seen below. The first two parts of Table 5 for predictor *X* and moderator *Z* appear to have flat, horizontal lines, and the last part for the interaction term appears to have a more slightly tiered pattern. Accordingly, these patterns reconfirm that, for the PLS product-indicator approach, sample size is more influential in determining significance for noninteraction terms and terms with large effect sizes, but the patterns also suggest that both sample size and the number of indicators are influential in determining significance for interaction terms. The equality of these latter influences will be revisited again when reviewing path estimation.

Indicators. The influence of indicators is promising. More indicators were generally significant, and it is clearly seen that fewer indicators (under four) often lead to nonsignificant results until either a larger sample size was used, such as 150 or over, or a true effect

was larger, such as $Z = 0.50$. Significance at 0.01 was achieved for the interaction term at 6 indicators—100 sample size, and 4 indicators—150 sample size, representing ideal threshold values for sample size and number of indicators in PLS.

Taken together, the results of significance levels for sample size and number of indicators suggest that appropriate detection of interaction terms require sample sizes of 100–150 and 4 or more indicators for each predictor and moderator constructs. The ideal threshold combinations outlined above imply that increasing the number of indicators when analyzing moderators is just as important as gathering more data. Determining significance is an important first step and essential when calculating the correct path estimates. The ability of the techniques to capture true path estimates is now explored.

Path Estimation

After reviewing significance, true score estimation can be addressed. Table 7 provides the path estimations and standard error results for the different combinations of sample size and number of indicators.

Single-Indicator Regression

Keeping in mind that a perfect estimation procedure should result in 0.30 for the x to y path, 0.50 for the z to y path, and 0.30 for the $x*z$ to y path, note that the Table 7 single-indicator regression results (Column 1), which are under conditions of measurement error, consistently (and significantly) underestimated these true effects. The reason, as discussed earlier, is that single-indicator regression does not explicitly take into account the attenuating effects of measurement error. At a sample size of 500, for example, path estimations for single-indicator regression are one-half to one-third of the true effects. Estimation of the interaction term, keeping in mind that none yielded significant results, was only 0.098—far from the 0.30 true score. Therefore, the single-indicator regression approach never seems to yield the true effects even when larger sample sizes are used.

Ideal PLS Interaction Terms

Estimation with the PLS product-indicator approach is promising. Specifically, the number of indicators

²⁵ Note, however, that other statistical benefits could result from this increase in sample size, such as greater statistical precision. See Appendix B for a discussion of those benefits.

Table 7 PLS Path Estimation from Monte Carlo Simulation (500 Runs per Cell)

Sample size	Indicators per construct						
	one item per construct ^a	two per construct (4 for interaction)	four per construct (16 for interaction)	six per construct (36 for interaction)	eight per construct (64 for interaction)	ten per construct (100 for interaction)	twelve per construct (144 for interaction)
20	x → y 0.186 (0.276) z → y 0.330 (0.286) x*z → y 0.162 (0.352)	x → y 0.186 (0.276) z → y 0.330 (0.286) x*z → y 0.162 (0.352)	x → y 0.215 (0.250) z → y 0.334 (0.264) x*z → y 0.250 (0.370)	x → y 0.219 (0.237) z → y 0.335 (0.251) x*z → y 0.276 (0.377)	x → y 0.217 (0.238) z → y 0.341 (0.254) x*z → y 0.267 (0.402)	x → y 0.220 (0.237) z → y 0.341 (0.251) x*z → y 0.305 (0.375)	x → y 0.217 (0.223) z → y 0.323 (0.259) x*z → y 0.308 (0.424)
50	x → y 0.130 (0.220) z → y 0.241 (0.298) x*z → y 0.104 (0.248)	x → y 0.195 (0.187) z → y 0.326 (0.218) x*z → y 0.172 (0.236)	x → y 0.232 (0.153) z → y 0.386 (0.159) x*z → y 0.274 (0.186)	x → y 0.251 (0.139) z → y 0.390 (0.154) x*z → y 0.276 (0.232)	x → y 0.263 (0.131) z → y 0.403 (0.142) x*z → y 0.304 (0.216)	x → y 0.264 (0.124) z → y 0.396 (0.142) x*z → y 0.320 (0.230)	x → y 0.267 (0.126) z → y 0.418 (0.134) x*z → y 0.333 (0.221)
100	x → y 0.140 (0.186) z → y 0.247 (0.270) x*z → y 0.114 (0.210)	x → y 0.208 (0.130) z → y 0.326 (0.195) x*z → y 0.169 (0.181)	x → y 0.256 (0.091) z → y 0.382 (0.143) x*z → y 0.256 (0.120)	x → y 0.260 (0.100) z → y 0.410 (0.120) x*z → y 0.282 (0.119)	x → y 0.274 (0.078) z → y 0.431 (0.097) x*z → y 0.304 (0.112)	x → y 0.283 (0.080) z → y 0.434 (0.097) x*z → y 0.308 (0.117)	x → y 0.276 (0.080) z → y 0.444 (0.092) x*z → y 0.332 (0.087)
150	x → y 0.145 (0.174) z → y 0.243 (0.269) x*z → y 0.102 (0.214)	x → y 0.256 (0.091) z → y 0.382 (0.143) x*z → y 0.256 (0.120)	x → y 0.245 (0.086) z → y 0.397 (0.122) x*z → y 0.242 (0.100)	x → y 0.261 (0.073) z → y 0.417 (0.104) x*z → y 0.277 (0.080)	x → y 0.265 (0.070) z → y 0.440 (0.085) x*z → y 0.291 (0.078)	x → y 0.271 (0.070) z → y 0.448 (0.080) x*z → y 0.298 (0.065)	x → y 0.280 (0.062) z → y 0.453 (0.075) x*z → y 0.303 (0.070)
200	x → y 0.151 (0.164) z → y 0.246 (0.263) x*z → y 0.102 (0.211)	x → y 0.199 (0.120) z → y 0.328 (0.183) x*z → y 0.176 (0.143)	x → y 0.243 (0.081) z → y 0.397 (0.118) x*z → y 0.242 (0.082)	x → y 0.259 (0.068) z → y 0.426 (0.092) x*z → y 0.267 (0.063)	x → y 0.273 (0.072) z → y 0.432 (0.096) x*z → y 0.300 (0.108)	x → y 0.275 (0.059) z → y 0.448 (0.073) x*z → y 0.291 (0.058)	x → y 0.280 (0.056) z → y 0.456 (0.067) x*z → y 0.300 (0.049)
500	x → y 0.146 (0.160) z → y 0.246 (0.258) x*z → y 0.098 (0.206)	x → y 0.198 (0.109) z → y 0.328 (0.176) x*z → y 0.165 (0.142)	x → y 0.242 (0.069) z → y 0.396 (0.110) x*z → y 0.222 (0.087)	x → y 0.257 (0.057) z → y 0.424 (0.084) x*z → y 0.248 (0.064)	x → y 0.268 (0.047) z → y 0.441 (0.068) x*z → y 0.262 (0.051)	x → y 0.271 (0.044) z → y 0.452 (0.058) x*z → y 0.269 (0.048)	x → y 0.278 (0.041) z → y 0.458 (0.053) x*z → y 0.277 (0.043)

Note. $x \rightarrow y$ refers to the mean of 500 path estimates from predictor x to criterion y (true score = 0.30; below dashed line denotes significance).
 $z \rightarrow y$ refers to the mean of 500 path estimates from the moderator variable z to criterion y (true score = 0.50; below solid line denotes significance).
 $x * z \rightarrow y$ refers to the mean of the 500 path estimates for the interaction effect of z on the path from x to y (true score = 0.30; below double lines denote significance).

Significance is denoted through the bolded lines with values below the lines being significant at 0.01. Dashed line is predictor X , solid line is moderator Z , and double line is interaction term $X * Z$.

^aSame as single-indicator multiple regression (population standard errors are within parentheses).

has a slightly greater positive impact on results than do larger sample sizes. This conclusion is based on the observations that, as shown in Table 7, the interaction term estimation (values below the double lines are significant) approaches the 0.30 true score at 8 indicators—100 sample size within 10% of the true score (i.e., above 0.27) and is close to estimation at 6 indicators—150 sample size. These cell combinations are slightly higher than the threshold ideals suggested above, with the 4 indicators—150 sample size combination falling about 15% away from the true score.

Number of Indicators or Sample Size

In the cells above and to the left of the ideal thresholds identified above, path estimations are well below the true score by at least 15%–20%. In the cells below and to the right of these threshold points in the table, the influence of sample size and number of indicators can be evaluated by the patterns of estimation. Increasing the sample size (i.e., estimations further down the columns) does not improve on these estimations and, in fact, can make them worse. For instance, the interaction path estimations for 6 and 8 indicators at a 500 sample size dropped to 0.248 and 0.262,

respectively, from 0.261 and 0.304 values at the thresholds. This pattern is generally consistent throughout the other columns as well.

Increasing the number of indicators (i.e., estimations across the rows), however, does improve estimations more consistently. Within a row, each increase in indicators generally produces an increase in the estimation, consistently moving toward or maintaining the true score. Thus, more indicators help estimate the “true” parameter more closely.

Power

Sample size alone did not appear to help much at uncovering the true parameter for interaction terms, and hence a strategy of gathering more data (i.e., increasing sample size) might not be helpful. Although this was observed in the data, it is important to recognize how sample size and the number of indicators contribute to statistical estimates through power. For a discussion of these influences on power, see Appendix B.²⁶

Inaccurate True Score Estimates

As seen in the literature review, attenuation of the estimates through measurement error is a condition that creates inaccurate true score estimates. This condition exists in the Monte Carlo data and had an effect on single-indicator regression, while only having a slight dampening effect on PLS estimates.

Single-Indicator Regression Biases. Estimation of the interaction term of $X * Z$ for the single-indicator regression analysis resulted in estimations around 0.10 (Table 7) when they should be 0.30. Such results could leave the researcher with the impression that an interaction effect is much smaller than its true score, or even that the interaction effect is possibly nonexistent. This interaction term estimation, in particular, would generally represent the worst among the estimations, because the interaction reliability will be necessarily smaller because it is a product of the reliabilities of the predictor and moderator indicators.

PLS Biases. While much better than single-indicator regression, we see that PLS still tends to underestimate the structural paths that connect constructs. At the same time, PLS tends to overestimate the measurement paths connecting constructs to their indicators. Loadings in these results are overestimated as Table 8²⁷ displays. The true loadings were set in the Monte Carlo analysis at 0.70 for the main-effects constructs of predictor X and moderator Z , and by implication at 0.49 for the interaction construct of $X * Z$. If we consider from Table 5 only the loadings for significant estimates, we can see that the loadings in Table 8 tend to be inflated by more than 10% in the two- and four-indicator situation. And, as conjectured by Chin (1995, p. 319), it is not until we use 10 to 12 indicators that a more accurate loading estimate is reached.

Thus, contrasted to other causal modeling techniques (e.g., LISREL), PLS tends to be more conservative in its estimates of theoretical (i.e., structural) paths and more positively biased towards its loading estimates. This implies caution against putting too much emphasis on PLS loadings when there are few indicators (i.e., ≤ 8). Among the significant estimates, we do see that, on average, the estimates for the interaction construct of $X * Z$ were closer to the 0.30 true effect than the estimates of predictor X to their true effect of 0.30. Our interaction construct thus had reduced estimation bias. With the initial results of the baseline Monte Carlo study in mind, a short exploration of two further comparisons—summed scales within regression and heterogeneous loadings—are undertaken.

Summed Regression—with Baseline Data

Using the same Monte Carlo data as in the baseline case, the practice of summing scales was employed for the 2- through 12-indicator cases. Multiple regression was then used to provide significance levels and true-score estimates. Interaction term results from this summed regression are shown in Table 9. These significance results demonstrate patterns more similar to the PLS product-indicator approach patterns than the single-indicator regression patterns (Table 7). However, summed regression still consistently under-

²⁶ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

²⁷ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

Table 9 Monte Carlo Path Estimation for Interaction Term ($X * Z$) Using Summated Regression (500 Runs per Cell) with Means, Population Standard Errors in Parentheses, and t -Stats, Respectively

Sample size	Indicators per construct					
	two per construct (4 for interaction) (std. Error) t value	four per construct (16 for interaction)	six per construct (36 for interaction)	eight per construct (64 for interaction)	ten per construct (100 for interaction)	twelve per construct (144 for interaction)
20	0.146 (.290) 0.502	0.205 (.259) 0.792	0.232 (.236) 0.982	0.229 (.248) 0.926	0.240 (.237) 1.015	0.243 (.231) 1.053
50	0.151 (.210) 0.717	0.210 (.159) 1.324	0.216 (.150) 1.436	0.238 (.145) 1.646	0.242 (.144) 1.688	0.268 (.130) 2.066
100	0.150 (.180) 0.830	0.212 (.130) 1.631	0.233 (.112) 2.070	0.249 (.090) 2.509	0.250 (.103) 2.435	0.270 (.091) 2.965
150	0.212 (.130) 1.631	0.209 (.121) 1.730	0.238 (.096) 2.480	0.255 (.083) 3.079	0.255 (.082) 3.118	0.260 (.079) 3.288
200	0.160 (.155) 1.035	0.205 (.105) 2.038	0.236 (.089) 2.668	0.252 (.095) 2.669	0.257 (.077) 3.314	0.270 (.064) 4.244
500	0.157 (.149) 1.054	0.211 (.098) 2.158	0.236 (.075) 3.159	0.250 (.061) 4.072	0.257 (.059) 4.365	0.265 (.052) 5.094

Note. True score = 0.30. Significance is denoted through the bolded double lines with values below the lines being significant at 0.01. $t = 2.36$, $P \leq 0.01$, one-tail.

estimated the true theoretical interaction term scores by more than 10% in 41 of 42 cells. Of the significant results, most are substantially below the correct score.

These conclusions are drawn, in part, from the significance levels and estimates shown in Table 9, where double lines denote patterns of significance for summated regression. Any interaction term value below the double line is significant at 0.01. The pattern of significance is somewhat similar to the PLS product-indicator approach, but less effective in detection with 8 indicator—100 sample size and 6 indicator—150 sample size cells being identified as ideal points. All values for the 2 and 4 indicators are not significant, and the 6 indicator—100 sample size cell did not emerge as significant as it was in the PLS baseline case.

Beyond the fact that summated regression detects the interaction true score at later points than does PLS, we can also compare the accuracy of both esti-

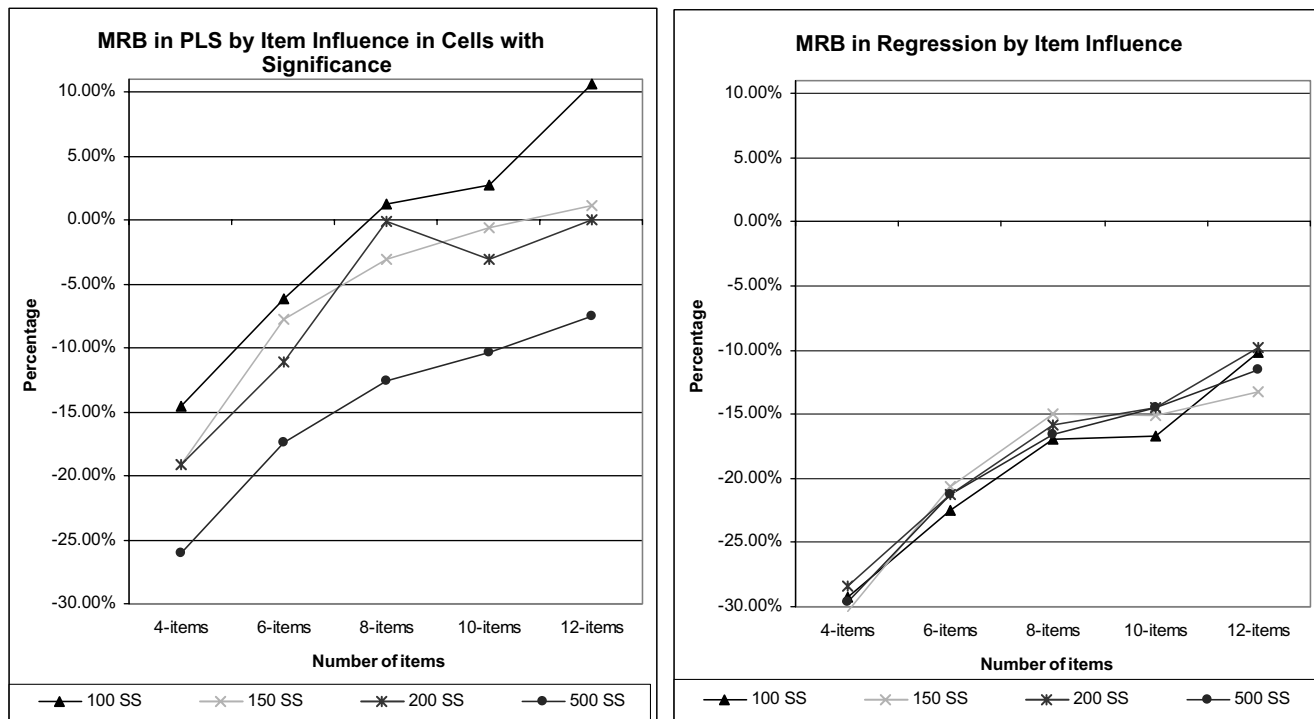
mates from the true value of 0.30 via the *mean relative bias* (MRB). Analogous to the formula provided by Reinartz et al. (2002), the average percentage bias for the $t = 500$ runs from the true population estimate X_{pop} of 0.3 can be determined as follows:

$$MRB = 100 * \frac{1}{t} \sum_{i=1}^t \frac{X_{pop} - X_i}{X_{pop}}$$

From Figure 4, no summated regression estimate reached the true interaction score of 0.30, and only two cells (12 indicators—200 and 100 sample sizes) came within 10% of that score. Most of the PLS estimates reached the 10% threshold at 6 indicators and were within the 5% range at 8 indicators.²⁸

²⁸ The sample size 500 is the only anomaly, which, although performing better than summated regression, was lower than the other sample sizes. We are unclear as to this difference beyond simulation artifact.

Figure 4 Mean Relative Bias in PLS and Regression Interaction Terms Highlighting the Influence of the Number of Items and Sample Size (0% Represents Zero Bias in the True Score Estimation)



In contrast to the interaction term results, the ability to detect the direct effects of predictor X and moderator Z were identical for both summated regression and PLS.²⁹ For strong effect sizes ($Z = 0.50$), both procedures were able to detect an effect at 4 indicators—50 sample size. For the moderate effect size ($X = 0.30$), more data (sample size ≥ 100) or more indicators (4+) was required to obtain significance. But these moderate effect sizes are not consistently estimated (i.e., within 10% of the true score) until eight or more indicators are used. Small sample sizes performed poorly once again until either a large effect size existed ($Z = 0.50$) or a large number of indicators were used.

Summated Regression—with Heterogeneous Loadings

To this point in the analysis, the data set created with the Monte Carlo simulation maintained a 0.70

loading pattern, which reflects a typical level found within the IS literature and provided a consistent benchmark for comparing sample size, effect size, and number of indicators. Comparing the summated regression-based technique to the new PLS product-indicator approach under varying loading patterns is the objective of the next simulation. The results of this comparison are shown in Table 10.

In this analysis, loadings were varied for the independent X variable and the moderating Z variable. By varying these loadings, conditions of heterogeneous constructs were simulated, reflecting the existence of more and less measurement error among a set of items, as shown in Table 10, first column. The dependent Y variable was maintained at 0.70 loadings to simplify the design. Furthermore, based on the previous Monte Carlo results, sample size was set at 100 and the number of indicators kept at either 6 or 8, which was the ideal range in the baseline case balancing both significance and accurate estimation. Average regression scores were calculated for the

²⁹ Because of page limitations, these results are not presented but are available from the authors upon request.

Table 10 Monte Carlo Simulation Comparing the Impact of Heterogeneous Loadings on the Interaction Estimate (in Bold) for the PLS Product Indicator Approach and Regression Using Averaged Scores with Population Standard Errors in Parentheses (Sample Size of 100, 500 Runs per Cell, 8 and 6 Indicators)

	Factor Loading Patterns for 6 items - pattern repeated for both X and Z constructs ^a	PLS Product Indicator Estimates (Std. Error)	Regression Estimates Using Averaged Scores (Std. Error)
A1	2 at .80, 2 at .70 2 at .60	x*z--> y 0.285 ** (0.108)	x*z--> y 0.234 * (0.112)
A2	3 at .80 3 at .70	x*z--> y 0.288 ** (0.091)	x*z--> y 0.246 ** (0.102)
A3	3 at .80 3 at .60	x*z--> y 0.283 ** (0.112)	x*z--> y 0.228 * (0.115)
A4	2 at .80, 2 at .60 2 at .40	x*z--> y 0.302 ** (0.123)	x*z--> y 0.218 * (0.124)
A5	3 at .80 3 at .40	x*z--> y 0.296 ** (0.123)	x*z--> y 0.213 * (0.128)
A6	3 at .70 3 at .60	x*z--> y 0.280 * (0.145)	x*z--> y 0.221 * (0.124)
A7	2 at .70, 2 at .60 2 at .30	x*z--> y 0.304 * (0.129)	x*z--> y 0.189 * (0.144)

	Factor Loading Patterns for 8 items - pattern repeated for both X and Z constructs ^a	PLS Product Indicator Estimates (Std. Error)	Regression Estimates Using Averaged Scores (Std. Error)
B1	4 at .80, 2 at .70 2 at .60	x*z--> y 0.293 ** (0.121)	x*z--> y 0.250 ** (0.102)
B2	4 at .80 4 at .70	x*z--> y 0.297 ** (0.114)	x*z--> y 0.234 * (0.109)
B3	4 at .80 4 at .60	x*z--> y 0.306 ** (0.107)	x*z--> y 0.251 * (0.099)
B4	4 at .80, 2 at .60, 2 at .40	x*z--> y 0.311 ** (0.112)	x*z--> y 0.240 * (0.109)
B5	6 at .80 2 at .40	x*z--> y 0.318 ** (0.124)	x*z--> y 0.223 * (0.118)
B6	4 at .70 4 at .60	x*z--> y 0.305 ** (0.115)	x*z--> y 0.234 * (0.112)
B7	4 at .70, 2 at .60 2 at .30	x*z--> y 0.314 ** (0.134)	x*z--> y 0.224 * (0.118)

x*z --> y refers to the mean of the 500 path estimates for the interaction effect of z on the path from x to y (true score = 0.30).

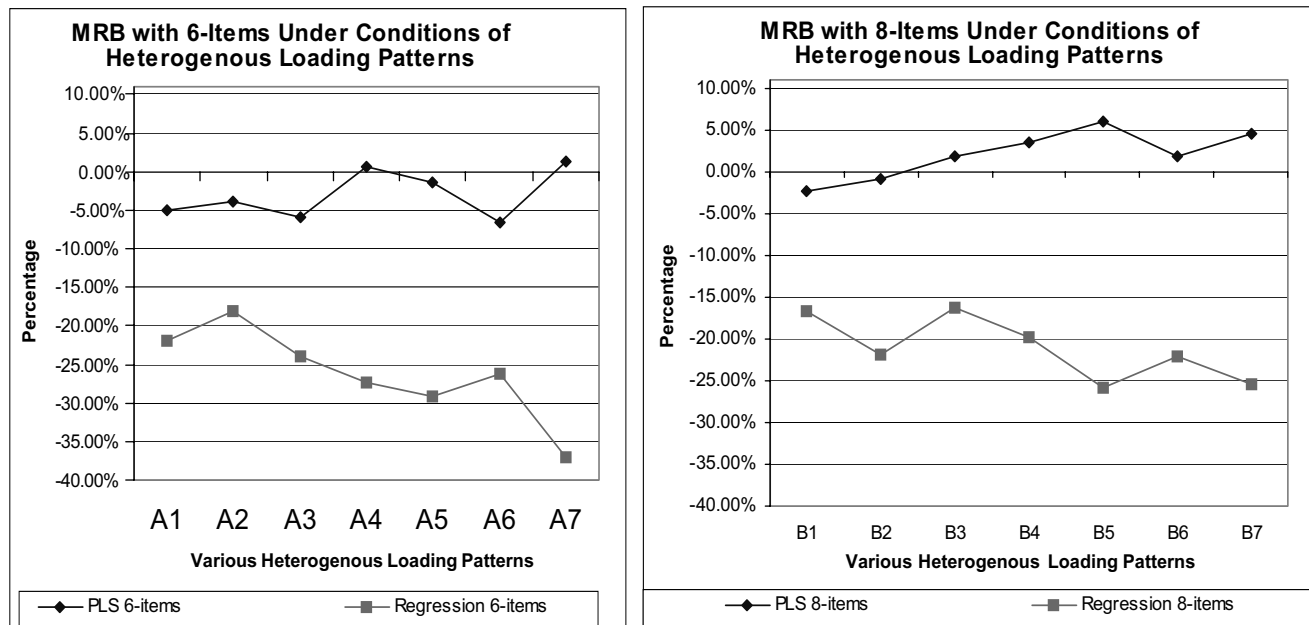
^aDependent variable loadings are held at 0.70 to simplify the design.

*Significant at 0.05; **significant at 0.01.

X and Y variables by averaging the indicators, respectively, and were calculated for the moderating Z variable by multiplying the averaged X and Y values. Even though scores could have been aggregated through various techniques, this method produced the best regression results.

The results in Table 10 show that in all instances of heterogeneous loadings for the eight-indicator situation, the PLS product indicator approach *consistently estimates* the 0.30 true effect of the interaction term (as seen in the bolded values and the MRB values within 5% of the true estimate, Figure 5), and these differ-

Figure 5 Comparison of PLS and Regression Interaction Terms MRB Under Conditions of Heterogeneous Item Quality (0% Represents Zero Bias in the True Score Estimation)



ences are not statistically significant. The summated regression approach, on the other hand, was both *significantly different* and *consistently underestimates* the 0.30 true effect by 16%–37%. No matter whether loadings vary a little as in the case with 0.80 and 0.70 loading combinations or vary a lot as in the case with 0.70, 0.60, and 0.30 loading combinations, PLS performed well, demonstrating its ability to handle measurement error and produce consistent results. For the six-indicator situation, as expected, the estimates tended to be lower. But again, we see an improvement of using PLS over summated regression. Here the estimates using summated regression ranged from 0.19 to 0.25, while the lowest PLS estimate was at 0.28. As in the homogeneous case, the patterns for the direct effects of predictor *X* and moderator *Z* under the summated regression approach were similar to earlier results.³⁰ This, then, leads us to conclude that when the reality of varying loadings patterns and lower reliability emerge, PLS is more accurate in its estimates of true scores than is summated regression.

³⁰ Because of page limitations, these results are not presented but are available from the authors upon request.

Summary

Taking all the results from sample size, number of indicators, effect size, and reliability into account, a minimum sample size of 150 with 4 indicators or a minimum sample size of 100 with 6 indicators appears best to balance the trade-offs for detection and accurate estimate. Researchers should always strive for the highest reliability possible in their measures. Unfortunately, lower or varied reliability is a reality of research in the IS field, one that is not likely to go away. The PLS product-indicator approach, demonstrated above, provides researchers with a technique that allows them to manage this reality. Variability in individual item reliability did not influence the PLS estimates in our eight-indicator simulation. With these threshold combinations, structural path estimates will be within 10% of the true effects, as shown in Table 7, and the overestimation of the measurement paths will be kept to a minimum, as shown in Table 8.

This conclusion, in combination with our findings in the literature review, may provide a possible explanation for the poor results of the past moderator

studies. Recall that, in our literature review, studies examining moderators employed an average sample size of 81.5 and an average number of three indicators. PLS improved on single-indicator measures and summated scales, and handled varying effect sizes, sample sizes, power levels, number of indicators, and reliabilities. Although we cannot conclude from our analysis how other effect sizes, loadings, and reliabilities would fare, these should be tested in the future to get an even greater appraisal of how PLS and summated regression estimates may vary. Smaller sample sizes and fewer indicators using the PLS product-indicator approach would likely be appropriate when using more reliable indicators (i.e., loadings higher than 0.70), but we cannot estimate by how much the estimates would rise.

Now that we have demonstrated through this Monte Carlo analysis that the PLS approach improves on regression-based techniques, the role of the PLS approach within IS theory is explored. Empirical data for the moderating influence of enjoyment on the well-known relationship between perceived usefulness and intention to use is now presented.

Study 2: The Moderating Effect of Enjoyment on the Perceived Usefulness/IT-Adoption Intention Relationship

This section presents the PLS product-indicator approach as applied to detecting the interaction effect of enjoyment on the perceived usefulness/IT-adoption intention relationship. In Davis' (1989) original presentation of this model, perceived usefulness and perceived ease of use were modeled as having direct effects on adoption intention. Later, Davis et al. (1992) note the difference between extrinsic and intrinsic sources of motivation to computer use in the workplace. While usefulness, an extrinsic source of motivation, had a significant effect on adoption intention, enjoyment, an intrinsic source of motivation defined as the extent to which the activity of using the computer is perceived to be enjoyable in its own right apart from any performance consequences, was felt to also have a direct effect. Their study found

that both perceived usefulness and enjoyment mediated the influence of perceived ease of use on intention. Thus, perceived ease of use was not included in predicting intention for this analysis.

The question becomes whether enjoyment also moderates the usefulness to the intention relationship. Consider the following example to illustrate the potential role of enjoyment as a moderator. The TAM model states that the stronger a person's belief in the usefulness of an information technology (IT), the more he/she would intend to use it. Yet, we also believe that the impact of this belief on IT-usage intention is *negatively* moderated by the level of enjoyment the individual has during his or her use of the IT. In essence, when the usage experience is more enjoyable, the impact of perceived usefulness on future intention to use is lower. Conversely, the less enjoyable one perceives the IT to be, the stronger the impact of one's perception of usefulness on intention to use. This phenomenon is based on a cognitive consistency argument in which the underlying theory is that when IT usage is extremely enjoyable, instrumental issues, such as perceived usefulness, ought not to come into one's decision-making criteria for future usage. In fact, for those people whose predominant purpose is enjoyment, more usefulness may be considered a detrimental feature, thereby negatively impacting intention to use. Thus, all else being equal, if we had two different groups of people in which the first group perceived the IT to be highly enjoyable and the second group perceived it to be highly unenjoyable, we would expect a low to negative correlation between perceived usefulness and IT-usage intention for the first group, and a high correlation for the second group. In this scenario, the dependent variable (Y) would represent IT-usage intention, and the predictor (X) and moderator (Z) variables would represent perceived usefulness and enjoyment, respectively.

Chin and Gopal (1995) similarly found that both enjoyment and relative advantage (which uses identical items to perceived usefulness) had an effect on group support system adoption intentions. In their article, they state "that there is also the possibility of interaction effects among the constructs that were not taken into account in this study. For example, Davis

et al. (1992) indicated that a positive interaction might exist between enjoyment and usefulness. Because of its similarity to usefulness, the relative advantage construct used in this study may also have an interaction effect with enjoyment" (p. 58).

To test the possibility of such an interaction effect, the perceived usefulness and enjoyment items were used to examine the adoption intention of electronic mail. See Appendix C³¹ for construct definition, items used, and organizational setting.

In formulating and testing for interaction effects using PLS, one needs to follow a hierarchical process similar to that used in multiple regression in which one compares the results of two models (i.e., one with and one without the interaction construct). Standardized indicators were chosen for this analysis because Likert scales were employed in this study, and the indicators were considered to be theoretically parallel. The standardizations were calculated using SPSS 9.0. For the analysis with the interaction construct (as depicted in Figure 2), it is necessary to include the two main effects constructs (in this study, perceived usefulness and enjoyment) to assess how the moderator construct, enjoyment, influences the impact of perceived usefulness on intention. When using the default standardized output from PLS, the standardized beta estimate of the main construct *X* (perceived usefulness) on dependent construct *Y* (intention) is interpreted as the amount of influence of *X* on *Y* when the moderator construct *Z* (enjoyment) is equal to zero. Likewise, the beta estimate from moderator construct *Z* to *Y* is interpreted as the amount of direct influence of *Z* on *Y* when *X* is equal to zero.

The standardized path estimate from the interaction construct informs us how a change in the level of the moderator construct *Z* (enjoyment) would change the influence of the main construct *X* (perceived usefulness) on dependent construct *Y* (intention). Thus, if *X* (perceived usefulness) has an estimated beta effect of *B* on *Y* (intention), a beta *M* from the interaction construct can be interpreted as a beta change to *B + M* for the estimated path from *X* (perceived usefulness)

to *Y* (intention) when *Z* (enjoyment) increases by one standard deviation from the baseline of zero.

You can also compare the squared multiple correlation (R^2) for this interaction model with the squared multiple correlation for the "main effects" model, which excludes the interaction construct. The difference between the squared multiple correlations is used to assess the overall effect size f^2 for the interaction where 0.02, 0.15, and 0.35 have been suggested as small, moderate, and large effects, respectively (Cohen 1988).³² It is important to understand that a small f^2 does not necessarily imply an unimportant effect. Even a small interaction effects can be meaningful under extreme moderating conditions, if the resulting beta changes are meaningful, then it is important to take these conditions into account.

The results of this study, as shown in Figure 6,³³ give a standardized beta of 0.449 from usefulness to intention, 0.227 from enjoyment to intention, and an interaction effect of -0.209 with a total R^2 of 0.50. Thus, these results imply that one standard deviation increase in enjoyment will not only impact intention directly by 0.227, but it would also decrease the impact of perceived usefulness to intention from 0.449 to 0.240. As expected, the main effects model, shown in Figure 7,³⁴ resulted in a slightly higher standardized beta and a smaller R^2 of 0.465. The interaction construct, therefore, has an effect size f of 0.07,³⁵ which is between a small and medium effect and is larger than found in most past IS studies. Even with a small-to-moderate effect size, these beta estimates help inform us of the conditions under which enjoyment becomes a dominant factor—equaling and potentially overshadowing perceived usefulness. For the group of people who perceive electronic mail to be extremely enjoyable, perceived usefulness will be a

³² $f^2 = [R^2(\text{interaction model}) - R^2(\text{main effects model})] / [1 - R^2(\text{main effects model})]$. Interaction effect sizes are small if 0.02, medium if 0.15, and large if 0.35 (Cohen 1988).

³³ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

³⁴ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

³⁵ $f^2 = (0.50 - 0.465) / (1 - 0.465) = 0.07$, a small to medium effect size.

³¹ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

less important factor on usage intention than is enjoyment. If there is a reasonable likelihood of encountering such a group, being aware of this interaction becomes important.

To assess whether the interaction effect and main effects were significant, a bootstrap resampling procedure (Efron and Tibshirani 1993) was performed. The results of 500 resamples indicate that all paths, weights and loadings, (as shown in Table 11) were significant at the 0.01 level.³⁶

The accuracy of the path estimates to the true effects must be assessed next. As noted earlier, the estimates of the structural paths tend to be more accurate as the reliability score for the estimated construct increases. To assess the reliability of the latent variable estimated by PLS, the composite reliabilities as suggested by Werts et al. (1974) were calculated and are presented in Table 11. Use of this formula, which does not assume equal loadings or error terms among the measures, typically provides more accurate estimates of the composite reliability. Overall, except for enjoyment with a three-indicator composite reliability of 0.85, the composite reliabilities of the other constructs are very high—at or above 0.96. In the case of enjoyment, if we employ the composite reliability as a bias correction factor, the 0.227 path between enjoyment and intention increases slightly to 0.246.

As a contrast, we summed the indicators and performed a moderated regression analysis instead. The result was a smaller interaction path of -0.140 ($p < 0.05$) with a correspondingly smaller R^2 of 0.422. Furthermore, the effect could not be considered significant at the 0.01 level. With the main effects R^2 of 0.404, this yields a substantially lower change in R^2 of 0.018 and an f^2 of 0.03. Contrast these numbers with those obtained using PLS (path of -0.209 with $p < 0.01$, f^2 of 0.07) and we see the theoretically weighted scales of PLS outperformed the summated scales by more than a factor of two. Thus, without the PLS procedure, we would not be able to conclusively

prove that enjoyment, an intrinsic motivation to use a computer, represents a significant and substantive moderating effect in the TAM model.

Discussion and Conclusion

This paper provided a new approach towards the assessment of interaction effects among continuous variables. IS research over the past 15 years has predominantly employed multiple regression- and ANOVA-based analytic techniques to investigate these interaction terms. Less than one-quarter of the interaction terms investigated in these works have been found to be significant, with only a handful of the articles providing an effect size estimate. As suggested, these cumulative results may be due to the analytic technique employed—specifically, multiple regression where this technique was demonstrated through the Monte Carlo simulation to often underestimate the moderator effect size by 16%–37%, depending on the number of indicators used.

As Cronbach (1987, p. 417) has urged, “further investigation of statistical power in studies of interaction and invention of more sensitive research strategies are much to be desired.” Following this sentiment, this study has provided an initial sense of the efficiency and effectiveness of a new PLS product-indicator approach. Through the use of the new approach, it has been shown that sample sizes of approximately 100 with eight indicators per main effect construct, and loadings of 0.70 are needed to detect an interaction effect and to yield reasonably consistent estimates. The combination of 150 sample size, 6 indicators, and 0.70 loadings also produced similar, significant results. In contrast, the average sample size and number of indicators in past IS studies were 81.5 and 3, respectively, much lower than the standards suggested by the Monte Carlo simulation. Increasing the number of indicators was shown to have a larger impact on consistent estimations than did increasing the sample size. Increasing the reliability of indicators will also help, but this typically emerges only after several attempts are made at building better questions and more unidimensional constructs.

It is important to understand that these suggested levels pertain only to detecting interaction effects. By

³⁶ We choose bootstrapping over the use of jackknifing because computational time was not a constraint and jackknifing is considered both less efficient and an approximation to the bootstrap (Chin 1998b, p. 320). Table 11 is available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

virtue of the fact that product indicators are multiplicatively less reliable than their respective indicators, the recommended sample size and indicator levels will always be larger than models with only direct effects. The heuristics for sample size and indicators using PLS typically make assumptions that the loadings of indicators are 0.70 or higher. In the case of modeling interaction effects, we must increase these requirements.

Several additional findings around known PLS biases, influences of different reliabilities, extensions for nonlinear indicators, and use of formative indicators are possible from these results. Appendix D³⁷ provides a more complete technical discussion of these insights.

Finally, it is important to highlight the results of the Monte Carlo simulation with heterogeneous loadings. It is typically the case that measures used in research can vary in the amount of measurement error. PLS is well suited for use in this situation because its primary objective is to differentially weight a set of items to produce the best predictive construct scores. While summated regression would, by default, treat all measures identically, our PLS product-indicator results demonstrate that discounting poor interaction terms can yield better overall estimates. However, the ability to sort through the poor performing items as they are applied within a particular predictive model is equally important. While not presented here because of page limitations, the PLS results via the loading estimates can help a researcher determine which items are of good quality and which need further improvement for future studies.

In summary, the new PLS product-indicator approach seems to yield promising results for researchers interested in assessing interaction effects. The Monte Carlo exercise demonstrated that single-indicator regression was inadequate for assessing interaction terms and that summated regression, while performing better, still underestimated the correct values by substantial margins. In particular, under conditions of heterogeneous loadings where the individual item reliabilities varied, the PLS product-indicator approach

came to the fore in retrieving the “true” population parameter, whereas regression resulted in at least a 16% underestimation. In estimating noninteraction terms that have higher reliabilities, summated regression and PLS performed equally well.

Study 2, which evaluated the PLS approach using an IS empirical data set on enjoyment and use intention again found that the PLS approach retrieved an interaction estimate 33% higher than the summated regression approach. While not as conclusive as the simulation, the effect size was found to be higher than in previous IS studies and twice the size estimated in the summated regression.

In all, it is hoped that the issues raised in this paper—such as appropriate sample size, multiple indicators, reliability, and power—will be part of the mindset and standard information provided in future research papers. Attention to these issues in future research should help the IS field build a cumulative body of knowledge with fewer problems than that found through the literature review. As stated at the start of the paper, it indeed might be the case that theoretical advancement of moderators has been impaired more by analytical techniques than by the lack of conceptualizing contingent factors. We believe that moderators’ roles within emerging theories are poised to have large effects on the field, if only we can improve on the analytical techniques to aid in this discovery.

Acknowledgments

This research was supported in part by a grant from the Social Sciences and Humanities Research Council of Canada whose assistance is gratefully acknowledged. The authors also wish to thank Theodora Lo for her assistance in the literature review and earlier simulation runs. An earlier version of this paper was presented at the 1996 International Conference on Information Systems.

References

- Anderson, J. C., D. W. Gerbing. 1988. Structural equation modeling in practice: A review and recommended two-step approach. *Psych. Bull.* 103 411–423.
- Ang, S., L. L. Cummings, D. W. Straub, E. P. Christopher. 1993. The effects of information technology and the perceived mood of the feedback giver on feedback seeking. *Inform. Systems Res.* 4(3) 240–261.

³⁷ Available at <http://www.informs.org/Pubs/Supplements/ISR/1526-5536-2003-02-SupplA.pdf>.

- Anson, R., R. Bostrom, W. Bayard. 1995. An experiment assessing group support system and facilitator effects on meeting outcomes. *Management Sci.* **41**(2) 189–208.
- Aiken, L. S., S. G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Sage Publications, Newbury Park, CA.
- Aubert, B. A., S. Rivard, M. Patry. 1994. Development of measures to assess dimensions of IS operation transactions. J. I. DeGross, S. L. Huff, M. C. Munro, eds. *Proc. 15th Internat. Conf. Inform. Systems*. Vancouver, British Columbia, 13–26.
- Bell, J. 1984. The effect of presentation form on the use of information in annual reports. *Management Sci.* **30**(2) 169–185.
- Bentler, P. M., C. P. Chou. 1988. Practical issues in structural modeling. J. S. Long, ed. *Common Problems/Proper Solutions, Avoiding Error in Quantitative Research*. Sage Publications, Newbury Park, CA, 161–192.
- Bergeron, F., L. Raymond, S. Rivard. 2001. Fit in strategic information technology management research: An empirical comparison of perspectives. *OMEGA* **29** 125–142.
- Bollen, K. A. 1989. *Structural Equation Models with Latent Variables*. Wiley, New York.
- , P. Paxton. 1998. Interaction of latent variables in structural equation models. *Structural Equation Modeling* **5**(3) 267–293.
- Bostrom, R. P., L. Olfman, M. K. Sein. 1990. The importance of learning style in end-user training. *MIS Quart.* 101–119.
- Brown, C. V., R. P. Bostrom. 1994. Organization designs for the management of end-user computing: Reexamining the contingencies. *J. Management Inform. Systems* **20**(4) 183–211.
- Busemeyer, J. R., L. E. Jones. 1983. Analysis of multiplicative combination rules when the causal variables are measured with error. *Psych. Bull.* **93** 549–552.
- Carmines, E. G., R. A. Zeller. 1979. *Reliability and Validity Assessment*. Sage University Paper Series on Quantitative Applications in the Social Science, Series no. 07-017. Sage Publications, Beverly Hills, CA.
- Cavaye, A. L. M. 1995. User participation in system development revisited. *Inform. Management* **29** 311–323.
- Chan, H. C., B. C. Y. Tan, K. W. Wei. 1994. The query cube: A framework for assessing user productivity with database information retrieval. *Proc. 15th Internat. Conf. Inform. Systems*. Vancouver, British Columbia, Canada.
- Chan, Y. E., S. L. Huff, D. W. Barclay, D. G. Copeland. 1997. Business strategic orientation, information systems strategic orientation, and strategic alignment. *Inform. Systems Res.* **8**(2) 125–149.
- Chidambaram, L., B. Jones. 1993. Impact of communication medium and computer support on group perceptions and performance: A comparison of face-to-face and dispersed meetings. *MIS Quart.* **17**(4) 465–491.
- Chin, W. Y. W. 1995. PLS is to LISREL as principal components analysis is to common factor analysis. *Tech. Stud.* **2** 315–319.
- . 1998a. Issues and opinion on structural equation modeling. *MIS Quart.* **22**(1) vii–xvi.
- . 1998b. The partial least squares approach for structural equation modeling. George A. Marcoulides, ed. *Modern Methods for Business Research*. Lawrence Erlbaum Associates, Mahwah, NJ, 295–336.
- . 2001. PLS-Graph Manual, Version 3.0, Unpublished.
- , A. Gopal. 1995. Adoption intention in GSS: Importance of beliefs. *Data Base Adv.* **26** 42–64.
- , B. L. Marcolin. 1995. The holistic approach to construct validation in IS research: Examples of the interplay between theory and measurement. *Admin. Sci. Assoc. Canada—23rd Conf., IS Proc.*, Vol. 16. Windsor, Ontario, Canada.
- , P. R. Newsted. 1999. Structural equation modeling analysis with small samples using partial least squares. Rick Hoyle, ed. *Statistical Strategies for Small Sample Research*. Sage Publications, Thousand Oaks, CA, 307–341.
- Chou, C., P. M. Bentler, A. Satorra. 1991. Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British J. Math. Statist. Psych.* **44** 347–357.
- Chung, W. Y., S. Iacono. 1994. Use of advanced features and perceptions of software quality: An experimental study. *Twenty-Seventh Hawaii Internat. Conf. System Sci. Hawaii*.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum, Hillsdale, NJ.
- , P. Cohen. 1983. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Compeau, D. R., C. A. Higgins. 1995. Application of social cognitive theory to training for computer skills. *Inform. Systems Res.* **6** 118–143.
- Cook, G. J. 1993. An empirical investigation of information search strategies with implications for decision support system design. *Decision Sci.* **24**(3) 683–697.
- Cronbach, L. J. 1987. Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psych. Bull.* **102** 414–417.
- Crossland, M. D., B. E. Wynne. 1994. Measuring and testing the effectiveness of a spatial decision support system. *Twenty-Seventh Hawaii Internat. Conf. System Sci.* Hawaii, IEEE Computer Society Press, Los Alamitos, CA.
- , J. N. Scudder, R. T. Herschel, B. E. Wynne. 1993. Measuring the relationships of task and cognitive style factors and their effects on individual decision-making effectiveness using a geographic information system. *Twenty-Sixth Hawaii Internat. Conf. System Sci.* IEEE Computer Society Press, Los Alamitos, CA.
- Culnan, M. J. 1987. Mapping the intellectual structure of MIS, 1980–1985: A co-citation analysis. *MIS Quart.* **11**(3) 340–353.
- Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quart.* **13** 319–340.
- , R. P. Bagozzi, P. R. Warshaw. 1992. Extrinsic and intrinsic motivation to use computers in the workplace. *J. Appl. Soc. Psych.* **22** 1111–1132.
- Davis, S. A., R. P. Bostrom. 1993. Training end users: An experimental investigation of the roles of the computer interface and training methods. *MIS Quart.* **17**(1) 61–85.
- DeSanctis, G., M. D’Onofrio, V. Sambamurthy, M. S. Poole. 1989. Comprehensiveness and restrictiveness in group decision

- heuristics: Effects of computer support on consensus decision making. *The Proc. Internat. Conf. Inform. System.* Boston, MA, 131–140.
- Dickson, G. W., G. DeSanctis, D. J. McBride. 1986. Understanding the effectiveness of computer graphics for decision support: A cumulative experimental report. *Comm. ACM* **29**(1) 40–47.
- , J.-E. L. Partridge, L. H. Robinson. 1993. Exploring modes of facilitative support for GDSS technology. *MIS Quart.* **17**(2) 173–194.
- Doll, W. J., G. Torkzadeh. 1989. A discrepancy model of end-user computing involvement. *Management Sci.* **35**(10) 1151–1171.
- , ———. 1991. The measurement of end-user computing satisfaction: Theoretical and methodological issues. *MIS Quart.* (March) 5–10.
- Duxbury, L. E., C. A. Higgins, S. Mills. 1992. After-hours telecommuting and work-family conflict: A comparative analysis. *Inform. Sys. Res.* **3**(2) 173–191.
- Eckel, N. 1987. The interaction between the relative accuracy of probabilistic vs. deterministic predictions and the level of prediction-task difficulty. *Decision Sci.* **18**(2) 206–217.
- Efron, B., R. J. Tibshirani. 1993. *An Introduction to the Bootstrap* (Monographs on Statistics and Applied Probability, #57). Chapman & Hall, New York.
- Fedorowicz, J., E. Oz, P. D. Berger. 1992. A learning curve analysis of expert system use. *Decision Sci.* **23**(4) 797–818.
- Fornell, C., F. L. Bookstein. 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *J. Marketing Res.* **19** 440–452.
- , Y. Yi. 1992. Assumption of the two-step approach to latent variable modeling. *Sociological Methods Res.* **20** 291–320.
- , P. Lorange, J. Roos. 1990. The cooperative venture formation process: A latent variable structural modeling approach. *Management Sci.* **36** 1246–1255.
- Franz, C. R. 1979. Contingency factors affecting the user involvement role in the design of successful information systems. Unpublished Ph.D. dissertation, University of Nebraska, Lincoln, NE.
- Gelderman, M. 1998. The relation between user satisfaction, usage of information systems and performance. *Inform. Management* **34** 11–18.
- George, J. F., G. K. Eason, J. F. Nunamaker, Jr., G. B. Northcraft. 1990. A study of collaborative group work with and without computer-based support. *Inform. Systems Res.* **1**(4) 394–415.
- Ginzberg, M. J. 1979. A study of the implementation process. *TIMS Stud. Management Sci.* **13** 85–102.
- Goodhue, D. 1986. IS attitudes: Toward theoretical and definition clarity. *Proc. Internat. Conf. Inform. Systems*, December 15–17. San Diego, CA, 181–194.
- , R. L. Thompson. 1995. Task-technology fit and individual performance. *MIS Quart.* **19** 213–236.
- Goslar, M. D., G. I. Green, T. H. Hughes. 1986. Decision support systems: An empirical assessment for decision making. *Decision Sci.* **17** 79–91.
- Greenwald, A. G., A. R. Pratkanis, M. R. Leippe, M. H. Baumgardner. 1986. Under what conditions does theory obstruct research progress? *Psych. Rev.* **93**(2) 216–229.
- Hale, D. P., G. M. Kasper. 1989. The effect of human-computer interchange protocol on decision performance. *J. Management Inform. Systems* **6**(1) 5–20.
- Hartwick, J., H. Barki. 1994. Explaining the role of user participation in information systems use. *Management Sci.* **40** 440–465.
- Hiltz, S. R., K. Johnson, M. Turoff. 1991. Group decision support: The effects of designated human leaders and statistical feedback in computerized conferences. *J. Management Inform. Systems* **8**(2) 81–108.
- Hughes, C. T., M. L. Gibson. 1991. Students as surrogates for managers in a decision-making environment: An experimental study. *J. Management Inform. Systems* **8**(2) 153–166.
- Igbaria, M., J. J. Baroudi. 1993. A short-form measure of career orientations: A psychometric evaluation. *J. Management Inform. Systems* **10**(2) 131–154.
- , ———. 1995. The impact of job performance evaluations on career advancement prospects: An examination of gender differences in the IS workplace. *MIS Quart.* **19**(1) 107–123.
- Ives, B., M. H. Olson. 1984. User involvement and MIS success: A review of research. *Management Sci.* **30**(5) 589–603.
- Jaccard, J., R. Turrisi, C. K. Wan. 1990. *Interaction Effects In Multiple Regression*. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series #07-072. Sage Publications, Newbury Park, CA.
- Jacobs, S. M., R. T. Keim. 1988. An experimental study in overcoming hypothesis-confirming search strategies in computerized information retrieval systems. *Proc. Internat. Conf. Inform. Systems*. Minneapolis, MN, 81–89.
- Jarvenpaa, S. L., J. J. Machesky. 1986. End user learning behavior in data analysis and data modeling tools. *Internat. Conf. Inform. Systems*. San Diego, CA.
- , V. S. Rao, G. P. Huber. 1988. Computer support for meetings of groups working on unstructured problems: A field experiment. *MIS Quart.* **12** 645–666.
- Jones, W. J., R. J. McLeod. 1986. The structure of executive information systems: An exploratory analysis. *Decision Sci.* **17** 220–249.
- Jöreskog, K. G., D. Sorbom. 1993. *PRELIS 2 User's Reference Guide*. Scientific Software International, Lincolnwood, IL.
- , H. Wold. 1982. *Systems Under Indirect Observation—Causality Structure Prediction*. North-Holland Publishing Company, Amsterdam, The Netherlands.
- , F. Yang. 1996. Nonlinear structural equation models: The Kenny-Judd model with interaction effect. G. A. Marcoulides, R. E. Schumacker, eds. *Advanced Structural Equation Modeling, Issues and Techniques*. Lawrence Erlbaum Associates, Mahway, NJ, 57–88.
- Kasper, G. M., R. P. Cerveny. 1985. A laboratory study of user characteristics and decision-making performance in end-user computing. *Inform. Management* **9** 87–96.
- Keil, M., R. Mixon, T. Saarinen, V. Tuunainen. 1994–95. Understanding runaway information technology projects: Results from

- an international research program based on escalation theory. *J. Management Inform. Systems* **11**(3) 65–85.
- Kenny, D. A., C. M. Judd. 1984. Estimating the nonlinear and interactive effects of latent variables. *Psych. Bull.* **96** 201–210.
- Kim, K. K., N. S. Umanath. 1992–1993. Structure and perceived effectiveness of software development subunits: A task contingency analysis. *J. Management Inform. Systems* **9**(3) 157–181.
- Kwon, T. H. 1990. A diffusion of innovation approach to MIS infusion: Conceptualization, methodology, and management strategies. *Proc. Internat. Conf. Inform. Systems*. Copenhagen, Denmark, 139–147.
- Li, F., P. Harmer, T. E. Duncan, S. C. Duncan, A. Acock, S. Boles. 1998. Approaches to testing interaction effects using structural equation modeling methodology. *Multivariate Behavioral Res.* **33**(1) 1–39.
- Lohmöller, J.-B. 1984. *LVPLS Program Manual: Latent Variables Path Analysis with Partial Least-Squares Estimation*. Zentralarchiv für empirische Sozialforschung, Köln, Germany.
- . 1989. *Latent Variable Path Modeling with Partial Least Squares*. Physica-Verlag, Heidelberg, Germany.
- Lohse, G. L., E. J. Johnson. 1996. A comparison of two process tracing methods for choice tasks. *Twenty-Ninth Hawaii Internat. Conf. System Sci.* IEEE, Hawaii.
- Mackay, J. M., S. H. Barr, M. G. Kletke. 1992. An empirical investigation of the effects of decision aids on problem-solving processes. *Decision Sci.* **23**(1) 648–672.
- McClelland, G. H., C. M. Judd. 1993. Statistical difficulties of detecting interactions and moderator effects. *Psych. Bull.* **114** 376–389.
- McKeen, J. D., T. Guimaraes, J. C. Wetherbe. 1994. The relationship between user participation and user satisfaction: An investigation of four contingency factors. *MIS Quart.* **18** 427–451.
- Mookerjee, V. S., B. L. Dos Santos. 1993. Inductive expert system design: Maximizing system value. *Inform. Systems Res.* **4**(2) 111–140.
- , M. V. Mannino, R. Gilson. 1995. Improving the performance stability of inductive expert systems under input noise. *Inform. Systems Res.* **6**(4) 329–356.
- Motiwall, J., F. Y. K. Pheng. 1982. Decision effectiveness and information use: Effects of cognitive style, complexity, and stress. *Proc. Internat. Conf. Inform. Systems*, December 13–15. Ann Arbor, MI, 137–151.
- Noreen, E. W. 1989. *Computer Intensive Methods for Testing Hypotheses, An Introduction*. John Wiley & Sons, New York.
- Nunamaker, J. F., Jr., A. R. Dennis, J. S. Valacich, D. R. Vogel. 1991. Information technology for negotiation groups: Generating options for mutual gain. *Management Sci.* **37**(10) 1325–1347.
- Nutt, P. C. 1986. Evaluating MIS design principles. *MIS Quart.* 139–155.
- Olfman, L., M. K. Sein, R. P. Bostrom. 1989. EUC training: Comparison of methods and the role of individual differences. *Proc. 22nd Annual Hawaii Internat. Conf. System Sci.* IEEE, Hawaii.
- Ping, J. R. A. 1995. A parsimonious estimating technique for interaction and quadratic latent variables. *J. Marketing Res.* **32** 336–347.
- . 1996. Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psych. Bull.* **119** 166–175.
- Powers, R. F., G. W. Dickson. 1973. MIS project management: Myths, opinions, and reality. *California Management Rev.* **15** 147–156.
- Premkumar, G., W. R. King. 1994. The evaluation of strategic information system planning. *Inform. Management* **26** 327–340.
- Reinartz, W. J., R. Echambadi, W. W. Chin. 2002. Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Res.* **37**(2) 227–244.
- Sambamurthy, V., G. DeSanctis. 1990. An experimental evaluation of GDSS effects on group performance during stakeholder analysis. *Proc. 23rd Annual Hawaii Internat. Conf. System Sci.* IEEE Computer Society Press, Hawaii.
- , R. Zmud. 1999. Arrangements for information technology governance: A theory of multiple contingencies. *MIS Quart.* **23**(2) 261–290.
- Santhanam, R., M. K. Sein. 1994. Improving end-user proficiency: Effects of conceptual training and nature of interaction. *Inform. Systems Res.* **5**(4) 378–399.
- Schonberger, R. J. 1980. MIS design: A contingency approach. *MIS Quart.* **4** 13–20.
- Sengupta, K., D. Te'eni. 1993. Cognitive feedback in GDSS: Improving control and convergence. *MIS Quart.* **17**(1) 87–113.
- Sharma, S., S. Durvasula, W. R. Dillon. 1989. Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *J. Marketing Res.* **XXVI** 214–221.
- Smith, K. W., M. S. Sasaki. 1979. Decreasing multicollinearity: A method for models with multiplicative functions. *Sociological Methods Res.* **8** 35–56.
- Stone, E. F. 1988. Moderator variables in research: A review and analysis of conceptual and methodological issues. *Res. Personal and Human Resources Management* **6** 191–229.
- Suh, K. S., A. M. Jenkins. 1992. A comparison of linear keyword and restricted natural language data base interfaces for novice users. *Inform. Systems Res.* **3**(3) 252–272.
- Tabachnick, B. G., L. S. Fidell. 1989. *Using Multivariate Statistics*, 2nd ed. Harper and Row, New York.
- Tait, P., I. Vessey. 1988. The effect of user involvement on system success: A contingency approach. *MIS Quart.* (March) 90–107.
- Tan, B. C. Y., K. K. Wei, R. T. Watson. 1993. Dampening status influence using a group support system: An empirical study. *Proc. Internat. Conf. Inform. Systems*. Orlando, FL, 77–89.
- Taylor, S., P. A. Todd. 1995. Understanding information technology usage: A test of competing models. *Inform. Systems Res.* **6**(2) 144–176.
- Todd, P., I. Benbasat. 1991. An experimental investigation of the impact of computer based decision aids on decision making strategies. *Inform. Systems Res.* **2**(2) 87–115.
- , ———. 1992. The use of information in decision making: An experimental investigation of the impact of computer-based decision aids. *MIS Quart.* **16** 373–393.

- Trice, A. W., M. E. Treacy. 1986. Utilization as a dependent variable in MIS research. *Proc. Internat. Conf. Inform. Systems*, December 15–17. San Diego, CA, 227–239.
- Umanath, N. S., R. W. Scamell, R. D. Sidhartha. 1990. An examination of two screen/report design variables in an information recall context. *J. Management Inform. Systems* 21(1) 216–241.
- Vogel, D., J. Lehman, G. Dickson. 1986. The impact of graphical displays on persuasion: An empirical study. *Internat. Conf. Inform. Systems*. San Diego, CA.
- Watson, R. T., T. H. Ho, K. S. Raman. 1994. A fourth dimension of group support systems. *Comm. ACM* 37(10) 45–55.
- Weill, P., M. H. Olson. 1989. An assessment of the contingency theory of management information systems. *J. Management Inform. Systems* 6 59–85.
- Werts, C. E., R. L. Linn, K. G. Jöreskog. 1974. Intraclass reliability estimates: Testing structural assumptions. *Educational Psych. Measurement* 34 25–33.
- Wilson, E. V., T. B. A. Addo. 1994. The joint effects of interactions between data display and task variables on task performance. *Internat. Conf. Inform. Systems*. CIS, Vancouver, British Columbia, Canada.
- Wold, H. 1982. Soft modeling—The basic design and some extensions. K. Jöreskog, H. Wold, eds. *Systems Under Indirect Observation II*. North-Holland Press, Amsterdam, The Netherlands, 1–53.
- . 1985. Partial least squares. S. Kotz, N. L. Johnson, eds. *Encyclopedia of Statistical Sciences*, Vol. 6. Wiley, New York, 581–591.
- . 1989. Introduction to the second generation of multivariate analysis. H. Wold, ed. *Theoretical Empiricism*. Paragon House, New York, vii–xi.
- Wood, J. G., J. T. Nosek. 1994. Discrimination of structure and technology in a group support system: The role of process complexity. *Internat. Conf. Inform. Systems*. Vancouver, British Columbia, Canada.

Izak Benbasat, Senior Editor. This paper was received on May 12, 1997, and was with the authors 29 months for 3 revisions.

Copyright 2003, by INFORMS, all rights reserved. Copyright of Information Systems Research is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.