Tutorial

Partial Least Squares (PLS) Structural Equation Modeling (SEM) for Building and Testing Behavioral Causal Theory: When to Choose It and How to Use It

—Feature by
PAUL BENJAMIN LOWRY AND JAMES GASKIN

***Abstract*—Problem:** *Partial least squares (PLS), a form of structural equation modeling (SEM), can provide much value for causal inquiry in communication-related and behavioral research fields. Despite the wide availability of technical information on PLS, many behavioral and communication researchers often do not use PLS in situations in which it could provide unique theoretical insights. Moreover, complex models comprising formative (causal) and reflective (consequent) constructs are now common in behavioral research, but they are often misspecified in statistical models, resulting in erroneous tests.* **Key concepts:** *First-generation (1G) techniques, such as correlations, regressions, or difference of means tests (such as ANOVA or t-tests), offer limited modeling capabilities, particularly in terms of causal modeling. In contrast, second-generation techniques (such as covariance-based SEM or PLS) offer extensive, scalable, and flexible causal-modeling capabilities. Second-generation (2G) techniques do not invalidate the need for 1G techniques however. The key point of 2G techniques is that they are superior for the complex causal modeling that dominates recent communication and behavioral research.* **Key lessons:** *For exploratory work, or for studies that include formative constructs, PLS should be selected. For confirmatory work, either covariance-based SEM or PLS may be used. Despite claims that lower sampling requirements exist for PLS, inadequate sample sizes result in the same problems for either technique.* **Implications:** *SEM's strength is in modeling. In particular, SEM allows for complex models that include latent (unobserved) variables, formative variables, chains of effects (mediation), and multiple group comparisons of these more complex relationships.*

Index Terms*—Causal inquiry, partial least squares (PLS), structural equation modeling (SEM), theory building, 1G statistical techniques, 2G statistical techniques.*

INTRODUCTION

The primary purpose of statistical techniques is to estimate the probability that the pattern of data collected could have occurred by chance rather than by the causes proposed by the theory being tested. These techniques should be carefully selected based on the type of data collected and should be carried out in the context of theory using measures derived from a theory. Not all portions of a theory are easily tested. There is much about a theory that a researcher must understand before employing statistical tests—for example, its axiomatic foundations and the internal consistency of its logic. Statistics have no value when testing a shoddily constructed theory that could be readily broken by simple logic or counterexamples.

Regrettably, sometimes the application of statistical analyses is not aligned with the theory being tested. When this occurs, it holds back the progress of scientific research and diminishes the relevance of our work. Although such activities may have the form of research, they lack its substance. The implications of this difference are aptly described in the following anecdote:

> In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he's the controller—and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're

Manuscript received August 22, 2012; revised August 12, 2013 and March 01, 2014; accepted March 03, 2014. Date of publication April 22, 2014; date of current version May 20, 2014.
P. B. Lowry is with the Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong P7912, China (email: Paul.Lowry.PhD@gmail.com).
J. Gaskin is with the Department of Information Systems, Marriott School of Management, Brigham Young University, Provo, UT 84602 USA (email: james.eric.gaskin@gmail.com).
This paper has downloadable supplementary material at http://ieeexplore.ieee.org. The file contains Appendices 1 to 3. The material is 83 kB in size.

IEEE 10.1109/TPC.2014.2312452

missing something essential, because the planes don't land. Adapted from a 1974 Caltech commencement address by Richard Feynman [1, p. 155].

We, as researchers, exhibit "cargo cult" behavior when we run statistical tests that we have seen others use in the hopes of obtaining results similar to those that others have obtained without fully understanding if the tests we are applying actually match the particular needs of our current study. Having the form and appearance of research through familiar statistical tests and reporting without an understanding of the appropriate application of these tests cannot produce results upon which we can place any confidence.

We hope to steer researchers away from the cargo cult of behavioral research by providing a nontechnical, approachable explanation of the tradeoffs between partial least squares (PLS), covariance-based-structural equation modeling (CB-SEM), and first-generation (1G) statistical techniques in analyzing complex causal models. Some papers already make useful strides toward providing simple "beginner's guides" to PLS [2] and CB-SEM [3]. The current tutorial extends these papers by offering additional analytical procedures not offered in those beginner's guides and drawing upon more recent debates in the literature regarding the relative merits and limitations of each analytic approach. We also attempt to write this tutorial in nonmathematical lay terms to make it more approachable. It behooves all researchers—whether they work under the disciplines of causal epistemology or whether they only review papers produced under those disciplines—to gain fundamental knowledge of second-generation (2G) statistical techniques so that they know which technique should be applied in which circumstance and understand the techniques' empirical and theoretical implications.

Statistical methods are sine qua non in testing the utility of causal theories, but they can also aid theory development. Partial least squares (PLS), a form of SEM, has much to offer communication and behavioral researchers in this regard. A number of articles provide technical details about SEM, comparing component-based PLS SEM to CB-SEM (e.g., [4], [5]–[7]). The technical nature of this literature, however, may distract behavioral and communication researchers from the relatively simple logic of PLS, dissuade them from selecting the technique, or worse, cause them to misappropriate the technique.

Importantly, we find little evidence that PLS is being used to its potential in the broad field of technical and management communication research. To assess the use of PLS, we examined the top 10 technical and management communication journals [8]. In this field, PLS was first used in 1998 in the *Journal of Computer-Mediated Communication* [9], but it was then virtually ignored for a decade. Later, three JCMC papers used PLS [10]–[12]. In IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION (TPC), PLS was first used in three papers by Kock et al. [13]–[15] A total of seven other TPC papers subsequently have used PLS [16]–[22]. More recently, PLS was used once in the *Journal of Business Communication* [23]. We found no use of PLS in *Business Communication Quarterly, Journal of Business and Technical Communication*, *Journal of Technical Writing and Communication*, *Management Communication Quarterly*, *Technical Communication, Technical Communication Quarterly*, or in *Written Communication*. In contrast, PLS has been used thousands of times in other behavioral research fields' articles, including information systems, management, and marketing. Clearly, the use of PLS is lacking in our field, and this could likely be because of the challenging empirical nature of PLS and similar SEM techniques. The key limitation of not using SEM is that it holds back communication researchers from a more complete understanding and testing of whole theoretical models that drive communication phenomenon—from communication quality, teamwork, virtual teams, coordination, interactivity, and engagement to deception. This is a particular pressing opportunity in communication research, since most of the established theoretical models—such as Social Cognitive Theory, Theory of Planned Behavior, technology acceptance model, Protection Motivation theory, and General Deterrence Theory—are too complex for full testing with traditional statistical techniques. As a result, most communication research studies test "parts" of theoretical models without testing the "big picture" of the underlying theory.

Given the paltry use of PLS in the broader field of technical and management communication, we seek to make PLS approachable to communication researchers who are not necessarily experts in statistics and SEM, but who are familiar with the concepts and language commonly used in 1G statistical techniques, such as linear regression and ANOVA. By approaching PLS from a nontechnical, measurement-theoretical perspective, we hope to clarify when it may or may not be appropriate to adopt PLS, how it may be conducted, how its findings can be interpreted, and how they can inform theory development and theory testing. To begin, we explain the advantages of 2G statistical techniques (such as SEM), define PLS as a specialized form of SEM, and contrast the relative merits of PLS with the merits of 1G techniques and CB-SEM. Then, we demonstrate PLS using a large dataset and a complex statistical model with formative and reflective indicators in both moderating and mediating relationships. We conclude with recommendations on the use of 1G and 2G statistical techniques.

KEY CONCEPTS

In this section, we provide the foundation for our tutorial by first explaining the key differences between 1G and 2G statistical techniques. We then discuss the nature and basics of causal modeling. This section concludes with a discussion of the advantages of SEM for testing causal models. These particular concepts were selected to provide the reader with sufficient background to understand the differences between 1G and 2G techniques, to better understand causal modeling, and to explain why SEM holds specific advantages for causal modeling.

Key Concept 1: 1G Statistical Analysis Techniques 1G techniques are statistical methods, such as correlations, regressions, or difference of means tests (e.g., ANOVA or t-tests), that are well suited to simple modeling scenarios.

Correlations are useful for exploratory research. That is, they can be used for noncausal exploration of how constructs may be related, on which future path modeling or more causal research can be based (e.g., [24]). Correlations are also fundamental to more complex projects in helping to provide measurement model statistics for regression or SEM (e.g., [25]) and in helping to establish that constructs in a model do not suffer from common methods bias.

Means tests are particularly useful for experimentation, where the focus of a study is to demonstrate causality by controlling for time, treatment, and control conditions, among other considerations. In experimentation, one typically needs to use means testing to demonstrate that the treatment condition indeed behaved differently than the control condition—a test often known as a *manipulation check*. Such scenarios often occur when communication researchers are concerned with treatment differences in communication-related constructs, such as group size, information communication technology (ICT) use, task structure, deception, coordination, and communication media (e.g., face-to-face versus online), as seen in [26]–[31]. Or, for example, even if one is using regression or SEM for overall path modeling, means tests may be used to establish expected individual- or national-level cultural differences between samples, as seen in [13] and [32]–[36].

Regression analysis is particularly well suited to simple models in which few IVs and DVs are involved and the data are highly normalized (e.g., [27]). Regression can also be used to test highly simple models for the existence of moderation (such as interaction effects) and mediation. Regression is also ideal for repeated measures (e.g., [32]). It is also ideal for communication research involving social network analysis that relies on logistic regressions (e.g., [37]).

However, 1G techniques offer limited modeling capabilities, particularly in terms of *causal* or complex modeling. Specifically, 1G techniques either cannot, or are ill suited to modeling latent variables, indirect effects (mediation), multiple group moderation of multiple effects, and assessing the “goodness” of the proposed (tested) model in comparison with the observed relationships contained in the data.

Key Concept 2: 2G Statistical Analysis Techniques 2G techniques (such as SEM) are statistical methods for modeling causal networks of effects simultaneously—rather than in a piecemeal manner. SEM offers extensive, scalable, and flexible causal-modeling capabilities beyond those offered by 1G techniques. 2G techniques do not invalidate the need for 1G techniques however.

One of the prime advantages of SEM is the ability to include latent (unobserved) variables in causal models. Thus, the researcher may model abstract constructs comprised of many indicators (observed variables), each of which is a reflection or a dimension of the latent construct. Another key advantage of SEM is that it enables the researcher to estimate complete causal networks simultaneously. For example, the effect of $\mathrm{A} \rightarrow \mathrm{B}$ can be estimated while also estimating the effects of $\mathrm{A} \rightarrow \mathrm{C}$ and $\mathrm{B} \rightarrow \mathrm{C}$, as well as the indirect effect of A on C through B. In addition, these effects can all be estimated across multiple groups (e.g., management versus nonmanagement employee), while controlling for potential confounds (e.g., firm size and firm performance). Although SEM has rarely been used in communication-related research, it has many potential applications, and it has been used before (e.g., [10], [11]–[23]).

The key point of 2G techniques is that they are superior for the complex causal modeling that dominates recent communication and behavioral research.

> [SEM is] viewed as a coupling of two traditions—an econometric perspective focusing on prediction and a psychometric emphasis that models concepts as latent (unobserved) variables that are indirectly inferred by multiple observed measures [which are also called indicators or manifest variables. [38, p. vii]

Key Concept 3: Statistical Analysis Techniques for Causal Modeling Causal modeling defines variables and estimates the relationships among them to explain how changes in one or more variables result in changes to one or more other variables within a given context. Professional communication researchers might use causal modeling to explain and predict the key factors that increase communication quality, to examine the key predictor of deception communication, to examine the key customer and website communication elements that predict customer loyalty, to examine key factors predicting information overload, to examine the factors that predict susceptibility to spear phishing, and so on (e.g., [10], [13], [16], [23]). Causal inference makes three primary assumptions [39]: (1) covariation, (2) absence of spurious relationships, and (3) temporal precedence.

Covariation means that the predictor and the predicted variable vary together—that is, a change in the predictor leads to an estimable and systematic change in the predicted variable. Such covariation is seen in all of the previous citations of SEM use in communication fields.

Absence of spurious relationships means that potential confounds are accounted for in a model. In a communications context, this may refer to including critical factors, such as trust, communication history, and common ground, in a model of communication quality.

Temporal precedence means that the predictor occurs prior to the predicted variable such that the relationship can truly be causal. For example, does quality communication result from frequency of interaction? Or, is communication quality a cause of frequent interaction? One must logically demonstrate through argument (or empirically) if it is the chicken or the egg. 1G and 2G techniques allow for discovering the covariation of variables and the absence of spurious relationships, although 2G techniques allow for more sophisticated assessment of potential confounds. Temporal precedence cannot be established through either statistical technique, but must be worked out logically or through empirical evidence.

At its core, basic variable modeling can be represented as the relationship between two variables, as in the common linear equation: $y = mx + b$, where y is the dependent (predicted or outcome) variable, x is the independent (predictor or indicator) variable, m represents the relationship (slope) between x and y, and b is the intercept along a 2D axis. 1G techniques, such as simple linear regression, are extensions of this basic equation. As such, these techniques suffer from three main limitations in modeling [2]: (1) the tested model structure must be simple,[1] (2) all variables must be observable[2] (such as not latent), and (3) estimation of error is neglected. Hence, while multiple x variables can be included in the linear equations, multiple y variables and additional variables (a, b, c, etc.) cannot be accounted for within the equation. Thus, such multiple equations must be run separately in order to assess more complex models. 2G modeling techniques offer a robust solution by running these equations simultaneously and interdependently such that the effects of all variables are estimated codependently and simultaneously rather than separately. Accordingly, 2G techniques are able to offer a "truer" picture of the interdependent relationships in a complex theoretical model.

Key Concept 4: Advantages of Structural Equation Modeling Over 1G Techniques Given this background, 2G statistical techniques provide many additional features unavailable in 1G techniques. 1G statistics test the plausibility of *a single theoretical proposition*. A *theoretical proposition* is a functional statement of cause and

[1]This limitation can also be overcome if one utilizes complex plugging that leverages bootstrapping [40].

[2]This limitation can be overcome to some extent by extracting latent variable scores (LVS) during a factor analysis. These LVS can then be used as "observed" or "composite" proxy variables (instead of the full factor) during subsequent tests of causal relationships (regressions). This approach has its limitations, however, since LVS does not account for error as well as a fully latent factor.

effect (e.g., changes in X cause changes in Y; Y is a function of X). Most theories, however, require more than a single proposition to explain observed variations in the phenomenon of fully interested. 1G techniques, therefore, can only test a complex theoretical model (the empirical manifestation of a testable theory) in fragments. 2G statistical techniques, such as SEM, can test the plausibility of an *entire collection of propositions comprising a causal theory* simultaneously [41]. SEM can model multiple independent variables (IV) and multiple dependent variables (DV), chains of causal effects and indirect effects, and the latent constructs that variables are meant to measure. *Latent constructs* are constructs that cannot be measured directly, but that can be estimated through proxies. System robustness, for example, would be considered a latent construct because it cannot be measured directly, but could be estimated with proxy variables, such as average time between failures, downtime, and data losses. In SEM, an observed variable (sometimes called a *manifest* variable) that is offered as one of a handful of proxies for an unobservable latent construct is called an *indicator*.

1G statistical techniques must test the convergent[3] and discriminant[4] validity of latent variables using separate statistical procedures than those used to test causal relationships between those variables. SEM simultaneously tests the validity of measures and the plausibility of a theory [42]. 1G statistical techniques cannot directly test mediated relationships—chains of a relationship among three or more constructs (e.g., changes in W cause changes in X, which cause changes in Y, which cause changes in Z). Instead, they must decompose them into tests of relationships among pairs of constructs. SEM, in contrast, can directly test theoretically proposed chains of cause and effect, which are called *complex multistaged models*.

1G statistical techniques cannot provide estimates of measurement error when theoretical models require that several measurement items be multiplied together to compute the value of a variable. Multiplying measurements causes compounding of measurement errors. This circumstance is called a *fixed-scale construction* problem. SEM can provide estimates of measurement error for each item in a multi-item scale and for the scale as a whole, even when items are multiplied together. Next, we explain each of these feature advantages in more detail.

(1) SEM Jointly Assesses Measurement and Theory: Causal research has a long tradition of measuring constructs indirectly through multiple measurement items [43], such as using a system satisfaction questionnaire because actual satisfaction cannot be directly measured or using a technology-acceptance questionnaire because actual attitudes cannot be directly measured. When indirect measures (variables offered as surrogates for a construct) are used to gather data, measurement error is virtually guaranteed. Therefore, it becomes important to establish discriminant and convergent validity of one's measurement instruments before testing the theoretical model [43]. 1G techniques cannot test instrument validity and nomology (the theoretical relationships tested by the theoretical model) simultaneously; instead, researchers must conduct separate assessments for each in a process known as the "two-step approach" [41].

Separation of the theory and measure can cause incorrect measurements, incorrect explanations, and incorrect predictions. Theory and measurement are conceptually integrated and are best dealt with together. Theory tells us what to measure, and the results can be accurately interpreted only in the context of the theory from which they have sprung [44].[5] Chin [44] demonstrated empirically that measures for one construct that were validated separately from a theory in a two-step approach did not necessarily remain valid when they were combined with measures of other constructs from the same theory. The validity of a simple four-item scale changed when it was combined in three different ways with other measures of other constructs.[6] 1G techniques for reliability analysis

[3]Convergent validity is the degree to which operationalization converges on other operationalizations with which it should be theoretically similar.

[4]Discriminant validity is the degree to which an operationalization diverges from other operationalizations from which it should be theoretically dissimilar.

[5]However, there are cases where the two-step approach is appropriate for empirical theory building. These cases include situations when substantive apriori theory does not exist, such as in a theoretical scale development, preliminary studies, and early confirmatory studies [7]. In such scenarios, the results can be considered only exploratory and should not be treated as generalizable [7]. However, as shown in the next section, the SEM technique of PLS can be superior for many of these scenarios because it allows for exploratory theoretical development with simultaneous analysis of measures.

[6]Chin believes one reason for this outcome is the: "multidimensional nature of the measures. In a separate components analysis, only the epistemic relationship between the indicators and construct are examined. But when a causal connection is made between constructs, the appropriateness of a set of measures relates not only to how well they tap into a construct, but also how well they predict (or are predicted by) another construct" [44, pp. 39-40].

force the researcher to either ignore or constrain measurement error in subsequent analysis of the theoretical model—causing the problem of fixed-scale construction, among other problems [45]. Fixed-scale construction is problematic because it removes information from the theoretical model. Having tested the interitem reliability of individual items, using, for example, a Cronbach's alpha test, 1G statistical approaches typically average or sum the items into a scale with a single score, making further analysis of measurement error impossible [45].

The more complex a theory becomes, the more serious the problems with a two-step approach become. SEM offers a solution to the two-step approach by simultaneously testing the convergent validity and discriminant validity of the scales used to measure theoretical constructs (called "the measurement model") and the proposed nomological links among theoretical constructs (called "the structural model") [43].[7]

(2) SEM Allows for Holistic Testing of Multistaged Models: Since 1G techniques can test only one theoretical proposition at a time, each theoretical proposition must be tested separately from the other propositions. Therefore, it is impossible to test mediated relationships (chains of causation) among constructs directly with 1G techniques. The piecemeal testing of these relationships can lead to inflated t-statistics (indication of significance), which increases the likelihood of Type I error (false positives). Piecemeal testing also diminishes the ability of the researcher to account for the overall variation in the model using the R^2 statistic (percent of variance explained in a dependent variable), which leads to underestimation of the magnitude of effects. This increases the likelihood of Type II errors (false negatives). "It is possible in regression, for example, to misinterpret the underlying causality in that no single run can parse out all the variance in complex research models" [41, p. 17].

SEM statistical models represent causal relationships as paths. A *path* is a hypothesized correlation between variables representing the causal and consequent constructs of a theoretical proposition. Each path, therefore, is a hypothesis for testing a theoretical proposition. If a theory proposes, for example, that perceived ease of use causes an intention to use a system [47], then SEM would represent that relationship as a path between the variables that measure ease of use and intention to use. Paths are often presented as arrows in diagrams of SEM statistical models, with the arrows pointing in the proposed direction of causation.

An SEM statistical model can have a path for every proposition in a theory. This inclusiveness allows for complete testing of multistaged theoretical relationships [41]. SEM

> maps paths to many dependent (theoretical or observed) variables in the same research model and analyze[s] all the paths simultaneously rather than one at a time. [41, p. 10]

SEM also allows researchers to test relationships among unobservable, latent constructs [38]. These features are particularly important for building theories because theories rarely involve simple one-way, single-stage relationships. Phenomena of interest often "occur in a complex network of causal links" [48, p. 33]. Hence, when endogenous constructs that represent effects rather than causes are added to a theoretical model, SEM techniques come to different conclusions that are almost always more accurate.[8] The interpretation gap between 1G and 2G techniques may cause "subtle or even gross differences between analytical inferences about statistical conclusion validity" [41, p. 20].

[7]The theoretical implications of this section should be especially troubling to empirical behavioral researchers because of the current state of our use of statistical techniques. As an example, Boudreau et al. [46] found three disturbing conditions: (1) studies not fully validating their instruments; (2) reliability analyses typically involving only Cronbach's alpha, which is itself a severely limited measure; and (3) rare use of advanced validation techniques for mature research streams. The last-mentioned condition is of particular concern because mature research streams often involve extensions of theory and application to different theoretical contexts, all of which demand revalidation of the instruments for appropriate theoretical interpretation. Boudreau et al. also show that many researchers are increasingly using existing instruments (which is desirable for reasons of efficiency) without fully validating them in the new theoretical context. Moreover, they often use them blindly, not considering previous validation controversies. Another problem is that by not creating new instruments to measure existing constructs, researchers lose the opportunity to establish nomological validity in mature research streams. Unsurprisingly, Boudreau et al. found that "published studies making use of second-generation statistical techniques (SEM) are much more likely to validate their instruments than published studies making use of first-generation statistical techniques" [46, p. 11].

[8]We do not claim that SEM is better at establishing causation; it is better at representing the complex network of causal links necessary for establishing causation to be in accordance with the theoretical model. Regardless of the statistical technique used, causation requires association, temporal precedence, and isolation [49]. Importantly, "statistical analysis alone cannot prove causation, because it does not establish isolation or temporal ordering" [41, p. 40].

(3) SEM Avoids Fixed-Scale Construction: Another theoretical problem in 1G techniques is that these techniques do not account for error measurement in the testing of nomological (causal) links from IVs to DVs [45]. Researchers using 1G techniques might assess measurement error through a reliability analysis and then neglect or constrain this error in a subsequent analysis, causing the problem of fixed-scale construction, among other problems [45]. Fixed-scale construction occurs when researchers create indices of averages, sums, or weighted averages across measurement items when they have multiple measures of a construct. Since there is no accounting for any measurement error of the indicators in this approach, the techniques themselves cannot reveal or account for differences in measurement error among measurement items. In such cases, the researcher is assuming *de facto* that the measurement error is trivial. This underlying assumption is unrealistic for causal research in which the most important variables are, of necessity, subject to measurement error [45]. This problem extends to the use of covariates (exogenous variables that co-vary), such as in step-down analysis and ANCOVA/MANCOVA, in which measurement error in covariates is not accounted for [4].

1G techniques also restrictively assume homogeneity of variance and covariances of all dependent variables, whereas PLS does not require this assumption [4]. SEM avoids the problem of fixed-scale construction by allowing all indicators of every theoretical construct to be fully represented and analyzed in the model. SEM simultaneously accounts for measurement error as a theoretical model is tested. SEM assumes that unexplained variance is due, in part, to measurement error; therefore, employing SEM decreases the likelihood of Type II errors (false negatives)[9] [4], [38]. SEM also allows covariates to be treated as constructs so that their measurement error can be accounted for [4], which further reduces the likelihood of Type II errors.

[9]Recall that Type I errors (false positives) involve rejecting the null hypothesis and accepting the alternative hypothesis when the null hypothesis is actually true. Conversely, Type II errors (false negatives) involve accepting the null hypothesis and rejecting the alternative hypothesis (because of a lack of statistical power) when the alternative hypothesis is actually true.

(4) SEM Better Tests Moderators: Many causal behavioral theories involve constructs that moderate the relationships between other constructs [50]. A moderator is a construct that affects the strength of a causal relationship between two other constructs. Since 1G techniques typically do not model measurement error and suffer from fixed-scale construction, they have greater difficulty in detecting moderation effects (e.g., lack the necessary statistical power), which manifest as interactions in statistical tests[10] [45]. Lacking empirical support, researchers may abandon useful theoretical propositions; this potentially undermines theoretical development. Depending on the variance of the data and the effect size, 1G methods can require sample sizes 4–10 times larger than would be necessary to detect an interaction using PLS for example [51].

To summarize, the problem with the two-step approach used in 1G techniques is not the number of steps required to establish factorial validity and test a theoretical model; complex SEM models may also require multiple steps for final validation and testing, especially when formative indicators are involved. The problem is that 1G techniques typically change the nature of the measures after they have been validated (e.g., fixed-scale construction) and do not account for the theoretical relationships among all measures in a model during analysis. SEM techniques overcome these issues because measures and theory are tested together and all of the indicators in the measures are fully accounted for, which avoids fixed-scale construction.

Key Lessons

In this section, we compare and contrast PLS with CB-SEM, discuss sampling issues, and provide a step-by-step example of analyzing a causal model using PLS. Specifically, the key lessons include: (1) appropriately choosing to use PLS and CB-SEM, (2) considering these sampling guidelines, and (3) using this empirical demonstration of PLS and video supplement to see PLS in use.

[10]Interaction effects involve moderator variables, which can be qualitative or quantitative in nature and affect the direction or strength of the relationship between an IV and DV [45]. These are in contrast to mediators, which connect the relationship (or mediate) between two variables and are more easily depicted with SEM.

Lesson 1: Appropriately Choose to Use PLS and CB-SEM There are two forms of SEM. One is covariance based and represents constructs through factors (CB-SEM); the other is least squares based or components based and represents constructs through components (PLS). Although most of the characteristics and advantages of CB-SEM also apply to PLS, PLS can provide advantages over 1G techniques and CB-SEM techniques for preliminary theory building, while CB-SEM has advantages over PLS in terms of model validation. PLS incorporates several statistical techniques that are not part of CB-SEM—such as principal components analysis, multiple regression, multivariate analysis of variance, redundancy analysis, and canonical correlation [45]—without inflating the t-statistic, as would happen if each analysis were conducted separately from the others. To ensure fixed-scale construction never occurs, "the PLS algorithm allows each indicator to vary in how much it contributes to the composite score of the [construct]" [51, p. 25]. PLS is also especially useful for models that have higher-order constructs (e.g., third- or fourth-order constructs) [52].[11] CB-SEM, however, allows for the comparison between observed and proposed covariance matrices, which enables assessment of the overall "fit" of the proposed causal model. We discuss these and other differences that can vastly affect theoretical modeling next. In particular, we present how these differences establish a case for the preference of PLS when engaging in theory development or exploratory causal modeling.

(1) Factor Indeterminacy: Perhaps most important, PLS has a different goal than CB-SEM. CB-SEM seeks to model the covariation of all the indicators [6] to demonstrate that the assumed research model (the null hypothesis) "is insignificant, meaning that the complete set of paths, as specified in the model that is being analyzed, is plausible, given the sample data" [41, p. 24]. The primary objective of PLS, in contrast, is to demonstrate that the alternative hypothesis is significant, allowing the researcher to reject a null hypothesis by showing significant t-values and a high R^2 [5] (as argued in [41, p. 24]).

[11]A latent construct that has a direct formative or reflective relationship with its indicators is called a *first-order construct*. A latent construct in formative or reflective relationships with other latent variables is called a *higher order* construct. The order of a given construct is determined by the number of paths one would have to traverse to obtain from the construct to an indicator.

The differences in the goals of CB-SEM and PLS, as well as the differences in the underlying calculations performed with these techniques, create a greater contrast between PLS and CB-SEM than many researchers may realize. Due to its different theoretical goal, CB-SEM can cope with imperfect measurements better than PLS can; this ability is, in many cases, useful. CB-SEM analysis, however, often ends with factor indeterminacy [53], which means that it produces more than one solution that is mathematically sound without providing a means to determine which of the several solutions corresponds to the hypothesis being tested.[12] As noted, "an infinite number of unobservables may bear the same pattern of correlations with observed variables and yet be only weakly or even negatively correlated with each other" [53, p. 449]. As a result, CB-SEM is very unreliable in the exploratory analysis required for theory building [53], [55]. However, CB-SEM is ideal for testing the full nomology of a known theory and testing general model fit. Using this technique, the proposed causal model is compared to the covariance matrix in order to determine if the proposed model is a sufficiently "good" (such as appropriate) way to model the relationships among the variables. PLS does not have this capability [38], [56].

Since CB-SEM can support a "large number of alternative, but statistically equivalent, models that can be supported by the same data" and because of "over-fitting" [41, pp. 40–41], it becomes difficult to argue causality using a CB-SEM analysis. For these reasons, CB-SEM should be used to test only well-established theories that are empirically validated. It can be used safely only for confirmatory analysis in which well-established theoretical arguments can be used to overrule competing explanations [53], [57]. Even robust theoretical arguments, however, are not always sufficient to resolve the indeterminacy problem because multiple hypotheses can still equally well account for the same data [53], [57]. Despite this fact, research is replete with articles that discuss new theories using LISREL and AMOS (two common CB-SEM tools) and make causal claims that the results cannot support.

[12]See [54] for excellent coverage of the problem of factor indeterminacy.

Thus, researchers can help themselves avoid unsupportable conclusions by using PLS "for exploratory analysis and for testing developmental theories" [53, p. 451]. PLS avoids factor indeterminacy by composing constructs from the factor scores and using these in subsequent calculations, yielding explicit factors scores [53]. Since PLS avoids factor indeterminacy, it can then be used for confirmatory and exploratory studies. Accordingly, PLS does not require the theory being tested to already have empirical support that is well established from other sources [41].[13]

(2) Data Distribution Flexibility: PLS also differs from CB-SEM in the way it deals with the unknowns in model estimation [50], [58]. The PLS approach to prediction occurs iteratively; each step minimizes the residual variance of the theoretical and observed dependent variables to obtain parameter estimates.[14] Once PLS has obtained the parameter estimates, it calculates the significance of each path in the model using a t-test. Unlike 1G techniques, PLS does not need to assume that the DVs conform to any particular distributions. As a result, it is robust to violations of multivariate normal distributions [41], whereas CB-SEM (which relies primarily on maximum likelihood estimation) assumes data normality [59]. As a result, PLS allows more flexibility in analyzing theoretical models [41]. Specifically, PLS can calculate t-values through a technique called bootstrapping if the data are normally distributed and samples are independent. If data distributions are not normal or samples are not independent, PLS can calculate t-values with a technique called jackknifing (a.k.a. blindfolding).[15] If a path is found to be statistically significant, then the null hypothesis for that path can be rejected, and the statistical model can be interpreted as providing empirical support for the hypothesis represented by the path. Support for the hypothesis, in turn, can be interpreted by the researcher as support for the theoretical proposition that the hypothesis was meant to test.

(3) Construct Specification: One modeling and theoretical limitation of CB-SEM (as with 1G factoring techniques) is that it assumes that one is using reflective indicators rather than formative indicators in a model. A *reflective* indicator is an observed variable that is assumed to be an *effect* of a latent construct. The underlying construct is assumed to cause the values that manifest in the observed variable. If a system is robust, for example, it may be assumed that the effect of that robustness will be low mean time between failures and short downtimes. A key implication of this assumption would be that changes in the latent construct would manifest as changes in *all* of its indicators [60]. The latent construct is said to determine its reflective indicators [61]. Since all reflective indicators of a latent construct are assumed to be caused by the construct, reflective indicators would have to co-vary. Therefore, measures of convergent validity would be important to ensure that variations in one indicator are consistent with variations in the other reflective indicators of the same latent construct.

In contrast, a *formative* indicator is a variable measuring an assumed *cause of* or a *component of* a latent construct. Under this conception, a

[13]Note, however, that we do not imply that the PLS method will develop a theory for the researcher. A theory is a collection of assumptions and propositions derived from those assumptions that offer a logically consistent explanation for variation observed in a phenomenon of interest. A statistical test can only demonstrate the strength of correlations between variables and indicates the probability that such correlations are a result of random chance. A variable is a way to measure a theoretical construct in a particular context. In exploratory research, variables are not yet linked to constructs, and constructs are not integrated into causal relationships derived from underlying assumptions. Thus, PLS can reveal the unexpected existence of correlations among variables, but the researcher still must derive a rigorous theory to explain such discoveries.

[14]A *parameter* is a numerical value that represents some literal aspect of a population of scores (in this case, the true model from the actual population). Since it is rare that one can actually measure an entire population of scores, we use estimates of the parameters (a.k.a., *parameter estimates*), which are the statistics that are computed from our samples.

[15]Bootstrapping and blindfolding are nonparametric techniques that are built into PLS to improve model estimation with PLS (one chooses to do one or the other when performing a PLS analysis). *Bootstrapping* is a way of computing sampling error and generating t-values by using the available data as a distribution. Bootstrapping assumes independent residuals (residuals are the discrepancy between the actual values and the estimated values); they can be swapped without undermining the estimates [50]. In contrast, blindfolding makes no assumptions of independence.
Blindfolding is a resampling procedure that generates "jackknifed" estimated means and standard deviations (by omitting part of the examined construct's data matrix and repeatedly estimating the model parameters based on the blindfold omission distance) as another way to determine path significance [50]. This process of leaving out and reconstructing data repeats itself until every data point is left out and reconstructed once [50]. This technique provides two results [50]: the generalized cross-validation criterion that can be used to evaluate the model and the results of the jackknife technique. The jackknife technique results are the distribution of the parameter estimates (individual parameter estimates' standard errors) and do not require distribution assumptions [50]. Accordingly, blindfolding is particularly useful in dealing with unknowns, requires no distributional assumptions or independence, and, so, fits perfectly with PLS [50]. More specifics on these two techniques can be found in [42].

latent construct is assumed to be defined by or a function of its indicators. A key implication of such an assumption would be that changes in the latent construct would not necessarily be matched by changes in all of its indicators. Changes in a single indicator, however, would be enough to predict a change in the latent construct. It would be possible for formative indicators to vary independently of, or even inversely with, one another [62]. Therefore, measures of convergent validity would not be meaningful for formative constructs. The quality of an information system, for example, might be defined by the speed, accuracy, and completeness of the information it provides to decision makers. It would be possible, however, to improve the speed of a system by reducing the completeness and accuracy of the information it provides. Thus, it would not be appropriate to frame speed, accuracy, and completeness as reflective indicators of system quality. However, it could be useful to frame them as formative indicators, since speed, quality, and completeness comprise system quality.

The CB-SEM and 1G assumption that all indicators are reflective can result in serious modeling errors [38] that produce inappropriate results. This is particularly salient to behavioral research where *mixed models*—those comprising reflective and formative indicators—are common. Thus, when a theoretical model includes formative indicators (or a mix of reflective and formative), it is important to use an appropriate statistical technique, such as PLS, that can account for both indicators in its statistical model.

(4) Moderation and Model Complexity: CB-SEM improves on many of the 1G problems of detecting interaction effects. However, CB-SEM techniques, such as LISREL and AMOS, are not as sensitive to moderator effects as PLS is [45], since PLS is generally better at dealing with measurement error. This, in turn, decreases the sample size requirement [45]. Chin et al. [45] show through a Monte Carlo simulation that PLS can be combined with a product-indicator approach to measure interactions and moderation effects more effectively than can be accomplished with CB-SEM. PLS software, such as SmartPLS [63], also includes specific design features to ease the analysis of interactions. However, analyzing categorical (grouped) moderators, such as gender, is far simpler in CB-SEM tools, such as AMOS. For example, AMOS has built-in design features that assist with multiple group analyses, whereas SmartPLS does not.

The other typical problem when using interaction variables as moderators with CB-SEM is that interactions dramatically increase the number of indicators and underlying complexity of a model, which CB-SEM is not well equipped to handle. When CB-SEM encounters complex models, it requires very large samples for estimation accuracy and is limited in working with only a relatively few variables in order to achieve convergence [53], [58]. This requirement leads to further problems, as when the number of measures, factors, and levels within factors increase in CB-SEM, the model requires a much larger number of free parameters[16] for estimations. This increases the chance of nonconvergence (model failure) and improper solutions [4]. As a result, CB-SEM should be used to analyze models with a maximum of 40–50 indicators, even with large sample sizes, in order to prevent model nonconvergence [45]. Again, this is a common problem in the behavioral research literature where several articles have used LISREL and AMOS to analyze much larger models.

When to Choose PLS or CB-SEM: To summarize this section, in choosing whether to use PLS or CB-SEM, one should initially consider whether the research is exploratory (building or testing a new theory) or confirmatory (testing a well-established theory). For exploratory work, PLS should be selected. For confirmatory work, either technique may be used. In Table I, we recommend when to select PLS or CB-SEM given particular modeling considerations—when the work is confirmatory.

Lesson 2: Consider These Sampling Guidelines
Scholarly studies often claim that sample size requirements vary across analytic approaches. For example, scholars frequently justify their use of PLS (as a "limited information" estimation procedure) due to its assumed ability to handle lower sample sizes [64]. However, increasing evidence exists that in many instances, PLS requires a comparable sample size to that used in other techniques [64]. One commonly used heuristic for determining the minimum required sample size in PLS is to multiply 10 times the "scale with the largest number of formative (such as causal) indicators" or to multiply 10 times "the largest number of structural paths directed at a particular construct in the structural model" [51, p. 39]. However, this heuristic has recently been criticized as too lenient and used too often in PLS research [52]. Thus, to be more accurate, "one needs to specify the effect size for

[16]An observation that is needed to define a model enough so that predictions can be made, but which must be determined by experiment or observation.

TABLE I
RECOMMENDATIONS ON WHEN TO USE PLS VERSUS CB-SEM

Model Requirement	PLS	CB-SEM
Includes interaction effects	Preferable, as it is designed for easy interactions	Difficult with small models, nearly impossible with large ones
Includes formative factors	Easier	Difficult
Includes multigroup moderators	Can use, but difficult	Preferable
Testing alternative models	Can use	Preferable, as it provides model fit statistics for comparison
Includes more than 40-50 variables	Preferable	Sometimes unreliable if it does converge; sometimes will not converge
Nonnormal distributions	Preferable (although it will still affect results, just to a lesser extent)	Should not be used; results in unreliable findings.
Nonhomogeneity of variance	Preferable (although it will still affect results, just to a lesser extent)	Should not be used; results in unreliable findings.
Small sample size	It will run (although it will still affect results negatively)	Unreliable if it does converge; often will not converge

each regression analysis and look up the power tables" [65, p. 327]. Furthermore, nonnormality of PLS does cause issues with power analysis and needs to be further considered [52]. Thus, before gathering data, it is important to determine the sample size necessary to achieve reasonable power. Accordingly, PLS users still need to follow basic statistical guidelines on power. Failure to do so, can lead to using inappropriate sample sizes [64]. Sample size still affects the stability of the estimates [38], particularly when dealing with complex interaction terms. For example, Goodhue et al. [66] compared the necessary sample sizes for 1G techniques, PLS, and CB-SEM. They found that sampling requirements actually did not significantly differ between these techniques with respect to achieving sufficient power. Thus, while PLS and 1G techniques may run when using lower samples (whereas CB-SEM often cannot run with low samples), the estimates may still be unstable and cannot be relied upon due to low power [66].

Lesson 3: Use this Empirical Demonstration of PLS and Video Supplement to See PLS in Use To further illustrate the potential strength of applying PLS to theory building in a communication context, we provide an example of PLS analysis using data collected from 346 participants in a large group-communication quasiexperiment using an audience response system (ARS) that took place over a year. Full details of the theory, model, procedures, and the experiment appear in [67]. For purposes of tutorial demonstration, we take the theory and the data as received. The measures are repeated in [67, App. 3]. Here, we only briefly overview the method procedures. A total of 346 undergraduate business majors at a large, public, university in Southern California participated, providing more than adequate apriori power. All participants were enrolled in one of two sections of the same introductory-level information systems course. A total of 60.7% of participants were male and 39.3% were female. Average age was 22 (SD 2.8). Average GPA was 3.01 (SD 0.44). Ethic distribution was Asian (49.7%), Caucasian (16.5%), Hispanic (15.0%), African (1.7%), and other/no response (17.1%).

This study employed a quasiexperimental nonequivalent groups design with multiple outcome measures, which is appropriate when random assignment is not possible. Two large sections met for two instructional quarters of the same course: large-group interaction without ARS (control group) and large-group interaction with ARS (treatment group). The ARS tool Classroom Performance System (CPS) by EInstruction, Inc. was used for this research. CPS provided all group members with a small, handheld, eight-button response pad that transmitted an infrared signal to a receiver connected to the facilitator's computer. The system's software recorded participant responses and graphically displayed results in real time. The study established the efficacy of ARS in increasing

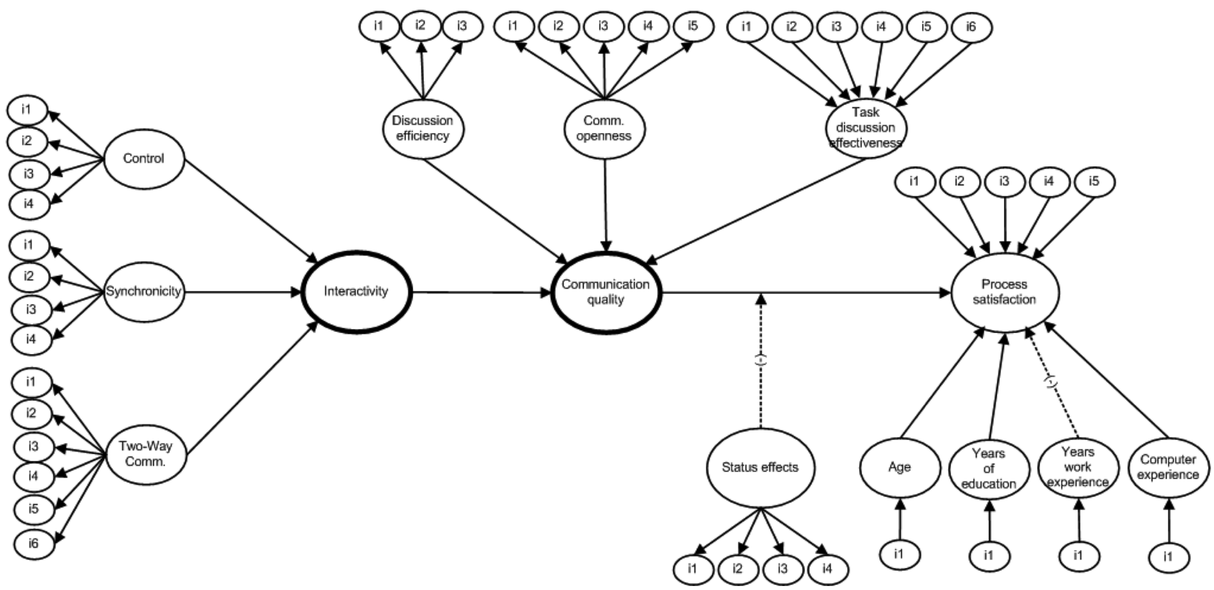


Fig. 1. Theoretical model used for demonstration.

interactivity and communication quality in large groups, compared to control groups who followed the same activities and procedures without use of ARS.

The data for the model that we will be testing were collected to test a theoretical model of *process satisfaction*. The model proposes that *communication quality* is a mediator between *interactivity* and *process satisfaction* and that *status* is an inverse moderator of the relationship between communication quality and process satisfaction. Since we present the model here only as a device for demonstrating the utility of PLS, we do not define the constructs or explain the theory in this tutorial; neither do we discuss the measures, nor do we defend the experimental procedures under which data were collected. For purposes of demonstration, we take the theory and the data as received. Details of the theory and the experiment appear in [67].

The model depicted in Fig. 1 includes formative and reflective indicators. Positive relationships are depicted with a solid line, while inverse relationships are depicted with a dotted line. In this model, interactivity is a second-order formative factor composed of the reflective constructs of control, synchronicity, and two-way communication. Communication quality is a second-order formative factor composed of the reflective constructs of communication openness and task discussion effectiveness and the formative construct of discussion efficiency. Process satisfaction is a first-order formative construct. Status is a first-order reflective construct that is proposed to negatively moderate the relationships between communication quality and process satisfaction. In addition, four covariates are proposed to affect process satisfaction. Finally, communication quality is proposed to be a full mediator in the relationship between interactivity and process satisfaction.

The interactivity model of process satisfaction has a number of characteristics that would make it a good candidate for PLS analysis. First, it is a mixed model of formative and reflective indicators, which 1G techniques and CB-SEM cannot test. Second, the model contains 41 indicators, which would be virtually impossible to test with 1G techniques because there are too many indicators and pathways, and this is pushing the limits of appropriateness for CB-SEM [41]. Finally, once the interaction items were included in the model (to test the proposed moderating relationship), our model grew to 159 indicators (not depicted in Fig. 1). This is much larger than the 40–50 indicator limitation established to prevent model nonconvergence in CBSEM [45]. Having established the model to be tested, we will now explain the testing process. For this analysis, we used SmartPLS version 2.0 [63], which is available freely to academics at http://www.smartpls.de. See [68] for a guide to setting up SmartPLS, including prepping, loading, and troubleshooting data.

Supplementary Videos for Our Tutorial: Since the use of a manuscript for communicating the mechanics of analytical procedures is quite limiting, we also include links for step-by-step video tutorials. We propose that this composite approach to a tutorial manuscript should increase the accessibility of the material in order to make following and replicating the procedures much easier. The value of supplementing traditional paper tutorials with online video tutorials is manifest in the heavily trafficked site we will be referencing (Gaskination's YouTube channel),[17] as set up by one of the coauthors of this tutorial. This channel is devoted almost exclusively to SEM video tutorials and demonstrations of procedures using CB-SEM and PLS approaches. The site is heavily used as a resource for researchers across the world conducting quantitative studies. The 75 videos have been viewed (currently) in 195 countries, around half a million times, for a total viewing time of nearly one million minutes. The site also acts as an active forum around the videos, currently with more than 2000 posts. The YouTube channel currently has eight videos dedicated to PLS analyses in SmartPLS, the software we will be using is in the empirical demonstration as will be shown. These nine videos have been viewed around 20,000 times since their posting less than 10 months ago.

Step 1: Model Specification: Before we run a PLS analysis, we need to configure the model in a way that will produce the kinds of results we need. To do this, we must carefully establish which indicators are formative and which are reflective [69], [70]. This is important because the tests to establish the factorial validity for reflective indicators are quite different than the approach used to validate formative indicators (see [62][18] for specifics). Incorrectly specifying the indicators can increase both Type I and Type II errors [62], [70]. In SmartPLS, to set a factor as formative, simply right click the factor and select "invert measurement model."

In our demonstration, we modeled the five indicators of communication openness as reflective because the various items are interchangeable (e.g., "It was easy to communicate openly to all members of this group," "Communication in this group was very open," and "When people communicated to each other in this group, there was a great deal of understanding."). Any change in communication openness should be matched by similar changes in all of its indicators. Conversely, we characterized the indicators of task discussion effectiveness as formative because the construct is a composite of these indicators, rather than a cause of them (e.g., "The context of the discussions was carelessly developed," "Participation in the discussions was unevenly distributed," "Ideas in the discussions were uncritically examined," and "The amount of information exchanged was sufficient"). Removing or replacing any of the items would change the meaning of the construct; they could reasonably be expected to vary inversely under some conditions. Since these indicators are not interchangeable, they must be framed as formative rather than reflective.

The degree to which the first-order constructs contribute to the second-order constructs is established by creating them as a molar model, as outlined and discussed in [71]. To do this in PLS, we created a second-order construct that contained all indicators of its first-order subconstructs and then ran a model with the subconstructs predicting the second-order construct. This "repeated indicator" approach works well when the second-order construct is either reflective or exogenous (such as predictor only). However, when the second-order construct is endogenous and formative, the repeated indicators in the second-order construct are perfectly predicted by the first-order constructs, which also contain those indicators. Therefore, all other potential effects from other predictors are effectively swamped out, and the R^2 for the second-order construct is 100%. To overcome this issue, we must take a two-step approach by first modeling the measurement model and obtaining the latent variable scores for the second-order construct (and all other top-level constructs). Then, we must create a new model that uses the latent variable scores as indicators of the constructs. A demonstration of this two-step approach is offered on Gaskination's YouTube Channel [72].[19]

Step 2: Establish Construct Validity of Reflective Constructs: An example of a factor analysis in SmartPLS is also offered on Gaskination's YouTube channel [73]. Establishing validity and testing the entire path model occurs in one pass using PLS by running a bootstrap of the model using 200 (or more) resamples.[20] In doing so, we performed a confirmatory factor analysis (CFA) that was

[17]This channel is available at http://www.youtube.com/user/Gaskination.

[18]In our tutorial, we intentionally have used the neutral term *construct* to refer to those that may have reflective or formative indicators because, as indicated in [62], researchers typically use the term *latent variables* to refer to constructs that have reflective indicators and the term *formative constructs* to refer to constructs that have formative indicators.

[19]"SmartPLS Formative 2nd order Constructs": http://www.youtube.com/watch?v=kPeUTKjMF7o.

[20]To do this in SmartPLS, simply select "bootstrapping" instead of "PLS algorithm" in the calculate menu.

executed as part of the PLS run. Following the procedures outlined in [43] and [74], we first established convergent validity for the reflective constructs by checking whether the measurement items were loaded with significant t-values on their theoretical constructs. All of our reflective indicators were significant at the α 0.05 level on this test. We then examined the t-values of the outer model loadings (an estimation of how much a particular indicator loads onto a construct in the model).[21] All of the outer loadings were significant at the 0.05 α level (see Table A1.1 in the online appendix, available online as downloadable supplementary material at http://ieeexplore.ieee.org). These results indicate strong convergent validity in our model for the constructs. SmartPLS also offers a measurement model table (akin to a pattern matrix) with the inner model loadings.[22]

To determine the discriminant validity of our indicators, we used two established techniques. First, we correlated the latent variable scores against the indicators. In SmartPLS, the latent variable scores can be found in the “Index Values for Latent Variables” portion of the “Index Values Results” section of the default report after running the PLS algorithm. In a separate spreadsheet that contains the individual items, we ran Pearson’s correlations of all the items against the latent variable scores. These correlations represent a confirmatory factor analysis in which the correlations are the actual loadings of the indicators on all of the constructs in the model. Although exact guidelines governing the validity of such results have not yet been established, we proceed based on the general rule that “all the loadings of the measurement items on their assigned constructs should be an order of magnitude larger than any other loading ... For example, if one of the measurement items loads with a 0.70 coefficient on its latent construct, then the loadings of all measurement items on any latent construct but their own should be below 0.60” [43, p. 93]. Anything outside these guidelines would constitute a violation in discriminant validity and would be dropped. Using latent variable scores, strong discriminant validity was established for all items except for the fifth item of synchronicity, which we therefore dropped (see Table A1.2, available online as downloadable supplementary material at http://ieeexplore.ieee.org).

To confirm the discriminant validity of our indicators further, we calculated the average variance extracted (AVE):

> Conceptually, the AVE test is equivalent to saying that the correlation of the construct with its measurement items should be larger than its correlation with the other constructs [43, p. 94]

which is similar to correlation tests with multitrait, multimethod (MTMM) matrices. To perform this test, we ran a correlation of each variable with each other variable and then compared these correlations to the square root of the AVE for each construct. The AVE is calculated in SmartPLS by computing the variances shared by the items of a particular construct[23] (See Table A1.3, which is available online as downloadable supplementary material at http://ieeexplore.ieee.org.) In the table, the AVE square roots are represented as the bold and underlined diagonal elements. The offdiagonal elements in Table A1.3, which is shown as downloadable supplementary material at http://ieeexplore.ieee.org, represent the correlations between the constructs. To establish discriminant validity further, the diagonal elements must be greater than the offdiagonal elements for the same row and column, not the AVE value itself. The AVE analysis showed very strong discriminant validity for all subconstructs, further confirming our choices of items to retain and drop.

Step 3: Establish the Reliability of the Reflective Constructs: *Reliability* refers to the degree to which a scale yields consistent and stable measures over time [75] and applies only to reflective indicators. PLS computes a composite reliability score (similar to Cronbach’s alpha in that they are both measures of internal consistency) as part of its integrated

[21] In SmartPLS, this becomes the default report when you run “bootstrapping” instead of the PLS algorithm. The data necessary to demonstrate convergent validity are found in the “outer loadings (mean, STDEV, tvalues)” section of the report. Specifically, one examines the t-values of each item, and the p-value of each t-value needs to be significant at the 0.05 alpha protection level (needing a t-value of about 1.96 or greater—absolute value) or the specific item demonstrates a lack of convergent validity on that factor.

[22] In SmartPLS, the pattern matrix can be found in the “cross loadings” portion of the “quality criteria” section in the default report after running the PLS algorithm.

[23] These are automatically generated by SmartPLS and can be found in the “overview” portion of the “quality criteria” section in the default report after running the PLS algorithm. This is also where you will find the composite reliability and Cronbach’s alpha scores.

model analysis. (See Table A1.4, available online as downloadable supplementary material at http://ieeexplore.ieee.org).[24] Each reflective construct in our research model demonstrated a level of reliability well above the recommended threshold of 0.70 [42].

Step 4: Establish Construct Validity of Formative Indicators: Researchers have traditionally used theoretical reasoning to argue the validity of formative constructs [69]. The procedures for determining the validity of reflective measures do not apply to formative indicators [62], [74], since formative indicators may move in different directions and can theoretically co-vary with other constructs. The concepts of construct validity and reliability, therefore, do not apply to formative constructs. Although statistical approaches are emerging to test the construct validity of formative indicators [62], [76], no single technique is universally accepted for validating formative measures. However, the modified MTMM approach, as presented in [76] and [77], is regarded by many as a promising solution. Therefore, we inspected the MTMM table produced by our PLS run[25] (Table A1.5, available online as downloadable supplementary material at http://ieeexplore.ieee.org) and concluded that convergent validity was highly likely. However, a simpler approach is just to ensure the indicator weights for formative constructs are roughly equal and all have significant t-values [78].[26]

Because multicollinearity poses a greater problem for formative indicators than for reflective indicators, we also used the approach suggested by Petter et al. [62] to assess formative validity, which involves testing the multicollinearity[27] among the indicators using regression. An example of how to detect multicollinearity in SPSS is shown here [79]. At a maximum, the variance inflation factor (VIF) for formative factors should be 10, but for a more rigorous test, they should be below 3.3 [62]. In our case, all of the VIFs of the indicators were below 3.3, indicating sufficient construct validity for our formative indicators. Had any indicators scored higher than 10, we would have had to drop them from the model. As with CB-SEM techniques, if items were dropped at any point during the factor analysis, we would have to restart the factor analysis because of the highly interrelated nature of variables in SEM analyses. Removing a single item can alter the inner and outer model loadings and weights. Thus, the final reported validity statistics should be those gathered once all changes to the structure of the measurement model are complete.

Step 5. Test for Common Methods Bias: Since the endogenous variables were collected at the same time and using the same instrument as the exogenous variables, we tested for common methods bias to establish that such bias did not distort the data we collected. However, we also acknowledge that there is increasing debate as to how serious this bias is [80]. This is an important consideration in most behavioral research; thus, it should be accounted for after construct validity is established. To do so, we used two approaches. First, we examined the exploratory, unrotated factor analysis to find the results of Harman's single-factor test for all of the first-order constructs using a standard statistical package. The aim of the test is to determine if a single factor emerges that explains the majority of the variance in the model. If so, then common method bias likely exists on a significant level. The result of our factor analysis produced 27 distinct factors, the largest of which accounted for only 29.89% of the variance of the

[24]These are automatically generated by SmartPLS and can be found in the "overview" portion of the "quality criteria" section in the default report after running the PLS algorithm.

[25]For reflective measures, loadings were used because they "represent the influence of individual scale items on reflective constructs; PLS weights represent a comparable influence for formative constructs" [77, p. 49]. For formative items, we created new values that were the product of the original item values by their respective PLS weights (representing each item's weighted score, found in the "outer weights" section of the PLS calculation results section of the default report after running the PLS algorithm). Then, we created a composite score for each construct by summing all of the weighted scores for a construct. Finally, we produced correlations of these values, providing intermeasure and item-to-construct correlations.
To test convergent validity, we checked whether all of the items within a construct highly correlate with each other and whether the items within a construct correlate with their construct value. This was true in all cases, leading to the inference of convergent validity. While we would ideally want interitem correlations to be higher within a given construct, this cannot be strictly enforced since there are exceptions depending on the theoretical nature of the formative measure [69], [77]. In addition, large matrices would introduce exceptions that are not necessarily meaningful and, thus, careful theoretical judgment needs to be used before removing any items [62]. Thus, we believe the most meaningful discriminant validity check with formative measures is to examine the degree to which items within a construct correlate to a given construct.

[26]In SmartPLS, the indicator weights and their t-values can be found in the "outer weights (means, STDEV, t-values)" portion of the bootstrapping section of the default report after running the bootstrapping algorithm.

[27]This is a statistical phenomenon when two or more predictors are highly correlated. Multicollinearity generally does not reduce the predictive power of the model itself, but it often reduces the validity of specific predictor results.

model. This suggested that our data did not suffer from common methods bias.

Due to the growing dispute about the merits of Harman's single-factor test [81], we corroborated the results from Harman's single-factor test by examining a correlation matrix of the constructs (using Pearson's correlations) to determine if any of the correlations were above 0.90 among the formative indicators. Had there been correlations that high, it would have given strong evidence that a common methods bias existed [82]. In no case did our correlations reach this threshold; thus, the likelihood of common methods bias is low.

Moving beyond the scope of this tutorial, more advanced approaches can be applied to test common methods bias further, should simpler techniques not provide clear results. A leading approach with PLS is to include a marker variable in the data collection that is unrelated to the model [83]. In this use, a researcher would correlate the data to the marker variable, and if the correlations are high, then common methods bias likely exists. Other approaches are extensively reviewed in [81] and [84].

Step 6. Test for Moderation Effects (If Applicable): Our theory proposed that the relationship between communication quality and process satisfaction was inversely (such as negatively) moderated by status—in other words, the lower a respondent's status, the more the respondent's process satisfaction would be affected by communication quality. Moderator relationships in a theory are tested statistically by checking for interaction effects among independent variables.

Whether moderators exist in a model is assessed by a hierarchical process similar to that used in 1G statistical techniques. First, two models—one with the moderator relationship and one without [45]—were constructed and compared. This process required two PLS runs—one for the baseline model and one for the interaction model. In creating the baseline model, the main effects of the interaction term need to be included, including status. We tested our model for its interaction term using the product-indicator (PI) approach proposed by Chin et al. [45] because this method is the most effective approach in identifying interaction terms in complex path models.[28,29]

Adding the PI interaction terms dramatically increased the number of indicators in the overall model to 159, rendering the model analyzable by PLS only. The interaction of status and communication quality was significant at an α protection level of 0.05 ($t = 2.85$). Adding in the interaction term decreased the beta coefficient of the path between communication quality and process satisfaction from the baseline model (from 0.798 to 0.699); the R^2 for process satisfaction also increased from 0.651 to 0.668. Consequently, in the interaction model, the negative path coefficient between status effect and PS was now significant ($t = 2.80$). Our significant interaction had an effect size of $f^2 = 0.05$, showing a small interaction effect[30]; however, even small effects using the product-indicator approach indicate important model relationships [45].

The interaction model is shown in Fig. 2. Variance is explained and indicated for each construct as

[28]This approach adds three critical improvements to measuring interaction effects. First, this approach models paths between each exogenous and endogenous construct—a critical step because "when the main effect variables are missing in the analysis, interaction path coefficients are not true interaction effects" [45, p. 196]. Second, it standardizes or centers the individual items for the moderation scores. "Standardizing or centering indicators helps avoid computational errors by lowering the correlations between the product indicators and their individual components" (pp. 198–199). Standardizing is used if it is thought that the indicators measure their constructs equally well. Since we had no theoretical reason to believe that there were unequal differences in the specific indicators, standardizing was our methodological choice. Third, no information is eliminated from the model. All of the interaction indicators stand alone without being summarized and are free to vary on their own to take advantage of PLS analysis.

[29]To add this interaction variable in SmartPLS, just right-click on the endogenous variable (in our case, process satisfaction) and then choose "create moderating effect." Then, select a predictor and a moderator (in our case, communication quality and status effects, respectively). This will automatically produce the interaction variable. A demonstration of this feature is offered on Gaskination's YouTube channel: http://www.youtube.com/watch?v=upEf1brVvXQ.

[30]To be conservative, we consider only the change in R^2, shown in the f^2 statistic, to be equivalent to effect size. This is because regression changes in β are less accurate indicators of effect size, especially if multicollinearity exists [85]. Since PLS and regression share similarities in how β is calculated, we also do not consider changes in β to be equivalent to effect size. We calculate f^2 as $[R^2\text{(interaction model)} - R^2\text{(main effects model)}]/[1 - R^2\text{(main effects model)}]$.

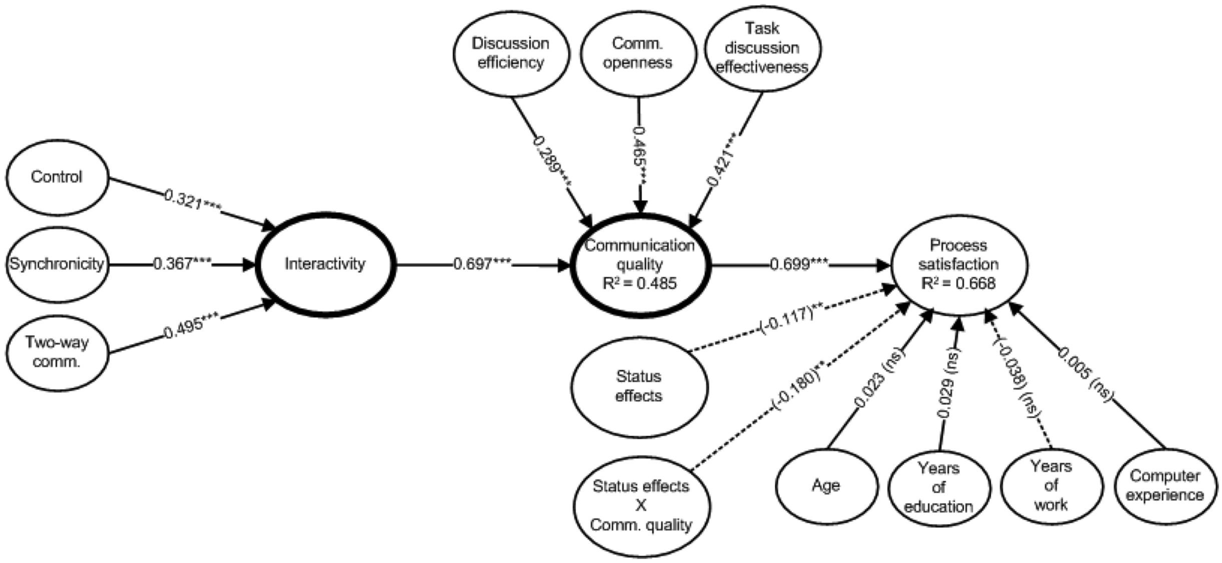


Fig. 2. Model with interactions of status effects using the PI approach.

R^2 (see[31]). The path coefficients or betas (βs)[32] are indicated numerically (from 0.000 to 1.000) on the paths between the two constructs, along with their direction and significance (again, negative relationships are noted with dotted lines). Numbers that appear on paths are the path coefficients (βs). Asterisks indicate significance levels: $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Gaskin [86] provides a video demonstration on how to conduct an interaction moderation analysis in SmartPLS.

Step 7. Test for Mediation (If Applicable): In addition to the moderation check, to establish the full nomological validity of our model, we performed a mediation check, a check that, by necessity, must be done in stages. A mediator is a construct in a causal chain between two other constructs. For example, if one were to assert that increasing system speed caused reduced cognitive load on users, thereby increasing user productivity, one would say that cognitive load mediates the relationship between system speed and user performance. Full mediation occurs when the IV no longer has a significant effect on the DV when the mediator is included in the model; partial mediation occurs when the IV still has a significant effect but when its effect is diminished when the mediator is included in the model [87].

In our model, communication quality is proposed as a mediator between interactivity and process satisfaction. That is, interactivity first increases communication quality, leading to an increase in process satisfaction. Interactivity specifically does not directly increase process satisfaction. Since we only had one mediator in our path model, we followed the simple test of mediation proposed by Baron and Kenny.[33,34] Comparing three different model runs in PLS corresponds to the three models required by Baron and Kenny's classic mediation test and comparing the resulting path coefficients and R^2s.

First, the unmediated path between interactivity process satisfaction (still including covariates and the interaction terms) had a significant β of 0.460 and produced an R^2 of 0.537 for process satisfaction. When the mediation relationship with communication quality was added, the new paths

[31]In SEM, variance explained (R^2s) for a particular DV represents the degree to which the percentage (0%–100%) of variance in the DV is accounted for by the IV(s) that predicts it. Thus, a key goal in SEM is to provide as high variances explained as possible for each DV. The less variance that is explained, the more that factors outside the model account for the DV's decrease of the explanatory power of the model.

[32]In SEM, a path coefficient is the partial correlation coefficient between the IV and DV, adjusted for other IVs. In short, it shows how much of an increase/decrease in the IV affects the DV.

[33]"A variable functions as a mediator when it meets the following conditions: variations in levels of the independent variable significantly account for variations in the presumed mediator (such as Path a), variations in the mediator significantly account for variations in the dependent variable (such as Path b), and when paths a and b are controlled, a previously significant relation between the independent and dependent variables is no longer significant, with the strongest demonstration of mediation occurring when Path c is zero" ([87, p. 1176]).

[34]Another method is to perform a bootstrap and then examine the "total effects (means, STDEV, t-values)" portion of the bootstrapping section of the default report. The t-statistic for the total effect will represent the "total effect" the predictor has on the dependent variable through the mediator (if no direct path is specified between the predictor and DV)—such as the mediated effect.

were significant (interactivity to communication quality had a β of 0.693 and communication quality to process satisfaction had a β of 0.648). Importantly, the direct path between interactivity and process satisfaction became nonsignificant with a β of 0.068. Meanwhile, the new R^2 for process satisfaction increased to 0.672. These results validated our model by providing strong evidence that communication quality acts as a full mediator and that predicting only a direct relationship between interactivity and process satisfaction is suboptimal and theoretically incorrect. An example of how to perform a mediation test in SmartPLS is available here [88].

Step 8. Assess the Predictive Power of the Model: Once the full model had been tested (establishing or rejecting the nomological validity of the model), we assessed the predictive power of the model (how well the model explains variance in the DVs), as demonstrated by the path coefficients and R^2s in the model (see Fig. 2). Chin [38] indicates that to demonstrate meaningful predictive power of a PLS model, one needs to show high R^2s and substantial and significant structural paths. To be "substantial," standardized paths need to be close to 0.20 (and ideally 0.30 or higher) to indicate that the model has meaningful predictive power (also see Fig. 2). In our test, the path between the interaction term and process satisfaction was lower, but again, even small interaction terms that are significant are important to a model [45]. Thus, we concluded that our overall model has excellent predictive power.

Step 9. Provide and Interpret Final Statistics: As the final step of the analysis of our model, we provide the measurement model statistics (Table A1.6, available online as downloadable supplementary material at http://ieeexplore.ieee.org) and a summary of the path coefficients and significance levels (Table A1.7, available online as downloadable supplementary material at http://ieeexplore.ieee.org).

Other Analyses and Concerns: SEM includes numerous other analyses. Although many of these are useful and commonly needed analyses, the aim of this tutorial is not to provide an exhaustive text on all possible SEM analyses, but to provide a simple tutorial for the <u>most</u> common and most needed analyses for standard SEM studies. Nevertheless, we offer a few additional guidelines in this section for the most common of these analyses.

Our illustrative study did not include multigroup moderation or moderated mediation; however, these types of relationships are increasingly included in research studies. The interested reader may refer to [89] for a video demonstration and explanation of how to conduct such analyses in SmartPLS. Mediated moderation is also becoming more common and can be conducted in the same way as normal mediated effects—with the only difference being that the exogenous variable in such a relationship is the interaction variable, rather than the regular independent (predictor) variable.

How to include control variables (sometimes called covariates or alternative hypotheses) is a point of confusion for many novice scholars. Control variables should be included for the express purpose of accounting for known or potential confounding effects on any construct in the model. The most common mistake regarding control variables is that the researcher will only control for the potential effects on the dependent variables. However, if a control variable is suspected to also affect a mediator or even an independent variable, these effects can be accounted for by drawing a line in SmartPLS from the control variable to the other variable. However, the inclusion of these effects should be theory driven.

IMPLICATIONS

Based on our empirical demonstration and our explanation of 1G and 2G techniques, we next offer several recommendations for and implications of using these techniques for research.

First, before performing data validation (ideally as part of design), researchers need to properly specify their models, determine which constructs are reflective and which are formative, and then use appropriate validation techniques specific to reflective and formative indicators. If interaction terms exist, a summated indicator (SI) approach should be used only with reflective indicators (because they are interchangeable) and when there is little to no measurement error. Meanwhile, the PI approach should be used with mixed models and for data that have measurement error. In our model, we had six reflective constructs and two formative ones. For our moderation, we utilized the PI approach because the moderation included a formative construct (communication quality).

Second, 1G regression techniques can often be used effectively to analyze path models under the following conditions: if an entire model is reflective, if a model is fairly simple (e.g., few paths and nodes), if the model is complex but measured

TABLE II
COMPARISON OF ANALYSIS TECHNIQUES FOR SEVERAL MODELING SCENARIOS

Modeling Scenario	First generation	CB-SEM	PLS
Estimate the effect of one observed* IV on one observed DV	Simplest	Also works but is overkill for such simple tests	Also works but is overkill for such simple tests
Estimate the effect of many observed IVs and one observed DV	Simplest	Also works but is overkill for such simple tests	Also works but is overkill for such simple tests
Repeated measures	Simplest	Difficult but possible	Not yet designed well for this
Estimate effects on multiple observed DVs	Also works (multivariate general linear model)	Works (also provides model fit statistics)	Also works (but no fit statistics)
Estimate effects between latent variables	Will not work	Works (also provides model fit statistics)	Also works (but no fit statistics)
Estimate indirect and total effects (mediation) through a chain of effects	Will not work	Works through bootstrapping	Works through bootstrapping
Compare multiple effects across multiple groups	Will not work	Simplest	Not yet designed well for this
Create simple observed interaction effects to predict single observed DV	Works	Also works but is overkill for such simple tests	Also works but is overkill for such simple tests
Create complex latent interaction effects	Will not work	Difficult but possible	Simplest
Model includes formative latent variables	Will not work	Not designed for this; difficult, but can use MIMIC models	Simplest
Hierarchical models (second- or third-order latent variables)	Will not work	Works well for reflective variables; possible for formative with MIMIC models	Works for reflective and formative variables
Inclusion of covariates or control variables	Works	Works; best for medium size models and when model fit is desired	Works; best for medium-size and complex models when model fit not desired
Modeling large, complex models with high number of latent constructs	Will not work	Typically will not work because of issues in creating Cartesian products; best for medium to small models	Will work but must adhere to sampling size requirements or results will be suspect

*such as, not latent

with little to no error, if there are large interaction terms that are reflective and measured with little error (in which case the SI approach should be suitable), if polynomial relationships exist [41], if there are nonlinear relationships [41], or if carefully validated data-transformation techniques need to be used to deal with highly influential outliers, heteroscedasticity, and the like. [41]. In fact, Marcoulides et al. [52] point out that when all constructs are considered reflective, any comparison between multiple regression and 2G techniques is trivial. We did not rely on 1G techniques because our model was complex and included formative constructs.

Third, 2G techniques should be used to analyze path models under the following conditions: if the entire model is formative (either CB-SEM or PLS can theoretically be used, but the analysis for CB-SEM is challenging and still has some issues [90]), if a model is mixed between formative and reflective indicators (in which case PLS should be used), if a model has small interaction terms measured with error (using the PI approach), or if a model has large interaction terms (many indicators) or terms with substantial error (in which case PLS is most appropriate). Table II offers our recommendations of which statistical approach to select given one of a few common causal-modeling scenarios. Our model included a very large interaction as well as some formative constructs; thus, PLS was the most appropriate analysis approach.

Fourth, to improve analysis with PLS (this is also applicable to 1G techniques and CB-SEM), Marcoulides and Saunders [64] suggest focusing on strong theoretical models, emphasizing data screening (including tests of normality and missing data), examining the psychometric properties of the variables in a model, using carefully measured variables, examining the "magnitude of the relationships and effects between the variables being considered in the proposed model" [64, p. vii], examining "the magnitude of the standard errors of the estimates considered in the proposed model and construct confidence intervals for the population parameters of interest" [64, p. vii], and reporting statistical power (for an example of a power analysis, see [91]). We have illustrated and described these in the context of our model, either in the demonstration above or in the online appendix, available online as downloadable supplementary material at http://ieeexplore.ieee.org.

Fifth, researchers must also be careful with model parameterizations (e.g., correlation matrix versus covariance matrix), as many researchers have mistakenly claimed a difference in models due to PLS when, in fact, the difference was due to the effects produced by the method of model parameterization [52]. CB-SEM is based on the covariance matrix, whereas PLS is based on the variances within blocks of variables. Due to the complexity of our final model (159 variables), the covariance matrix would have been unwieldy. This provided another reason to rely on PLS for our model.

As a final consideration, it is important to only use standard SEM tools for linear relationships, as most current tools for PLS and CB-SEM are limited by their ability to only probe linear relationships. To explore curvilinear relationships, special tools and different interpretation approaches must be taken. The discussion of curvilinear relationships and tools to assess them is beyond the scope of this tutorial. For a good guide to such approaches, we recommend the work of Ned Kock [92]–[94], developer of WarpPLS. This tutorial—intended as a beginner's guide to SEM issues and to PLS analyses in particular—also does not address some of the numerous other more advanced issues related to SEM, such as, but not limited to, lateral collinearity (see [95]), heteroscedasticity, multivariate outliers, model fit, factor rotation, and measurement invariance. To include all possible analyses in SEM would require a textbook of space. We have provided in the main text, and supplemented by the appendix, what we suggest will be sufficient instruction simply to analyze a fairly complex model in PLS.

In closing, it is our contention that researchers who appropriately apply the concepts outlined in this tutorial, including the basics of PLS analyses, will be more competent, effective, and competitive in their causal research. Incompatible approaches to statistical analysis can undermine a discipline's maturity. When researchers test a theory with incompatible statistical techniques (or use an appropriate technique in an incorrect way), they are more likely to be misled by their findings, even if their data are solid. This could taint theoretically sound arguments and advance theoretically unsound arguments. Over the long term, this could undermine the reputation of a field and even undermine general causal inquiry.

Although behavioral and communication researchers with a solid statistical understanding will be better able to probe current research questions, the development of sound theory and the derivation of sound hypotheses must guide the drawing of causal inferences from any statistical technique. By carefully considering the theoretical implications of their statistical choices, causal researchers can eliminate cargo cult research in communication and behavioral literature.

ACKNOWLEDGMENTS

The authors appreciate extensive feedback and developmental reviews of this work by Detmar Straub and Robert O. Briggs. The authors also appreciate feedback from Wynne W. Chin, Carol Saunders, George A. Marcoulides, David Wilson, Jordan Barlow, Laura Rawlins, and Joseph Sowa.

REFERENCES

[1] J. Wilson, "Responsible authorship and peer review," *Sci. Eng. Ethics*, vol. 8, pp. 155–174, 2002.
[2] M. Haenlein and A. M. Kaplan, "A beginner's guide to partial least squares analysis," *Understanding Stat.*, vol. 3, pp. 283–297, 2004.
[3] A. Diamantopoulos, "Modelling with LISREL: A guide for the uninitiated," *J. Market. Manage.*, vol. 10, pp. 105–136, 1994.
[4] R. P. Bagozzi and Y. Yi, "On the use of structural equation models in experimental designs," *J. Market. Res.*, vol. 26, pp. 271–284, 1989.
[5] D. W. Barclay, C. A. Higgins, and R. Thompson, "The partial least squares approach to causal modeling: Personal computer adoption and use as an illustration," *Technol. Studies*, vol. 2, pp. 285–324, 1995.
[6] C. Fornell, P. Lorange, and J. Roos, "The cooperative venture formation process: A latent variable structural modeling approach," *Manage. Sci.*, vol. 36, pp. 1246–1255, 1990.
[7] C. Fornell and Y. Yi, "Assumptions of the two-step approach to latent variable modeling," *Sociol. Meth. Res.*, vol. 20, pp. 291–320, 1992.
[8] P. B. Lowry, S. Humphreys, J. Malwitz, and J. C. Nix, "A scientometric study of the perceived quality of business and technical communication journals," *IEEE Trans. Prof. Commun.*, vol. 50, no. 4, pp. 352–378, Dec. 2007.
[9] D. S. Staples, J. S. Hulland, and C. A. Higgins, A self-efficacy theory explanation for the management of remote workers in virtual organizations, *J. Comput.-Mediated Commun.,* vol. 3.
[10] A. Lawson-Body and M. Limayem, The impact of customer relationship management on customer loyalty: The moderating role of web site characteristics, *J. Comput.-Mediated Commun.,* vol. 9.
[11] Y. Kang and S. Kim, "Understanding user resistance to participation in multihop communications," *J. Comput.-Mediated Commun.*, vol. 14, pp. 328–351, 2009.
[12] L. Goel, S. Prokopec, and I. Junglas, "Coram Populo—In the presence of people: The effect of others in virtual worlds," *J. Comput.-Mediated Commun.*, vol. 18, pp. 265–282, 2013.
[13] N. Kock, R. Parente, and J. Verville, "Can Hofstede's model explain national differences in perceived information overload? A look at data from the US and New Zealand," *IEEE Trans. Prof. Commun.*, vol. 51, no. 1, pp. 33–49, Mar. 2008.
[14] N. Kock, R. Chatelain-Jardon, and J. Carmona, "An experimental study of imulated web-based threats and their impact on knowledge communication effectiveness," *IEEE Trans. Prof. Commun.*, vol. 51, no. 2, pp. 183–197, Jun. 2008.
[15] H. Yujong and D. J. Kim, "Understanding affective commitment, collectivist culture, and social influence in relation to knowledge sharing in technology mediated learning," *IEEE Trans. Prof. Commun.*, vol. 50, no. 3, pp. 232–248, Sep. 2007.
[16] W. Jingguo, T. Herath, C. Rui, A. Vishwanath, and H. R. Rao, "Phishing susceptibility: An investigation into the processing of a targeted spear phishing email," *IEEE Trans. Prof. Commun.*, vol. 55, no. 4, pp. 345–362, Dec. 2012.
[17] C. Sangmi, S. Bagchi-Sen, C. Morrell, H. R. Rao, and S. J. Upadhyaya, "Internet and online information privacy: An exploratory study of preteens and early teens," *IEEE Trans. Prof. Commun.*, vol. 52, no. 2, pp. 167–182, Jun. 2009.
[18] C. Scott and S. Sarker, "Examining the role of the communication channel interface and recipient characteristics on knowledge internalization: A pragmatist view," *IEEE Trans. Prof. Commun.*, vol. 53, no. 2, pp. 116–131, Jun. 2010.
[19] A. M. Croteau, L. Dyer, and M. Miguel, "Employee reactions to paper and electronic surveys: An experimental comparison," *IEEE Trans. Prof. Commun.*, vol. 53, no. 3, pp. 249–259, Sep. 2010.
[20] Y. Jie, J. Zhenhui, and H. C. Chan, "The influence of sociotechnological mechanisms on individual motivation toward knowledge contribution in problem-solving virtual communities," *IEEE Trans. Prof. Commun.*, vol. 54, no. 2, pp. 152–167, Jun. 2011.
[21] O. Turetken, A. Jain, B. Quesenberry, and O. Ngwenyama, "An empirical investigation of the impact of individual and work characteristics on telecommuting success," *IEEE Trans. Prof. Commun.*, vol. 54, no. 1, pp. 56–67, Mar. 2011.
[22] J. Lin, C. H. Chuan, and X. Lingling, "A tale of four functions in a multifunctional device: Extending implementation intention theory," *IEEE Trans. Prof. Commun.*, vol. 55, no. 1, pp. 36–54, Mar. 2012.
[23] J. Mayfield and M. Mayfield, "The relationship between leader motivating language and self-efficacy: A partial least squares model analysis," *J. Bus. Commun.*, vol. 49, pp. 357–376, 2012.
[24] R. Privman, S. R. Hiltz, and W. Yiran, "In-group (us) versus out-group (them) dynamics and effectiveness in partially distributed teams," *IEEE Trans. Prof. Commun.*, vol. 56, no. 1, pp. 33–49, Mar. 2013.
[25] C. M. Fuller, K. Marett, and D. P. Twitchell, "An examination of deception in virtual Teams: Effects of deception on task performance, mutuality, and trust," *IEEE Trans. Prof. Commun.*, vol. 55, no. 1, pp. 20–35, Mar. 2012.

[26] G. Giordano and J. F. George, "The effects of task complexity and group member experience on computer-mediated groups facing deception," *IEEE Trans. Prof. Commun.*, vol. 56, no. 3, pp. 210–225, Sep. 2013.
[27] E. A. Gomez and N. Elliot, "Measuring mobile ICT literacy: Short-message performance assessment in emergency response settings," *IEEE Trans. Prof. Commun.*, vol. 56, no. 1, pp. 16–32, Mar. 2013.
[28] P. B. Lowry, T. L. Roberts, J. Nicholas, and C. Romano, "What signal is your inspection team sending to each other? Using a shared collaborative interface to improve shared cognition and implicit coordination in error-detection teams," *Int. J. Human-Comput. Studies*, vol. 71, pp. 455–474, 2013.
[29] T. L. Roberts, P. B. Lowry, and P. D. Sweeney, "An evaluation of the impact of social presence through group size and the use of collaborative software on group member "voice" in face-to-face and computer-mediated task groups," *IEEE Trans. Prof. Commun.*, vol. 49, no. 1, pp. 28–43, Mar. 2006.
[30] P. B. Lowry, J. F. N. , Jr., A. Curtis, and M. R. Lowry, "The impact of process structure on novice, virtual collaborative writing teams," *IEEE Trans. Prof. Commun.*, vol. 48, no. 4, pp. 341–364, Dec. 2005.
[31] P. B. Lowry, T. L. Roberts, N. C. Romano, Jr., P. Cheney, and R. T. Hightower, "The impact of group size and social presence on small-group communication: Does computer-mediated communication make a difference?," *Small Group Res.*, vol. 37, pp. 631–661, 2006.
[32] B. Hendriks, F. van Meurs, H. Korzilius, R. Le Pair, and S. Le Blanc-Damen, "Style congruency and persuasion: A cross-cultural study into the influence of differences in style dimensions on the persuasiveness of business newsletters in Great Britain and the Netherlands," *IEEE Trans. Prof. Commun.*, vol. 55, no. 2, pp. 122–141, Jun. 2012.
[33] F. Xiang and T. M. Rajkumar, "The role of national culture and multimedia on first impression bias reduction: An experimental study in US and China," *IEEE Trans. Prof. Commun.*, vol. 56, no. 4, pp. 354–371, Dec. 2013.
[34] D. Zhang, P. B. Lowry, L. Zhou, and X. Fu, "The impact of individualism-collectivism, social presence, and group diversity on group decision making under majority influence," *J. Manage. Inf. Syst.*, vol. 23, pp. 53–80, 2007.
[35] P. B. Lowry, J. Cao, and A. Everard, "Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: The case of instant messaging in two cultures," *J. Manage. Inf. Syst.*, vol. 27, pp. 163–200, 2011.
[36] L. Zoonky and L. Younghwa, "Emailing the boss: Cultural implications of media choice," *IEEE Trans. Prof. Commun.*, vol. 52, no. 1, pp. 61–74, Mar. 2009.
[37] P. Burton, W. Yu, V. R. Prybutok, and G. Harden, "Differential effects of the volume and diversity of communication network ties on knowledge workers' performance," *IEEE Trans. Prof. Commun.*, vol. 55, no. 3, pp. 239–253, Sep. 2012.
[38] W. W. Chin, "Issues and opinion on structural equation modeling," *MIS Quart.*, vol. 22, pp. vii–xvi, 1998.
[39] B. Schneider, M. Carnoy, J. Kilpatrick, W. H. Schmidt, and R. J. Shavelson, "Estimating causal effects using experimental and observational designs," The Governing Board of the American Educational Research Association Grants Program, Washington DC, 2007.
[40] A. F. Hayes, *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, USA: Guilford, 2013.
[41] D. Gefen, D. W. Straub, and M. Boudreau, "Structural equation modeling and regression: Guidelines for research practice," *Commun. AIS*, vol. 4, pp. 1–78, 2000.
[42] W. W. Chin, "The partial least squares approach to structural equation modeling," in *Modern Methods for Business Research*, G. A. Marcoulides, Ed. Mahwah, NJ, USA: Erlbaum, 1998, pp. 295–336.
[43] D. Gefen and D. W. Straub, "A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example," *Commun. AIS*, vol. 16, pp. 91–109, 2005.
[44] W. W. Chin, "The holistic approach to construct validation in IS research: Examples of the interplay between theory and measurement," in *Proc. ASAC*, Windsor, ON, Canada, 1995, pp. 33–43.
[45] W. W. Chin, B. L. Marcolin, and P. R. Newsted, "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," *Inf. Syst. Res.*, vol. 14, pp. 189–217, 2003.
[46] M.-C. Boudreau, D. Gefen, and D. W. Straub, "Validation in information systems research: A state-of-the-art assessment," *MIS Quart.*, vol. 25, pp. 1–16, 2001.
[47] F. D. Davis, "User acceptance of information technology: System characteristics, user perceptions and behavioral impacts," *Int. J. Man-Mach. Studies*, vol. 38, pp. 475–487, 1993.
[48] J. Hage and B. F. Meeker, *Social Causality*. Boston, MA, USA: Unwin Hyman, 1987.
[49] T. D. Cook and D. T. Campbell, *Quasi-Experimentation: Design and Analysis for Field Settings*. Chicago, IL, USA: Rand McNally, 1979.
[50] J.-B. Lohmoller, "Basic principles of model building," in *Theoretical Empiricism*, H. Wold, Ed. New York: Paragon House, 1989, pp. 1–25.
[51] W. W. Chin, B. L. Marcolin, and P. R. Newsted, "A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and voice mail emotion/adoption study," in *Proc. 17th Int. Conf. Inf. Syst.*, Cleveland, OH, USA, 1996, pp. 21–41.
[52] G. A. Marcoulides, W. W. Chin, and C. S. Saunders, "A critical look at partial least squares modeling," *MIS Quart.*, vol. 33, pp. 171–175, 2009.
[53] C. Fornell and F. L. Bookstein, "Two structural equation models: LISREL and PLS applied to consumer exit-voice theory," *J. Market. Res.*, vol. 19, pp. 440–452, 1982.

[54] J. H. Steiger and P. H. Schönemann, "A history of factor indeterminacy," in *Theory Construction and Data Analysis in the Behavioral Sciences*, S. Shye, Ed. San Francisco, CA, USA: Jossey-Bass, 1978.
[55] W. W. Chin and P. A. Todd, "On the use, usefulness and ease of use of structural equation modeling in MIS research: A note of caution," *MIS Quart.*, vol. 19, pp. 237–246, 1995.
[56] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a silver bullet," *J. Market. Theory Practice*, vol. 19, pp. 139–152, 2011.
[57] S. A. Mulaik, "Comments on the measurement of factorial indeterminacy," *Psychometrika*, vol. 41, pp. 249–262, 1976.
[58] H. Wold, "Soft modeling—The basic design and some extensions," in *Systems Under Indirect Observation II*, K. Joreskog and H. Wold, Eds. Amsterdam, the Netherlands: North-Holland Press, 1982, pp. 1–53.
[59] J. F. Hair, C. M. Ringle, and M. Sarstedt, "From the special issue guest editors," *J. Market. Theory Practice*, vol. 19, pp. 135–138, 2011.
[60] A. Diamantopoulos, "Viewpoint: Export performance measurement: Reflective versus formative indicators," *Int. Market. Rev.*, vol. 16, pp. 444–457, 1999.
[61] K. Bollen, "Multiple indicators: Internal consistency of no necessary relationship?'," *Qual. Quantity*, vol. 18, pp. 377–385, 1984.
[62] S. Petter, D. W. Straub, and A. Rai, "Specifying formative constructs in information systems research," *MIS Quart.*, vol. 31, pp. 623–656, 2007.
[63] C. M. Ringle, S. Wende, and S. Will. (2005). SmartPLS 2.0 (M3) Beta. [Online]. Available: http://www.smartpls.de
[64] G. A. Marcoulides and C. Saunders, "PLS: A silver bullet?," *MIS Quart.*, vol. 30, pp. 1–8, 2006.
[65] W. W. Chin and P. R. Newsted, "Structural equation modeling analysis with small samples using partial least squares," in *Statistical Strategies for Small Sample Research*, R. H. Hoyle, Ed. Thousand Oaks, CA, USA: Sage, 1999, pp. 307–341.
[66] D. Goodhue, W. Lewis, and R. Thompson, "PLS, small sample size, and statistical power in MIS research," presented at the 39th Hawaii Int. Conf. Syst. Sci., Kona, HI, USA, 2006.
[67] P. B. Lowry, N. C. Romano, J. L. Jenkins, and R. W. Guthrie, "The CMC interactivity model: How interactivity enhances communication quality and process satisfaction in lean-media groups," *J. Manage. Inf. Syst.*, vol. 26, pp. 159–200, 2009.
[68] J. Gaskin. (2012). SmartPLS new project, load and troubleshoot data. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[69] A. Diamantopoulos and H. M. Winklhofer, "Index construction with formative indicators: An alternative to scale development," *J. Market. Res.*, vol. 38, pp. 269–277, 2001.
[70] C. B. Jarvis, S. B. MacKenzie, and P. M. Podsakoff, "A critical review of construct indicators and measurement model misspecification in marketing and consumer research," *J. Consum. Res.*, vol. 30, pp. 199–218, 2003.
[71] W. W. Chin and A. Gopal, "Adoption intention in GSS: Relative importance of beliefs," *ACM SIGMIS Database*, vol. 26, pp. 42–64, 1995.
[72] J. Gaskin. (2012). SmartPLS formative second-order constructs. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[73] J. Gaskin. (2012). SmartPLS factor analysis. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[74] D. W. Straub, M. C. Boudreau, and D. Gefen, "Validation guidelines for IS positivist research," *Commun. AIS*, vol. 14, pp. 380–426, 2004.
[75] D. W. Straub, "Validating instruments in MIS research," *MIS Quart.*, vol. 13, pp. 146–169, 1989.
[76] G. M. Marakas, R. D. Johnson, and P. F. Clay, "The evolving nature of the computer self-efficacy construct: An empirical investigation of measurement construction, validity, reliability, and stability over time," *J. Assoc. Inf. Syst.*, vol. 8, pp. 16–46, 2007.
[77] K. D. Loch, D. W. Straub, and S. Kamel, "Diffusing the internet in the Arab world: The role of social norms and technological culturation," *IEEE Trans. Eng. Manage.*, vol. 50, no. 1, pp. 45–63, Feb. 2003.
[78] C. M. Ringle, M. Sarstedt, and D. W. Straub, "Editor's comments: A critical look at the use of PLS-SEM in MIS quarterly," *MIS Quart.*, vol. 36, pp. iii–xiv, 2012.
[79] J. Gaskin. (2011). Detecting multicollinearity in SPSS. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[80] R. P. Bagozzi, "Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations," *MIS Quart.*, vol. 35, pp. 261–292, 2011.
[81] P. M. Podsakoff, S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: A critical review of the literature and recommended remedies," *J. Appl. Psychol.*, vol. 88, pp. 879–903, 2003.
[82] P. A. Pavlou, H. Liang, and Y. Xue, "Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective," *MIS Quart,*, vol. 31, pp. 105–136, 2007.
[83] M. K. Lindell and D. J. Whitney, "Accounting for common method variance in cross-sectional research designs," *J. Appl. Psychol.*, vol. 86, pp. 114–121, 2001.
[84] S. B. MacKenzie and P. M. Podsakoff, "Common method bias in marketing: Causes, mechanisms, and procedural remedies," *J. Retail.*, vol. 88, pp. 542–555, 2012.
[85] T. A. Carte and C. J. Russell, "In pursuit of moderation: Nine common errors and their solutions," *MIS Quart.*, vol. 27, pp. 479–501, 2003.

[86] J. Gaskin. (2012). SmartPLS interaction moderation. Gaskination's Statistics [Online]. Available: http://youtube.com/Gaskination
[87] R. B. Baron and D. A. Kenny, "The moderator mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," *J. Personal. Soc. Psychol.*, vol. 51, pp. 1173–1182, 1986.
[88] J. Gaskin. (2012). SmartPLS mediation, bootstrap and Sobel test. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[89] J. Gaskin. (2013). SmartPLS: Multigroup moderation and moderated mediation. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[90] A. Diamantopoulos, "Incorporating formative measures into covariance-based structural equation models," *MIS Quart.*, vol. 35, pp. 335–358, 2011.
[91] J. Gaskin. (2013). Post-hoc power analysis in SmartPLS and AMOS. Gaskination's Statistics. [Online]. Available: http://youtube.com/Gaskination
[92] N. Kock, "Using WarpPLS in e-collaboration studies: An overview of five main analysis steps," *Int. J. e-Collab.*, vol. 6, pp. 1–11, 2010.
[93] N. Kock, "Using WarpPLS in e-collaboration studies: Mediating effects, control and second order variables, and algorithm choices," *Int. J. e-Collab.*, vol. 7, pp. 1–13, 2011.
[94] *"WarpPLS 3.0 User Manual,"* N. Kock, ScriptWarp Systems, Laredo, TX, USA, 2012.
[95] N. Kock and G. S. Lynn, "Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations," *J. Assoc. Inf. Syst.*, vol. 13, 2012.

Paul Benjamin Lowry received the Ph.D. degree in Management Information Systems from the University of Arizona, Tucson, AZ, USA. He is an associate professor of Information Systems and the Associate Director of the MBA Program at the City University of Hong Kong, Hong Kong, China. His research interests include behavioral information security, human–computer interaction, e-commerce, and scientometrics.

James Gaskin received his Ph.D. degree in Management Information Systems from Case Western Reserve University, Cleveland, OH, USA. He is an assistant professor of Information Systems at Brigham Young University, Provo, UT, USA. His research interests include organizational genetics, human-computer interaction, and research and teaching methodologies. He is well known for his YouTube video tutorials for structural equation modeling (SEM).

ONLINE APPENDIX 1: MODEL STATISTICS

***** Note to editors and reviewers:** Because of length restrictions, this appendix is NOT intended for publication in the printed version (unless portions are desired in print); instead, we intend to make it available via an online Web site supplement.

Table A1.1 provides evidence of the significance of the loading or weight of each item within each latent variable. All items on the latent variable should have a significant *t*-statistic in order to demonstrate adequate convergent validity. As shown below, our data meets this criteria.

Table A1.1 T-statistics for Convergent Validity

Construct (latent variable)	Indicator	t-statistic
Two-way communication	two1	15.52***
	two2	19.24***
	two3	19.53***
	two4	21.67***
	two5	14.10***
	two6	19.67***
Synchronicity	syn1	12.43***
	syn2	34.86***
	syn3	30.50***
	syn4	44.17***
	syn5	8.61***
Control	ctl1	25.45***
	ctl2	40.07***
	ctl3	7.09***
	ctl4	7.54***
Discussion efficiency	eff1	34.12***
	eff2	81.78***
	eff3	78.05***
Communication Openness	op1	41.19***
	op2	46.15***
	op3	49.78***
	op4	55.03***
	op5	33.74***
Status effects	stat1	4.67***
	stat2	5.91***
	stat3	6.01***
	stat4	7.03***

*$p < 0.05$, ** $p < 0.01$, ***$p < 0.001$

Table A1.2 is a matrix of loadings and cross-loadings for all reflective items in the model. The loadings of the items in this table should be greater for the latent variable to which they theoretically belong than for any other latent variable. Discriminant validity is adequate if the cross-loadings (with other latent variables) are more than the absolute value of 0.100 distant from the loading on the primary latent variable [1]. For example, in Table A1.2, the ctl1 item loads with a value of 0.823 onto the Control latent variable, but loads onto the other variable with values no greater than 0.500. The strong loading on Control indicates that ctl1 is more strongly correlated with ctl2-4 than it is with the other items in the table. Item syn5 is removed because it is also loads fairly strongly on Two-way and loads only slightly more strongly on Synchronicity. It is beyond the minimum 0.100 threshold; however, removing it in this case will actually improve discriminant validity since its low loading on Synchronicity is also bringing down the average loading for that latent variable.

Table A1.2 Loadings of the Measurement Items

	Control	Two-way	Synchronicity	Openness	Efficiency	Status
ctl1	0.823	0.492	0.411	0.282	0.203	-0.102
ctl2	0.826	0.441	0.419	0.203	0.194	-0.072
ctl3	0.567	0.318	0.279	0.221	0.211	-0.113
ctl4	0.590	0.320	0.256	0.064	0.109	-0.042
two1	0.498	0.715	0.400	0.234	0.240	-0.201
two2	0.457	0.751	0.523	0.260	0.318	-0.178
two3	0.328	0.722	0.473	0.225	0.298	-0.172
two4	0.448	0.769	0.542	0.255	0.327	-0.186
two5	0.326	0.672	0.354	0.134	0.241	-0.214
two6	0.377	0.727	0.449	0.314	0.277	-0.170
syn1	0.333	0.443	0.674	0.278	0.299	-0.122
syn2	0.407	0.481	0.833	0.285	0.380	-0.140
syn3	0.413	0.526	0.817	0.278	0.437	-0.153
syn4	0.409	0.528	0.846	0.248	0.364	-0.139
syn5	0.274	0.409	*0.583	0.235	0.257	-0.039
op1	0.295	0.377	0.338	0.871	0.577	0.230
op2	0.317	0.377	0.378	0.883	0.598	0.182
op3	0.193	0.214	0.267	0.885	0.522	0.193
op4	0.237	0.244	0.297	0.900	0.572	0.241
op5	0.173	0.228	0.243	0.845	0.535	0.231
eff1	0.220	0.324	0.385	0.572	0.864	0.156
eff2	0.245	0.387	0.420	0.596	0.935	0.153
eff3	0.222	0.356	0.460	0.578	0.930	0.163
stat1	-0.076	-0.213	-0.135	0.209	0.119	0.904
stat2	-0.088	-0.221	-0.173	0.244	0.143	0.912
stat3	-0.151	-0.245	-0.162	0.223	0.163	0.916
stat4	-0.096	-0.251	-0.101	0.204	0.217	0.851

***** Item removed to improve discriminant validity

Discriminant validity is also demonstrated by comparing the square root of the average variance extracted to the correlations with other latent variables[2]. If the diagonal values are greater than any other correlation, then this establishes adequate discriminant validity. If this threshold is not met (i.e., a correlation is stronger than the diagonal value) then the AVE is lower than the shared variances with other latent variables. This means that the model will need to be reevaluated to determine if items with either low loadings or high cross-loadings (such as syn5 in Table A1.2) can be dropped in order to increase the AVE or decrease the shared variance with another latent variable.

Table A1.3. Discriminant Validity through the Square Root of AVE **(on diagonal)**

	(1)	(2)	(3)	(4)	(5)	(6)
Control (1)	**(0.712)**					
Two-way comm (2)	0.547	**(0.727)**				
Synchronicity (3)	0.469	0.605	**(0.809)**			
Comm Open (4)	0.271	0.324	0.337	**(0.877)**		
Discuss eff (5)	0.255	0.391	0.455	0.640	**(0.910)**	
Status effects (6)	-0.117	-0.260	-0.168	0.245	0.179	**(0.896)**

Table A1.4 provides the composite reliability of each reflective latent variable. Reliability is a measure of internal consistency required in reflective (internally correlated) latent variables. No such expectation is placed upon formative latent variables. To establish reliability, the composite reliability measures should be greater than 0.700 [3].

Table A1.4 Composite Reliability

Construct (latent variable)	Composite reliability
Control	0.800
Two-way communication	0.870
Synchronicity	0.883
Communication Openness	0.943
Discussion efficiency	0.828
Status effects	0.942

Table A1.5 MTMM Analysis Table

	V1	V2	V3	V4	V5	V6	V7	V8	TKD	V9	V10	V11	V12	V13
V1														
V2	0.725													
V3	0.620	0.615												
V4	-0.503	-0.538	-0.715											
V5	0.337	0.298	0.471	-0.363										
V6	0.284	0.287	0.429	-0.344	0.345									
V7	0.267	0.263	0.427	-0.404	0.224	0.352								
V8	0.211	0.136	0.293	-0.247	0.509	0.268	0.299							
TKD	0.644	0.590	0.758	-0.590	0.765	0.508	0.510	0.751						
V9	0.306	0.257	0.437	-0.352	0.564	0.320	0.264	0.633	0.674					
V10	0.496	0.498	0.660	-0.511	0.480	0.377	0.439	0.369	0.676	0.499				
V11	0.476	0.445	0.644	-0.513	0.449	0.416	0.433	0.368	0.654	0.472	0.762			
V12	0.208	0.177	0.318	-0.277	0.419	0.200	0.274	0.437	0.486	0.518	0.469	0.425		
V13	0.324	0.250	0.448	-0.321	0.564	0.320	0.288	0.549	0.647	0.667	0.487	0.524	0.630	
SATP	0.457	0.409	0.629	-0.487	0.638	0.413	0.411	0.612	0.800	0.843	0.803	0.752	0.684	0.853

Table A1.6 offers the descriptive statistics (means and standard deviations) for, and correlations among, all first order latent variables in the model.

Table A1.6 Measurement Model Statistics (n = 346)

Construct	μ	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control (1)	4.97	1.14							
Two-way communication (2)	5.13	1.20	0.547						
Synchronicity (3)	5.12	1.22	0.469	0.605					
Task discussion effectiveness (4)	4.97	1.18	0.338	0.508	0.402				
Communication openness (5)	4.50	1.72	0.271	0.324	0.337	0.418			
Discussion efficiency (6)	4.78	1.74	0.255	0.391	0.455	0.486	0.640		
Status effects (7)	2.21	1.59	-0.117	-0.260	-0.168	-0.065	0.245	0.179	
PS (8)	5.18	1.34	0.364	0.490	0.429	0.733	0.519	0.582	-0.062

Table A1.7 offers an evaluation of the structural model, including path coefficients (regression weights) and *t*-values. Strong and significant paths in the expected direction indicate support for that hypothesized path.

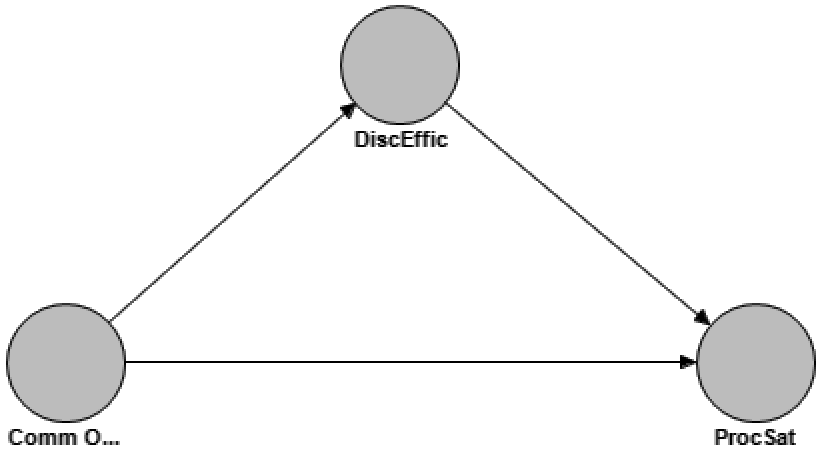
Table A1.7 Summary of Path Coefficients and Significance Levels

Hypotheses and corresponding paths	Expected sign	Path coefficient	*t*-value ($df = 345$)
Interactivity → communication quality (CQ)	+	0.697	13.63***
Communication quality → process satisfaction (PS)	+	0.699	15.29***
Status effects negatively moderates the relationship between CQ and PS	-	(-0.180)	2.85**
Two-way communication is a first-order factor of interactivity	+	0.495	44.18***
Synchronicity is a first-order factor of interactivity	+	0.367	34.90***
Control is a first-order factor of interactivity	+	0.321	28.38***
Discussion efficiency is a first-order factor of CQ	+	0.289	26.50***
Communication openness is a first-order factor of CQ	+	0.465	29.65***
Task discussion effectiveness is a first-order factor of CQ	+	0.421	49.30***

* indicates significant paths: *$p < 0.05$, ** $p < 0.01$, ***$p < 0.001$

ONLINE APPENDIX 2: SUPPLEMENTAL PLS ANALYSES

In this appendix, we offer two additional illustrative PLS analyses that were not relevant to our theoretical model in the main text but can be readily in other contexts. To reduce complexity and to offer a more parsimonious illustration of these two additional analyses, we simplify the model for this appendix. The model we will be testing across groups as well as for moderated mediation includes only three latent constructs: Communication Openness, Discussion Efficiency, and Process Satisfaction as shown in the figure below. The categorical moderator used in these examples is an artificial gender variable created solely for illustrative purposes. Thus, no theoretical claims should be made based on these results, which arefor illustration only.



BETWEEN GROUP COMPARISONS

Moderating by group membership (such as gender, religion, nationality, and so forth.) is a common modeling need. However, existing tools for conducting PLS analyses are not well-designed for such tests. We therefore offer here an illustrative example of how to conduct this kind of analysis. A video demonstration of multi-group moderation, as well as moderated mediation in SmartPLS, is available here at Gaskination [4].

1. Split the data into two datasets based on the values of the moderator. In this example, we are using gender as the moderating variable. Therefore, we create two datasets: one for male (n=188) and one for female (n=155).
2. Load both datasets into SmartPLS and then run a bootstrap analysis on each dataset using the same model. To switch between datasets, right-click the dataset and select "Use Data for Calculation". Open the default report after running the bootstrap for each gender.
3. Using the formula (shown below) provided on Wynne Chin's PLS FAQ website (http://disc-nt.cba.uh.edu/chin/plsfaq.htm), calculate the *t*-statistic for the difference between the effects. The formula requires the sample size of each group, as well as the regression weights and the standard errors for the path being tested. This fairly complex formula has been converted into an Excel function by James Gaskin (available here: http://www.kolobkreations.com/Stats%20Tools%20Package.xlsm) that also converts the *t*-statistic into a two-tailed probability value.

$$t = \frac{Path_{sample\_1} - Path_{sample\_2}}{\left[\sqrt{\frac{(m-1)^2}{(m+n-2)} * S.E.^2_{sample1} + \frac{(n-1)^2}{(m+n-2)} * S.E.^2_{sample2}}\right] * \left[\sqrt{\frac{1}{m} + \frac{1}{n}}\right]}$$

The results of testing our simple model (using 500 resamples) are shown in the table below. The results indicate that the effect of communication openness on discussion efficiency is significantly stronger for males. The effect between communication openness and process satisfaction is also significantly stronger for males. However, the effect between discussion efficiency and process satisfaction is significantly stronger for females.

	Comm Open → Disc Effic		**Comm Open → Proc Sat**		**Disc Effic → Proc Sat**	
	Female	**Male**	Female	**Male**	**Female**	Male
Sample Size	158	188	158	188	158	188
Regression Weight	0.3167	**0.5865**	0.1576	**0.3400**	**0.4603**	0.2410
Standard Error (S.E.)	0.0421	0.0541	0.0436	0.0580	0.0404	0.0672
t-statistic	3.8367		2.4440		2.6773	
p-value (2-tailed)	0.0001		0.0150		0.0078	

MODERATED MEDIATION

Extending the above analysis to test for moderated mediation is fairly simple. We can actually use the same approach, but instead of looking at the regression weight and standard error for the direct effect, we will use the ones for the total effect. In this analysis, we will test whether the effect of communication openness on process satisfaction, mediated by discussion efficiency, is moderated by gender such that the total effect of communication openness on process satisfaction is significantly different for males and females. Using the same steps as above (split data, bootstrap, open default report), we then input the regression weight and standard error for the *total effects* into Chin's formula. The result for our model is shown below. The results indicate that the mediated effect is stronger for males than for females, and this difference is significant at $p<0.05$.

	Comm Open → Disc Effic → Proc Sat	
	Female	**Male**
Sample Size	158	188
Regression Weight	0.3020	0.4824
Standard Error (S.E.)	0.0409	0.0684
t-statistic	2.166	
p-value (2-tailed)	0.031	

APPENDIX 3. MEASURES

Table A1: Measures Used in this Research		
Latent variable (type)	Items	Measure notes
Interactivity (second-order)	Subconstruct: control (reflective): Ctnl1: I felt that I had a great deal of control over my communication in this group. Ctnl2: While I was involved in this group, I could choose freely what I wanted to hear/read and say/contribute. * Ctnl3: While involved in this group, I had absolutely no control over my communication. Ctnl4: While involved in this group, my actions determined the kind of experiences I had. Subconstruct: two-way communication (reflective): Two1: The facilitator effectively gathered group members' feedback. Two2: The group environment facilitated two-way communication between group members and the facilitator. * Two3: It was difficult to offer feedback to the facilitator. Two4: The facilitator made me feel he/she wanted to listen to the group members. * Two5: The facilitator did not at all encourage group members to communicate. Two6: The group environment gave group members the opportunity to communicate. Subconstruct: synchronicity (reflective): Synch1: The facilitator processed my input very quickly. Synch 2: Getting information from the facilitator was very fast. Synch 3: In the group environment I was able to obtain the information I wanted without any delay. Synch4: When I communicated with the facilitator, I felt I received instantaneous information.	This version directly from [5]; they adapted original measures from [6] to make them consistently in past tense and more general to a group interaction (not a Web site interaction).
Discussion efficiency	Eff1: To what extent would you agree that this group interaction was result oriented?	This version directly from [5]; original from [7].

(reflective)	Eff2: The time spent in the group interaction was efficiently used. Eff3: Issues raised in the group interaction were discussed thoroughly.	
Task discussion effectiveness (formative)	*Taskd1. The discussions, were ineffective *Taskd2. The context of the discussions was carelessly developed Taskd3. Issues were examined effectively *Taskd4. Participation in the discussions was unevenly distributed *Taskd5. Ideas in the discussions were uncritically examined Taskd6. The amount of information exchanged was sufficient	This version directly from [5]. Instrument application: use as is to measure communication task discussion effectiveness on the group level for the posttest. Original from [8]; used all original items except 1 and 3 as these overlap with discussion quality. Changed original anchors to the degree of agreement on 7-point Likert-like scale.
Process satisfaction (formative)	Satp1: Our group discussion process was efficient. * Satp2: Our group discussion process was uncoordinated. * Satp3: Our group discussion process was unfair. Satp4: Our group discussion process was understandable. Satp5: Our group discussion process was satisfying.	This version directly from [5]; original from [9]. Re-anchored on a seven-point scale from five-point scale. Original anchors were inefficient / efficient, uncoordinated / coordinated, unfair / fair, confusing / understandable, dissatisfying / satisfying; changed to only using first part of anchor with respondent indicating how strongly they agreed or disagreed (made items 1, 4, 5 positive). For classroom use “decision-making” was termed as “discussion process.”
Openness (reflective)	Open1: It was easy to communicate openly to all members of this group. Open2: Communication in this group was very open. Open3: When people communicated to each other in this group, there was a great deal of understanding. Open4: It was easy to ask advice from any member of this group. *Open5: We needed to adapt our style of communication to effectively communicate.	This version directly from [5]; adapted from [10]; changed to past tense.
Status Effects (reflective)	Stat1: Some group members tried to intimidate others, e.g. by talking loudly, using aggressive gestures, making threats, etc. Stat2: Some group members tried to use their influence, status, or power so as to force issues on the other group members. Stat3: I felt inhibited from participating in the interaction because of the behavior of other group members. Stat4: I experienced pressure, either to conform to a particular viewpoint or to not contradict others.	This version directly from [5]; original from [7]. Expended from a five-point scale to a point-point scale (strongly agree to strongly disagree).

* reverse coded

All items use a 1-to-7-point Likert-type scale anchored on “strongly disagree . . .strongly agree.”

BIBLIOGRAPHY FOR ONLINE APPENDICES

[1] D. Gefen and D. W. Straub, "A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example," *Communications of the AIS,* vol. 16, pp. 91-109, 2005.

[2] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research,* vol. 18, pp. 39-50, 1981.

[3] W. W. Chin, "The partial least squares approach to structural equation modeling," in *Modern Methods for Business Research*, G. A. Marcoulides, Ed. Mahwah: Lawrence Erlbaum Associates, 1998, pp. 295-336.

[4] J. Gaskin, "SmartPLS: Multigroup moderation and moderated mediation " in *Gaskination's Statistics*, 2013.

[5] P. B. Lowry, N. C. Romano, J. L. Jenkins, and R. W. Guthrie, "The CMC interactivity model: How interactivity enhances communication quality and process satisfaction in lean-media groups," *Journal of Management Information Systems,* vol. 26, pp. 159-200, 2009.

[6] Y. Liu, "Developing a scale to measure the interactivity of websites," *Journal of Advertising Research,* vol. 43, pp. 207-218, 2003.

[7] R. M. Davison, "An Instrument for Measuring Meeting Success," *Information & Management,* vol. 32, pp. 163-176, 1997.

[8] J. K. Burgoon, M. Burgoon, K. Broneck, E. Alvaro, and J. F. Nunamaker Jr., "Effects of synchronicity and proximity on group communication," in *Annual Meeting of the National Communication Association*, New Orleans, Louisiana, USA, 2002.

[9] B. C. Y. Tan, K.-K. Wei, and J.-E. Lee-Partridge, "Effects of Facilitation and Leadership on Meeting Outcomes in a Group Support System Environment," *European Journal of Information Systems,* vol. 8, pp. 232-246, 1999.

[10] K. H. Roberts and C. O'Reilly, "Measuring organizational communication," *Journal of Applied Psychology,* vol. 59, pp. 321-326, 1974.