

INVESTIGATING TWO CONTRADICTORY VIEWS OF FORMATIVE MEASUREMENT IN INFORMATION SYSTEMS RESEARCH¹

By: **Gimun Kim**
College of Business Administration
Konyang University
Nonsan
KOREA
gmkim@konyang.ac.kr

Bongsik Shin
Information and Decision Systems
San Diego State University
San Diego, CA 92182
U.S.A.
bshin@mail.sdsu.edu

Varun Grover
Department of Management
Clemson University
101 Sirrine Hall
Clemson, SC 29634-1305
vgrover@clemson.edu

Abstract

The use of formative measurement in the field of Information Systems has increased, arguably due to statistical tools (e.g., PLS) that can test such models. However, in the literature, there exist two contradictory views on the potential deficiency of formative measurement. While opponents who are critical of formative measurement argue that there are native weaknesses of the formative approach in model estimation, propo-

nents who are in favor of using formative measurement counter that opponents' research methods in measurement model specification are flawed. The goal of this work is to empirically test these opposing views on whether the alleged estimation instability of formative measurement is due to measurement model misspecification or simply the shortcoming of formative measurement. To assess the integrity of arguments of both parties, we adopt a research design in which four different cases are tested in terms of interpretational confounding and external consistency. We find that regardless of whether there is a specification issue, formative measures can lead to misleading outcomes. Based on the results, we offer guidelines that researchers may adopt in planning and executing data analysis with structural equation modeling. Given that the use of formative measurement is at a critical juncture in the IS field, we believe that the guidelines in this research note are important to promote appropriate use of the approach rather than relegate it to a bandwagon effect.

Keywords: Formative measurement, formative indicators, measurement models, measurement instability, external consistency, interpretational confounding, information systems measures

Introduction

Structural equation modeling (SEM) has become the preferred data analysis tool for empirical research in Information Systems. Recently, the use of formative measurement in the IS field has increased, arguably due to the statistical tools (e.g., PLS) that can test such models. For example, the online search of the EBSCO database using the key words of “formative construct” and “formative measure” resulted in 17

¹Joe Valacich was the accepting senior editor for this paper. Traci Carte served as the associate editor.

out of 1,487 *MIS Quarterly* articles and all but one were published after 2001. The same search in *Information Systems Research* had seven publications out of 410 articles, and all but one was published after 2003. The acceptance of formative measurement by the IS community has been generally favorable and the choice of whether to use formative or reflective indicators has largely been left to the researchers' subjective discretion. Given the growing use of formative measurement models in IS research, the IS community needs to take stock of theoretical and empirical discussions on the utilization of formative measurement in order to guide its appropriate use at this critical juncture.

Most studies conducted in non-IS disciplines focus on promoting and improving formative measurement as a viable alternative (or supplement) to reflective measurement. Very recently, serious debates began to emerge between opponents and proponents of formative measurement. A few recent studies explored threats formative measurement might pose to the reliability of its estimation and argued that there is a native instability in the estimation of formative measurement (e.g., Howell et al. 2007; Wilcox et al. 2008). In response, a counterargument, Bollen (2007) highlighted that there were flaws in the research method used in those studies, notably misspecification in measurement model tested on the basis of theoretically disconnected data. So, is the issue with formative measurement a theoretical one—or is it empirical? The implications for the IS field are quite distinct for each of these alternatives. Emphasizing the need for more empirical studies, Diamantopoulos et al. (2008) acknowledged that “literature has only recently started to pay serious attention to formative measurement models and empirical applications are still rare.” The current situation motivated our research.

The goal of this work is to empirically test the competing views on whether the alleged estimation instability of formative measurement is due to the inherent shortcomings of formative measurement models or not. To assess the integrity of two competing views in an objective manner, we constructed a research design in which four different test cases are derived from the cross-section of *measurement model specification types* (formatively specified versus reflectively specified model) and *theorized relationships* (formatively versus reflectively theorized relationships between a construct and its measures). Then, the stability of measurement model pertaining to each test case was tested with real survey data. The stability of measurement model is assessed from two different perspectives: (1) interpretational confounding—the reliability of a measurement model estimation given the change of endogenous variables, and (2) external consistency—the effect of weakened consistency between measures of a formatively designed construct and dependent variables on overall model fit.

Overall, the analysis of test cases indicate that formative measurement itself could entail risks to the integrity of empirical findings, *regardless of whether there exists measurement model misspecification*. Unlike the reflectively designed measurement model that can be estimated independently, the formatively designed one relies on other dependent variables for the estimation of weights between the construct and its indicators. This inability to act independently during model estimation and validity testing seems to make the formative measurement model vulnerable to estimation corruption. From the findings, we offer recommendations that IS researchers can use as practical guidelines in designing their research models and executing data analysis based on the formative paradigm.

We start with a brief mathematical illustration about the difference between reflective and formative measurement models (for more discussion of formative measurement, refer to Diamantopoulos et al. 2008; Petter et al. 2007) and then discuss the contradicting views of the research community on the estimation reliability of formative measurement associated with interpretational confounding and external consistency.

Formative Measurement

The measurement model explains the relationship between a latent concept (variable) and its corresponding indicators (Edwards and Bagozzi 2000), and, depending on the causal direction, it becomes either a reflective (from a latent construct to its indicators) model or a formative (from indicators to their latent construct) model. The reflective measurement model bases itself on the classical test theory (Bollen and Lennox 1991) in which measured items represent the effects of a latent concept (η) and therefore the latent concept (η) constitutes a common cause of measured indicators (x_i). Accordingly, each measured indicator (x_i) becomes a function of the latent variable (η) and the concomitant measurement error (see Formula 1) in which x_i becomes a dependent variable that is explained by the latent construct (η). Here, x_i is the i^{th} indicator of the latent construct η , λ_i is called loading, and ε_i represents the measurement error of x_i .

$$x_i = \lambda_i \eta + \varepsilon_i \quad (\text{Formula 1})$$

in which $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$ and $\text{cov}(\eta, \varepsilon_i) = 0$ for all i .

Meanwhile, the relationship between measures and the formatively designed construct is shown in formula 2, in which γ_i represents the contribution of x_i indicator to the latent construct (η) and ξ becomes construct-level error (or disturbance)

term. Formula 2 is a multiple regression in which the latent variable as a dependent variable is determined by the indicators as explanatory variables.

$$\eta = \sum_{i=1}^n \gamma_i x_i + \xi \quad (\text{Formula 2})$$

in which $\text{cov}(x_i, \xi) = 0$ for all i .

Theoretical Issues in Debate

A number of studies have been conducted on the theoretical aspects of formative measurement, mostly in non-IS disciplines. Most studies have focused on promoting and improving formative measurement as a viable alternative (or supplement) to reflective measurement (Diamantopoulos et al. 2008). Special attention has been placed on the subject of misspecifying formative indicators as reflective (e.g., Diamantopoulos and Sigauw 2006; Jarvis et al. 2003; MacKenzie et al. 2005). In the IS-field, Petter et al. (2007) conducted a similar study on that issue. Serious disputes, however, began to emerge in 2007 and 2008 on the viability of formative indicators for structural equation modeling (e.g., Bagozzi 2007; Bollen 2007; Howell et al. 2007; Wilcox et al. 2008). Especially, proponents who are in favor of using formative measurement and opponents who are critical of it have differing explanations on the estimation reliability of the formative measurement model. Among the central issues opponents raise are on interpretational confounding and external consistency.

Interpretational confounding “occurs as the assignment of empirical meaning to an unobserved variable which is other than the meaning assigned to it by an individual *a priori* to estimating unknown parameters” (Burt 1976, p. 4). Researchers agree that interpretational confounding is a potential problem in both effect (reflective) and causal (formative) indicators, and that if the coefficients linking indicators and the latent variable significantly change depending on the endogenous variables in a model, then this is an evidence of interpretational confounding (Bollen 2007; Howell et al. 2007). Interpretational confounding, therefore, represents the situation in which the empirically observed meaning between a latent variable and its measures differs from the nominal meaning expected under the original specification. With interpretational confounding between empirical and nominal meanings of a construct, measure weights of a latent variable are distorted and subsequently the generalizability of the empirical test becomes problematic (Bagozzi 2007).

External consistency refers to the similarity in correlations of individual measures of a construct with other variables (Carver 1989). Specifically, this means that measures of a construct have a similar (positive/negative, significant/non-significant) relationship with the antecedents and consequences of the construct. External consistency is achieved when the measures of a construct correlate with the measures of other constructs (e.g., antecedents or consequences) in proportion to their correlation with the construct (Anderson and Gerbing 1982).

Opponents' View on the Issues

To the opponents of formative measurement, models with formative indicators are inherently more prone to interpretational confounding than are models with reflective indicators. They argue that this contamination is largely due to the fact that the formative indicator weights (or gammas) cannot be estimated without referencing other endogenous variables (Wilcox et al. 2008). In this situation, the estimation of a formatively designed construct can become a function of both its indicators and other extraneous elements such as the number and nature of dependent constructs and their measures, making it difficult to maintain the interpretational consistency of the formatively designed construct (Bagozzi 2007). When interpretational consistency varies contingent on dependent variables included or excluded, formative measurement becomes ambiguous (Edwards and Bagozzi 2000). Reflective measurement also carries the risks of estimation bias (Burt 1976). However, there exists a significant unitary (therefore replaceable) relationship among the reflective measures. The coefficients of reflective measures can be estimated without relying on other variables, thereby reducing the risk of such contamination.

Also, opponents argue against formative measurement based on issues of external consistency. Since formative indicators do not necessarily share a common theme, they do not have the same types of linkages with the antecedents and consequences of the construct (Coltman et al. 2008; Petter et al. 2007), implicitly assuming weak external consistency between formative indicators of a construct and indicators of other constructs. The lack of a common theme among formative indicators may not make a formatively designed construct an effective point variable that relays measured effects to dependent constructs in a consolidated manner (Howell et al. 2007). The opponents' argument is that because the formative measurement model is essentially a multiple regression, the formatively designed construct can act as a unitary entity (i.e., a point variable) only when there

is consistency of association strengths between the measures of exogenous and endogenous constructs. Therefore, external consistency may be the only way to make a formatively designed construct mediate the relationship between its indicators (or measures) and other constructs (Howell et al. 2007). When there is a lack of external consistency, a formatively designed construct placed as a point variable could become an information bottleneck in relaying the effect of its indicators on the dependent constructs. When a formatively designed construct merely reflects a composite value that best predicts the dependent variable rather than becoming an effective point variable, this may make it difficult to achieve model fit (Wilcox et al. 2008). In this situation, the discovery of “fit” models for a formatively measured construct may be an accidental outcome.

Proponents' View on the Issues

Unlike opponents, proponents believe that formative measurement is a reliable alternative to reflective measurement (e.g., Jarvis et al. 2003; MacKenzie et al. 2005; Petter et al. 2007). Bollen (2007), a proponent, argues that interpretational confounding and the problem associated with external consistency raised above do not occur if the model (regardless of formative or reflective) is correctly specified. He points out two conditions for correct model specification: (1) having a true measurement model in which all formative indicators cause a single common latent construct, and (2) using an over-identified model in which a formatively designed construct has at least two endogenous constructs. In other words, model estimation becomes unstable if the true measurement model and the empirical measurement model are discrepant or the formatively designed construct is a part of the under-identified model with only one endogenous construct. He argues that the opponents' criticism is flawed because their results are grounded on the estimation of such misspecified models.

Given the differing views on the reliability of formative measurement, researchers agree that there is a need for further empirical scrutiny to resolve the contending perspectives. For instance, Diamantopoulos et al. (2008, p. 14) acknowledged that

Finally, there is a debate on whether formative measurement is really necessary, that is, whether it should be used in the first place. Bagozzi (2007, p. 236), for example, states that, formative measurement can be done but only for a limited range of cases and under restrictive assumptions, while

Howell et al. (2007, p. 216; see also Wilcox et al. 2008) argue that “formative measurement is not an equally attractive alternative [to reflective measurement].” Although there are those (including the authors and Bollen 2007) who feel that, despite its various shortcomings, formative measurement is indeed a viable alternative to reflective measurement based on conceptual grounds, further theoretical and methodological research is necessary to finally settle this debate. Time will tell.

In summary, proponents of formative measurement indicate that correct model specification is the key. To them, opponents have created correlation matrices between indicators without considering whether such correlation matrices are consistent with a formative indicator model (theory-based true measurement model). This means there is discordance between theoretical and empirical models. The empirical approach to create models from an arbitrarily manipulated correlation matrix results in structurally misspecified measurement models. Thus, it is preferable to formulate a model first and then ask whether its structure is consistent with a correlation matrix rather than the other way around. Further, unlike Bollen's argument above, key empirical studies of opponents (e.g., Bagozzi 2007; Howell et al. 2007; Wilcox et al. 2007) are usually based on over-identified models. Opponents, on the other hand, claim that Bollen's arguments are contradictory—as many proponents agree that formative indicators need not have the same nomologic net, the same antecedents and consequences (Jarvis et al. 2003; Podsakoff et al. 2003), or may not necessarily share a common theme (e.g., MacKenzie et al. 2005).

As shown, the debate on the reliability of formative measurement has just begun and there are opposing views not only between the two schools of thought but also within the proponent camp. Considering that studies grounded on formative measurement are growing in the IS field, the debate needs to be settled soon because the development path of this debate will set the forms and conditions for the use of formative measurement by the IS community.

Research Method

Research Method Strategies

Our goal in this research note is to provide a complete assessment of these opposing perspectives through careful empirical scrutiny. In order to enhance the validity of our study, the research methodology should be designed to empi-

rically assess the two contradicting perspectives in an objective manner. We adopt six design strategies to overcome the methodical limitations raised by each camp.

- (1) Empirical studies by opponents used simulated correlation matrices among hypothetical measures without regard to their theoretical foundations in terms of structural relationship among measures and constructs in place (Bollen 2007; Diamantopoulos et al. 2008). To neutralize the bias this may pose to the model estimation, we first define theoretically supported measurement models and use a real survey dataset, rather than a simulated one, for analysis (see Appendix A).
- (2) Proponents' consistent position has been that formative indicators do not need to share a common theme (Petter et al. 2007). To test the reliability of empirical study on this topic, therefore, it is important to maintain the fundamental assumption of formative measurement models in which formative indicators influence the common latent construct. This study, therefore, is conducted on formative measurement models that are theoretically grounded (see the "Study Variables" section).
- (3) As for the external consistency of formative indicators, while proponents implicitly believe in the reliability of formative measurement, opponents argue that a formatively designed construct cannot function as an effective point variable and produces biased fit indices. To assess the integrity of two differing viewpoints, our study develops baseline theoretical models that include a formatively designed construct and two endogenous constructs. Then, the models are estimated with varied correlations between the indicators of the formatively designed construct and those of endogenous constructs (see the "Test Models" and "Empirical Demonstration" sections).
- (4) Proponents agree that the assessment of formative measurement should be based on over-identified models. Under-identified models cannot offer the reliability in model estimation. Accordingly, this research is designed to use over-identified models in which one formatively designed construct relates to two reflectively designed endogenous constructs (see Figures 1 and 2 in the "Test Models" section).
- (5) It is argued that a misspecified formative measurement model faces the problems of interpretational confounding and external consistency (Bollen 2007), whereas opponents suggest that those problems are prevalent with

formative measurement. Accordingly, both correctly specified models and misspecified models need to be compared to assess the alternative positions (see the "Research Design" section).

- (6) Opponents argue that formative indicators are natively more prone to interpretational confounding than reflective indicators. The proponents' stance is that both approaches are viable as long as there is no model misspecification. Accordingly, this study compares two measurement approaches (formative and reflective) in their estimation reliability (see the "Research Design" section).

Based on these considerations, we developed a research design (described below) to empirically validate the contradicting views of the two camps.

Research Design

The research design was framed so that the estimation reliability of formative measurement versus reflective measurement, and that of misspecified measurement model versus correctly specified model, could be cross-validated and the consistency of the cross-validation could be observed. This framing resulted in four test cases: a formatively theorized construct is correctly specified as a formative one (case 1); a formatively theorized construct is misspecified as a reflective one (case 2); a reflectively theorized construct is correctly specified as a reflective one (case 3); and a reflectively theorized construct is misspecified as a formative one (case 4). See Table 1 for all possible outcomes of the measurement perspective.

If interpretational confounding is triggered by measurement model misspecification as argued by proponents, then the problem should only be recognized in case 2 and case 4, which contain misspecified measurement models. Also, the problem of external consistency (not an issue for reflective models) should be observed with the misspecified formative models as in case 4, but not with the correctly specified formative models as in case 1. On the other hand, if interpretational confounding is due to the inherent instability of formative measurement as suggested by opponents, then it should be present with only formatively specified models in case 1 and case 4, but not in case 2 and case 3. Also, if the external consistency problem is consistent with the opponents' view of formative measurement, then it should occur with all the formatively specified models in case 1 and case 4, regardless of whether a model is misspecified or not.

Table 1. Four Test Cases

	Formatively Specified	Reflectively Specified
Formatively Theorized (example: IT infrastructure flexibility)	Correct specification (case 1)	Misspecification (case 2)
Reflectively theorized (example: Relational knowledge)	Misspecification (case 4)	Correct specification (case 3)

Study Variables

For the empirical tests of the above cases, a survey dataset that included 243 responses is used. Two exogenous constructs are assessed: *IT infrastructure flexibility* as a formatively theorized construct and *relational knowledge* as a reflectively theorized construct. Their definitions follow:

- *IT infrastructure flexibility*: Ability of a firm's IT infrastructure to enable quick development and support of various system components.
- *Relational knowledge*: Ability of IT staff to communicate and work effectively with people in other functional areas.

These constructs were measured according to index construction procedures (e.g., Coltman et al. 2008; Diamantopoulos and Siguaw 2006; Diamantopoulos and Winklhofer 2001) and conventional scale development procedures (e.g., DeVellis 2003; Netemeyer et al. 2003) (refer to Appendix A for a general description of the survey design, measure items, and data collection).

Judging from the general attributes of the formative perspective (Coltman et al. 2008; Jarvis et al. 2003; Petter et al. 2007), for IT infrastructure flexibility there is a conceptually clear formative relationship between the construct and its indicators. First of all, the four survey items represent *connectivity*, *compatibility*, *application functionality*, and *data transparency*, the key theoretical facets of a firm's IT infrastructure flexibility (Byrd and Turner 2000; Broadbent et al. 1996; Duncan 1995; Davenport and Linder 1994; Gibson 1993) (refer to Appendix A for a description of the items), and therefore are conceptually divergent (*not interchangeable*). Second, the indicators form the concept space of IT infrastructure flexibility and their addition or deletion moderates the construct's conceptual domain. Third, the four indicators shape the degree of IT infrastructure flexibility, but not the reverse (*direction of causality*). Finally, although formative indicators are not expected to have specific patterns or magnitude of intercorrelations—that is, formative indica-

tors might correlate positively or negatively or lack any correlation (Diamantopoulos et al. 2008)—their internal consistency may not be high because of the theoretical discreteness among indicators (Coltman et al. 2008). As shown in Table 2, indicator correlations of IT infrastructure flexibility are moderate, ranging from 0.23 through 0.44.

Relational knowledge is a reflectively theorized construct considering the following four theoretical and empirical criteria. The four indicators (refer to Appendix A) share a common theoretical theme in that they reflect the behavioral capacity influenced by an IT person's relational knowledge and thus are *interchangeable*; the addition or removal of an indicator has little effect on the content validity of the construct; variation in the four indicators is caused by variation in the construct but not the reverse (*direction of causality*); and, with the theoretical homogeneity among indicators, high internal consistency is expected (*indicator correlations*). As shown in Table 2, relational knowledge revealed high correlations, ranging from 0.77 to 0.88.

In addition to the two exogenous variables, five endogenous variables are considered to construct test models. Their definitions follow:

- *Financial performance*: A firm's general financial performance over the past 3 years.
- *IT performance*: Perceived IT contribution to business outcomes.
- *Business process performance*: Operational efficiency of inter- and intra- organizational processes.
- *Effectiveness in IT planning*: Quality of planning for the introduction, utilization, and adjustment of IT.
- *Effectiveness in IT coordination*: Quality of coordination on IT issues between IT people and users.

Test Models

Figures 1 and 2 are structural models formed to test the four cases presented in Table 1. Considering the issue of model identification and scaling, each test model is composed of one

Table 2. Measure correlations of exogenous constructs

IT infra. Flexibility	x1	x2	x3	x4	Relational Knowledge	x1	x2	x3	x4
x1	1.00				x1	1.00			
x2	0.24	1.00			x2	0.83	1.00		
x3	0.36	0.44	1.00		x3	0.77	0.83	1.00	
x4	0.23	0.38	0.36	1.00	x4	0.79	0.84	0.88	1.00

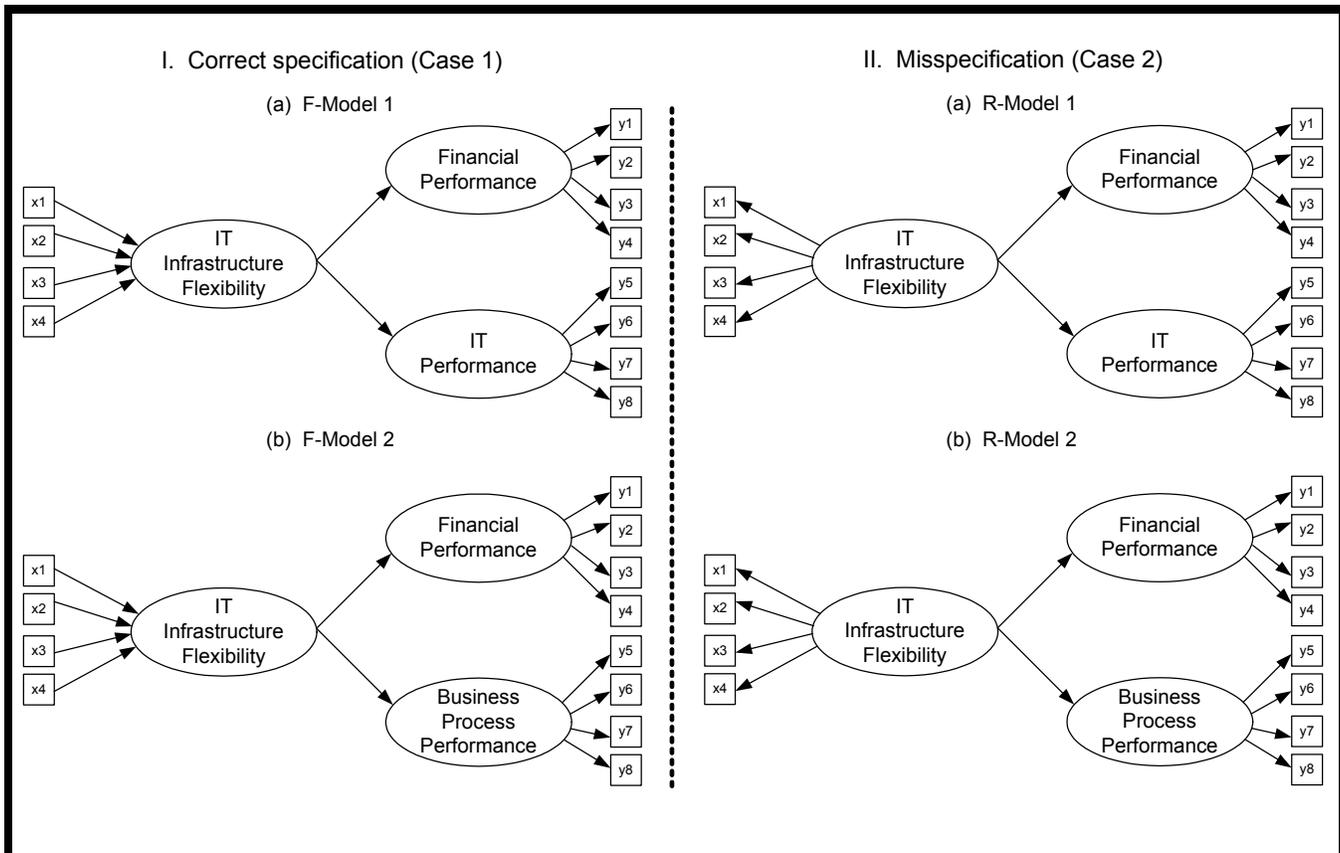


Figure 1. Research Model: IT Infrastructure Flexibility

exogenous construct and two endogenous constructs. As a formatively designed construct cannot be identified alone when the covariance-based SEM is used, Jarvis et al. (2003, p. 214) offered three approaches that enable the identification of a formatively designed construct: (1) a formatively designed construct identified by relying on two reflective dependent constructs; (2) a formatively designed construct identified by having at least two reflective indicators; and (3) a formatively designed construct identified by having at least one reflective indicator and one reflective dependent

construct. They have a commonality in which at least two reflectively designed constructs or reflective indicators should be defined to identify a formatively designed construct.

These models also have higher reliability in terms of scaling than conventional scaling approaches (e.g., fixing the measure weight of a formative indicator; setting a path coefficient from a formatively designed construct to an outcome variable to unity; assuming error variance of a formatively designed construct is zero). The conventional scaling approaches entail

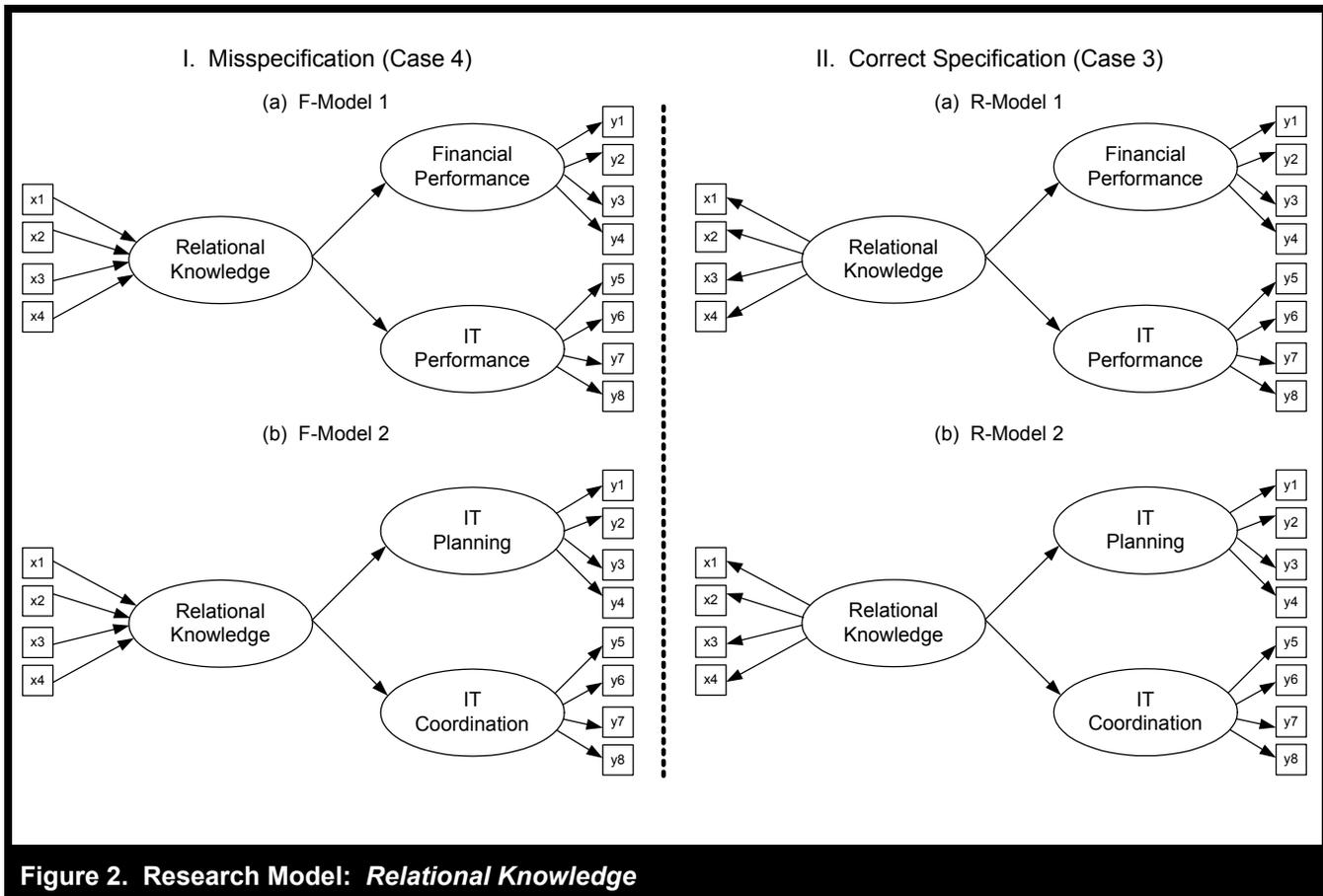


Figure 2. Research Model: *Relational Knowledge*

substantial arbitrariness in the estimation of measure weights, making them subject to bias in statistical inference and therefore inadequate for objective model comparison (Diamantopoulos et al. 2008; Franke et al. 2008). Among the three options suggested by Jarvis et al., the first option is more representative of IS research models and, therefore, is employed as our test model as shown in Figures 1 and 2.

To examine interpretational confounding, both reflective models and formative models are used in Figure 1 (test models with IT infrastructure flexibility as the exogenous construct) with case 1 (correct specification of a formatively theorized construct) and case 2 (misspecification). Figure 2 (test models with relational knowledge as the exogenous construct) also has case 3 (correct specification of a reflectively theorized construct) and case 4 (misspecification). Figures 1 and 2 use different sets of endogenous constructs. Here, the magnitude of interpretational confounding is examined by comparing weights of the exogenous construct between the two F-models and between the two R-models. Meanwhile, the effect of external consistency is studied only

with F-models because it is not an issue for reflective models with high indicator correlations (internal consistency) and subsequent stable external consistency. The correlation matrices in Appendix B (IT infrastructure flexibility) and Appendix C (relational knowledge) are used for the model estimation.

Empirical Demonstration

In this section, we empirically examine which of the research models are exposed to the problems of interpretational confounding and external consistency. To observe the consistency of model estimation results, models are tested with both LISREL and PLS that use covariance and component-based estimation respectively. Measure weights and path coefficients estimated by both LISREL and PLS are used for the discussion of interpretational confounding. The issue of external consistency is examined by reviewing the effect of weakened external consistency between the measures of

exogenous constructs and those of endogenous constructs on the overall model fit. As PLS is unable to estimate the model fit of a formatively designed construct, only LISREL is utilized for the examination of external consistency effect.

Interpretational Confounding

The degree of interpretational confounding can be observed by comparing the moderation in the standardized coefficients of a measurement model when its dependent variable(s) changes.

Case 1 (Correct Formative Model Specification)

Table 3 summarizes the estimation results of two correctly specified formative models (F-Model 1 and F-Model 2) for test of case 1 and comparable, misspecified reflective models (R-Model 1 and R-Model 2) for test of case 2, both with IT infrastructure flexibility as the exogenous construct. The correlation matrices in Appendix B are the basis for the coefficient estimation of the models. The results (path coefficients and fit indices) confirm the nomological validity of the formative model, which refers to congruity between the theoretically hypothesized relationship and empirical relationship between constructs. Comparison of two F-models reveals wide changes in standardized measure weights of IT infrastructure flexibility with both PLS and LISREL. For example, the value of $\gamma_{(1,2)}$ drops from 0.23/0.14 in F-Model 1 to 0.14/0.07 in F-Model 2 with PLS/LISREL. Also, the value of $\gamma_{(1,3)}$ increases from 0.45/0.25 in F-Model 1 to 0.60/0.37 in F-Model 2 with PLS/LISREL. The weight of $\gamma_{(1,4)}$ also experience noticeable divergence between the two F-models.

In F-Model 1, the formatively designed construct is a function of x_1 , x_3 , and x_4 , rather than a function of all measures (as per the theoretical meaning) in both PLS and LISREL. When a dependent variable was replaced in F-Model 2, IT infrastructure flexibility becomes a function of only x_3 (PLS) and of x_1 and x_3 (LISREL). This seems to indicate the estimation instability of formative measures with the choice of dependent constructs. This demonstrates that, even when the construct and measures are correctly specified, their relationship changes depending on the choice of endogenous variables.

Case 2 (Misspecified Reflectively)

Here, although the relationship between IT infrastructure flexibility and its measures are formative in nature, they are

misspecified to have reflective relationships. Given the measurement model misspecification and that the experiment is conducted with much weaker internal correlations (refer to Table 2) than the regular reflective measures, the loadings of the reflectively estimated IT infrastructure flexibility construct should be unstable if proponents are right. The standardized loadings in R-Model 1 and R-Model 2, however, were either identical or changed little with both PLS and LISREL, indicating the measurement model estimation was unaffected by the change of a dependent variable, by the misspecification of the exogenous construct, and by the low indicator correlations of the exogenous variable. This demonstrates that the relationship between reflective measures and their underlying construct remains stable *even when the relationship between the construct and its measures are misspecified*.

In terms of model fit, misspecified reflective models (R-Model 1 and R-Model 2) yielded model fit estimates matching those of adequately specified formative models (F-Model 1 and F-Model 2) (see Table 3, Panel B). What is notable is that F-Model 2 has only two and one significant measure weights with LISREL and PLS respectively. Nonetheless, F-Model 2 achieved comparable fit with R-Model 1 and R-Model 2, whose measure weights are all statistically significant. Furthermore, F-Model 2 attained model fit comparable to F-Model 1, which has three significant measure weights in both LISREL and PLS.

Case 3 (Correct Reflective Model Specification)

Table 4 summarizes the estimation results of the two correctly specified reflective models (R-Model 1 and R-Model 2) for test of case 3 and two comparable, misspecified formative models (F-Model 1 and F-Model 2) for test of case 4, both with relational knowledge as the exogenous construct. The correlation matrices in Appendix C are the basis for the coefficient estimation. It should be noted that relational knowledge is designed to be reflective and therefore its estimation as a formative model results in measurement model misspecification. When the test is conducted on reflective R-models, the standardized loadings of the exogenous construct are either identical or changed little, with all four indicators statistically significant with both LISREL and PLS, indicating that there is virtually no interpretational confounding.

Case 4 (Misspecified Formatively)

When the reflective variable is misspecified as a formatively designed construct, corresponding F-models show consider-

Table 3. Model Estimation and Fit Indices

Panel A: Exogenous Variable: IT Infrastructure Flexibility								
	PLS				LISREL			
	F-Model 1	F-Model 2	R-Model 1	R-Model 2	F-Model 1	F-Model 2	R-Model 1	R-Model 2
Dependent variables	Fin. Perform. IT Perform.	Fin. Perform. Bus. Proc. Perform.	Fin. Perform. IT Perform.	Fin. Perform. Bus. Proc. Perform.	Fin. Perform. IT Perform.	Fin. Perform. Bus. Proc. Perform.	Fin. Perform. IT Perform.	Fin. Perform. Bus. Proc. Perform.
Standard Weights /Loadings (t-value) of exogenous construct								
(1,1)	0.44 (4.11*)	0.41 (1.97)	0.66 (10.92*)	0.66 (5.75*)	0.28 (3.50*)	0.25 (2.77*)	0.50 (7.13*)	0.48 (6.78*)
(1,2)	0.23 (1.97)	0.14 (0.57)	0.71 (13.85*)	0.69 (6.53*)	0.14 (1.82)	0.07 (0.76)	0.58 (8.45*)	0.57 (8.26*)
(1,3)	0.45 (3.47*)	0.60 (2.73*)	0.79 (24.92*)	0.81 (15.38*)	0.25(2.96*)	0.37 (3.51*)	0.68 (10.15*)	0.73(10.71*)
(1,4)	0.27 (2.15*)	0.17 (0.71)	0.67 (12.73*)	0.64 (5.96*)	0.17 (2.30*)	0.11 (1.27)	0.53 (7.62*)	0.51 (7.27*)
Beta (path coefficients)								
(2,1)	0.28 (4.83*)	0.27 (2.98*)	0.27 (0.49)	0.27 (3.28*)	0.47 (6.22*)	0.48 (5.62*)	0.42 (5.54*)	0.39 (5.11*)
(3,1)	0.52 (9.55*)	0.44 (4.14*)	0.51 (9.63*)	0.43 (3.94*)	0.90 (6.17*)	0.77 (5.53*)	0.67 (8.69*)	0.56 (6.98*)
*p < 0.05								
Panel B: Exogenous Variable: IT Infrastructure Flexibility								
Dependent variables	F-Model 1	F-Model 2	R-Model 1	R-Model 2				
	Fin. Performance IT Performance	Fin. Performance Bus. Proc. Perform	Fin. Performance IT Performance	Fin. Performance Bus. Proc. Perform				
Fit Index (LISREL Only)								
Chi-sq(df)	85.49 (46)		83.17 (46)					
p-value	0.0004		0.0006					
RMSEA	0.060		0.058					
GFI	0.944		0.946					
NFI	0.970		0.965					
CFI	0.987		0.983					

Notes: RMSEA = root-mean-square error of approximation, GFI = Goodness of Fit Index, NFI = Normed Fit Index, CFI = Comparative Fit Index

able fluctuations in the relationship between the measures and constructs. The path coefficients, meanwhile, confirm the nomological validity with congruity between the theoretically hypothesized relationship and empirical relationship of constructs. Nonetheless, in both LISREL and PLS, relational knowledge is a function of x4 indicator in F-Model 1 and of x1, x2, and x4 measures in F-Model 2, rather than a function of all measures. Also, with the change of endogenous variables, measure weights experience considerable changes.

Despite the misspecification of the relationship between the formative exogenous construct and its indicators and the subsequent weak tie between theoretical and empirical meanings, F-models yield more favorable model fit estimates than R-models. The result resembles the cases of IT infrastructure flexibility in which F-models, despite their weak ties between formative measures and the construct, achieves a fit no worse than that of R-models.

External Consistency

To see the effect of weak external consistency on overall model fit, a simulation method is used for which the degree of external consistency is artificially worsened from the original model (baseline model) by changing correlations between formative measures and the measures of endogenous constructs. Here, we would like to emphasize that arbitrary changes of correlation strengths make it difficult for model conversion when real survey data are used. For instance, lowering already low correlations, raising already high correlations, or partial adjustments of correlation strengths triggers abnormality of the entire correlation matrix, preventing the structural model from conversion. The natural data therefore fundamentally differ from made-up data that had been used by most past studies for simulated experiments (e.g., Howell et al. 2007). Accordingly, we had to go through many trials and errors to find correlation adjustments that allow model con-

Table 4. Model Estimation and Fit Indices

Panel A: Exogenous Variable: Relational Knowledge								
	PLS				LISREL			
	F-Model 1	F-Model 2	R-Model 1	R-Model 2	F-Model 1	F-Model 2	R-Model 1	R-Model 2
Dependent variables	Fin. Perform IT Perform	IT Planning IT Coord'n	Fin. Perform IT Perform	IT Planning IT Coord'n	Fin. Perform IT Perform	IT Planning IT Coord'n	Fin. Perform IT Perform	IT Planning IT Coord'n
Standardized Weights /Loadings (t-value) of exogenous measures								
(1,1)	0.01 (0.04)	0.37 (2.76*)	0.25(13.46*)	0.27 (26.75*)	0.03 (0.29)	0.31 (3.06*)	0.86 (16.66*)	0.87 (16.93*)
(1,2)	0.31 (1.22)	0.36 (2.69*)	0.27 (22.45*)	0.28 (35.76*)	0.08 (0.70)	0.29 (2.45*)	0.91 (18.24*)	0.91 (18.43*)
(1,3)	-0.23 (-0.84)	-0.20 (-1.25)	0.26 (17.99*)	0.25 (27.82*)	-0.10 (-0.90)	-0.22 (-1.82)	0.92 (18.70*)	0.91 (18.43*)
(1,4)	0.92 (3.67*)	0.52 (2.82*)	0.30 (25.06*)	0.28 (41.43*)	0.43 (2.81*)	0.36 (2.90*)	0.94 (19.22*)	0.93 (19.04*)
Beta (path coefficients)								
(2,1)	0.17 (2.17*)	0.66 (15.78*)	0.17 (2.93*)	0.64 (15.15*)	0.36 (4.51*)	0.95 (9.23*)	0.17 (2.54*)	0.68 (9.79*)
(3,1)	0.48 (9.92*)	0.47 (8.12*)	0.46 (9.09*)	0.47 (8.20*)	1.17 (4.49*)	0.69 (9.18*)	0.49 (7.21*)	0.54 (7.72*)
*p < 0.05								
Panel b: Exogenous Variable: Relational Knowledge								
	F-Model 1	F-Model 2	R-Model 1	R-Model 2				
Dependent variables	Fin. Performance IT Performance	IT Planning IT Coordination	Fin. Performance IT Performance	IT Planning IT Coordination				
Fit Index (LISREL)								
Chi-sq(df)	76.30 (46)	61.23 (46)	134.06 (52)	144.44 (52)				
p-value	0.0033	0.0658	0.0000	0.0000				
RMSEA	0.052	0.037	0.081	0.086				
GFI	0.950	0.960	0.915	0.915				
NFI	0.980	0.987	0.964	0.964				
CFI	0.993	0.997	0.978	0.978				

Notes: RMSEA = root-mean-square error of approximation, GFI = Goodness of Fit Index, NFI = Normed Fit Index, CFI = Comparative Fit Index

vergence. Because of such restrictions in utilizing real data, tested models have slightly different levels of adjustment in the correlation strength.

Given that there are four measures (x1–x4) for the exogenous construct and eight measures (y1–y8) for two endogenous constructs of a test model, two different approaches are used to weaken external consistency. In the first approach, we changed correlations to make x1/x2 relate more strongly to y1–y4, and x3/x4 relate more weakly to y1–y4 from the baseline model. Then we reversed the situation in which x1/x2 relate weakly to y5–y8, and x3/x4 relate more strongly to y5–y8. In the second approach, the correlations between x1–x4 and y1–y8 are all increased but at differentiated strengths. For example, those between x1/x2 and y1–y4 are increased more than those between x3/x4 and y1–y4. Then, those between x1/x2 and y5–y8 are increased less than those between x3/x4 and y5–y8. Both approaches produced corre-

lation matrices whose external consistency is weaker than that of the baseline model. Then, the correlation matrices are applied to case 1 and case 4 because the issues of external consistency are germane to formatively specified models.

Case 1 (Correct Formative Model Specification)

Four tested scenarios, with correlations changed from the baseline models (all resulting in weakened external consistency), are summarized in Table 5. Two of them (scenarios 1 and 2) are derived from F-Model 1 and the other two (scenarios 3 and 4) from F-Model 2 in Figure 1. It can be seen that Scenario 2 becomes worse in external consistency than scenario 1 from the baseline F-Model 1. Also, scenario 4 has a weaker external consistency than scenario 3 from the baseline F-Model 2.

Table 5. Four Scenarios for External Consistency Test

F-Model	Scenarios	Endogenous Constructs	Increase/Decrease in Correlations from Baseline
F-Model 1 in Figure 1	1	Financial performance IT performance	(x1,x2 — y1~y4): +0.05, (x3, x4 — y1~y4): -0.05 (x1,x2 — y5~y8): -0.05, (x3, x4 — y5~y8): +0.05
	2	Financial performance IT performance	(x1, x2 — y1~y4): +0.1, (x3, x4 — y1~y4): -0.1 (x1, x2 — y5~y8): -0.1, (x3, x4 — y5~y8): +0.1
F-Model 2 in Figure 1	3	Financial performance Business process performance	(x1, x2 — y1~y4): +0.1, (x3, x4 — y1~y4): +0.05 (x1, x2 — y5~y8): +0.05, (x3, x4 — y5~y8): +0.1
	4	Financial performance Business process performance	(x1, x2 — y1~y4): +0.2, (x3, x4 — y1~y4): +0.05 (x1, x2 — y5~y8): +0.05, (x3, x4 — y5~y8): +0.2

F-Model 1: IT infrastructure flexibility (x1~x4); financial performance (y1~y4); IT performance (y5~y8)

F-Model 2: IT infrastructure flexibility (x1~x4); financial performance (y1~y4); business process performance (y5~y8)

Table 6. External Consistency and Model Fit with IT Infrastructure Flexibility

	F-Model 1			F-Model 2		
	Baseline model	Scenario 1	Scenario 2	Baseline model	Scenario 3	Scenario 4
Fit Index						
Chi-square (df)	85.49 (46)	88.11(46)	113.48(46)	85.49 (46)	89.96(46)	121.68(46)
P-value	0.0004	0.0002	0.0000	0.0004	0.0001	0.0000
RMSEA	0.060	0.062	0.078	0.060	0.063	0.082
GFI	0.944	0.943	0.928	0.944	0.942	0.923
NFI	0.970	0.969	0.959	0.970	0.965	0.955
CFI	0.987	0.985	0.974	0.987	0.982	0.970

Notes: Financial performance and IT performance are dependent variables for F-Model 1.

Financial performance and Business process performance are dependent variables for F-Model 2.

If the proponents’ arguments are right, the degree of external consistency should not affect overall model fit because the formatively defined construct behaves as an effective point variable. The summary of estimation (Table 6), however, reveals that the model fit of four scenarios gets worse from the baseline model in both F-Models, coinciding with the weakening external consistency. This is consistent with the simulation result of Howell et al. (2007). Despite the fact that the models are correctly specified formative models, the results are at odds with the proponents’ view.

Case 4 (Misspecified Formatively)

As with case 1, the correlations between the measures of exogenous and endogenous constructs were changed through four scenarios to worsen the external consistency from the base line models and then their estimation results are compared. Table 7 summarizes the changes of correlations from

the two formatively defined base-line models in Figure 2. It can be seen that scenario 2 is worse than Scenario 1 in external consistency for F-Model 1. Likewise, scenario 4 is worse than scenario 3 in external consistency for F-Model 2.

As seen in Table 8, all fit indices from the first scenario of F-Model 1 became worse off than those of the baseline model, reflecting the weakened external consistency. The second scenario, however, showed an almost perfect model fit despite the fact that the model had a poorer external consistency than the other two models (baseline and scenario 1). This pattern repeats in F-Model 2 in which scenario 4 results in a far better model fit than baseline and scenario 3. The unexpected reversal of association between the degree of external consistency and model fit conflicts with the findings by Howell et al. (2007). This seems to reflect that the artificially manufactured dataset is generally not bound by restrictions of data manipulations that the real survey dataset faces. Overall, the result underscores that the model fit is considerably affected

Table 7. Four Scenarios of External Consistency Test

F-Model	Scenarios	Endogenous Constructs	Increase/Decrease in Correlations from Baseline
F-Model 1 in Figure 2	1	Financial performance IT performance	(x1,x2 — y1~y4): +0.2, (x3, x4 — y1~y4): +0.1 (x1,x2 — y5~y8): +0.1, (x3, x4 — y5~y8): +0.2
	2	Financial performance IT performance	(x1, x2 — y1~y4): +0.4, (x3, x4 — y1~y4): +0.1 (x1, x2 — y5~y8): +0.1, (x3, x4 — y5~y8): +0.4
F-Model 2 in Figure 2	3	Effectiveness in IT planning Effectiveness in IT coordination	(x1, x2 — y1~y4): +0.05, (x3, x4 — y1~y4): -0.05 (x1, x2 — y5~y8): -0.05, (x3, x4 — y5~y8): +0.05
	4	Effectiveness in IT planning Effectiveness in IT coordination	(x1, x2 — y1~y4): +0.1, (x3, x4 — y1~y4): -0.1 (x1, x2 — y5~y8): -0.1, (x3, x4 — y5~y8): +0.1

F-Model 1: Relational knowledge (x1~x4); financial performance (y1~y4); IT performance (y5~y8)

F-Model 2: Relational knowledge (x1~x4); effectiveness in IT planning (y1~y4); effectiveness in IT coordination (y5~y8)

Table 8. External Consistency and Model Fit with Relational Knowledge

	F-Model 1			F-Model 2		
	Baseline model	Scenario 1	Scenario 2	Baseline model	Scenario 3	Scenario 4
Fit Index						
Chi-square (df)	76.30(46)	121.93(46)	46.50(46)	61.23(46)	127.70(46)	33.96(46)
P-value	0.0033	0.0000	0.4517	0.0658	0.0000	0.9056
RMSEA	0.052	0.083	0.007	0.037	0.086	0.000
GFI	0.950	0.923	0.969	0.960	0.919	0.977
NFI	0.980	0.972	0.962	0.987	0.970	0.971
CFI	0.993	0.982	0.997	0.997	0.980	1.000

Note: Financial performance and IT performance are dependent variables for F-Model 1.

Effectiveness in IT planning and effectiveness in IT coordination are dependent variables for F-Model 2.

by the degree of external consistency and their relationship can be unpredictable. This unpredictability can make it difficult for the formatively designed construct (regardless of accuracy in model specification) to behave as a reliable point variable.

results and shows that the problems of interpretational confounding and weakened external consistency only occurred among formatively specified cases (case 1 and case 4). If measurement model misspecification was the cause, the problems should have been found in misspecified cases (case 2 and case 4), but not in case 1 or case 3 which include correct model specification.

Analysis Results and Root Causes

The analysis indicates that *regardless of correct or incorrect specification, all formative models revealed problems associated with interpretational confounding and weakened external consistency*. Interpretational confounding was found in both formative models that include misspecified relational knowledge and correctly specified IT infrastructure flexibility. The results were consistent in both LISREL and PLS. Problems associated with reduced external consistency also existed for the two formative models regardless of the accuracy of model specification. Table 9 summarizes the test

The fluctuation of standardized measure weights and their statistical significance of formative models were considerably larger than their reflective counterparts. This was consistent throughout the cases regardless of correctness in the measurement model specification. For example, in F-Model 1 with IT infrastructure flexibility, the formatively specified construct is a function of x1, x3, and x4, rather than a function of all measures (as per the theoretical meaning). When a dependent variable was replaced in F-Model 2, IT infrastructure flexibility becomes a function of only x3 (PLS) and of x1 and x3 (LISREL). This seems to confirm the estimation corruption of formative measures based on the choice of the depen-

Table 9. Cases with Interpretational Confounding and External Consistency Problems		
	Formatively Specified	Reflectively Specified
Formatively Theorized (IT Infrastructure Flexibility)	<p>Correct specification (case 1)</p> <ul style="list-style-type: none"> ✓ Problems with interpretational confounding observed ✓ Problems with external consistency and model fit observed 	<p>Misspecification (case 2)</p> <ul style="list-style-type: none"> ✓ Problems with interpretational confounding not observed
Reflectively Theorized (Relational Knowledge)	<p>Misspecification (case 4)</p> <ul style="list-style-type: none"> ✓ Problems with interpretational confounding observed ✓ Problems with external consistency and model fit observed 	<p>Correct specification (case 3)</p> <ul style="list-style-type: none"> ✓ Problems with interpretational confounding not observed

dent constructs. As a result, the empirical meaning of the relationship between formative measures and their construct became arbitrary and therefore detached from nominal meaning and theoretical presumptions.

Regardless of the accuracy of model specification, the analysis indicates an association between the level of external consistency and model fit for formative measures. Results suggest that the unpredictability of model fit (mainly its deterioration) due to weakened external consistency is endemic to the formative approach. The worsening external consistency makes it difficult for a formatively designed construct to behave as a point variable that conveys the effects of formative measures to dependent constructs and, subsequently, the model fit suffers. The test also reveals that even in the situation in which the formatively designed construct cannot effectively behave as a point variable because of weakened external consistency, the overall model fit can improve (refer to scenario 4, F-Model 2 in Table 8). This inconsistency between the level of external consistency and model fit suggests that model estimation becomes arbitrary and, therefore, the estimation reliability of formative models is questionable.

It is also noted that formative models tend to yield more favorable path coefficients (beta values) than reflective models. This was consistently observed for the two formative models regardless of their correctness in specification (see Table 3, Panel A and Table 4, Panel A). Also, the overall model fit was either comparable to or better than that of reflective models (refer to Table 3, Panel B and Table 4, Panel B). Despite that there is a disjoint between nominal and empirical meanings (e.g., reduction in significant measure weights), the higher beta values and model fit is observed.

This appears to indicate that the estimation of formative measure weights is done in a manner that optimizes path coefficients and overall model fit, which may even compromise the structural integrity between formative measures and their construct. Therefore, the formative model may conveniently raise the probability of obtaining a better model fit and path coefficients, resulting in a higher chance of type I error in which the insignificant model fit or path is declared as significant. Thus, formative approach may cause IS researchers to reach incorrect conclusions regarding the hypothesis testing.

The root cause of these problems may be due to the fact that the exogenous construct with formative indicators is undefined and inestimable without relying on reflective dependent variables. Accordingly, the formative measure weights of constructs become the function of the number and nature of endogenous constructs and their measures (Bagozzi 2007). In the course of estimation, therefore, measure weights are resolved in a manner to optimize the overall fit and path coefficients of a model. This may not only trigger the corruption of measure weights, but also inflate the estimates. In addition to the weight contamination problem, a construct's inability to mediate the effect of all formative measures on dependent constructs seems to further aggregate estimation bias. The formatively defined construct is a linear combination of measures and its instability as a point variable that represents the measures seems to get worse when the degree of external consistency deteriorates. The analysis of the four test cases indicate that the alleged weaknesses of formative models in terms of interpretational confounding and external consistency are not caused by measurement model misspecification (e.g., reflective to formative, formative to reflective), but by the innate frailty in parameter estimation.

Recommendations

Our analysis reveals that formative measurement could pose fundamental problems in its estimation stability. We would hasten to add however, that this effort is just the beginning of, hopefully, a fruitful debate between the two opposing perspectives. Nonetheless, with the increasing adoption of formative measurement in the IS research, we hope that the time is right for a cautionary note on the blanket adoption of formative techniques, partially due to their apparent ease of execution through tools like PLS.

So, where do we go from here? Based on our findings, some may consider that an appropriate path for the field is to avoid formative measurement. However, this idea would be premature given that further theoretical and empirical studies are warranted to settle the debate (Diamantopoulos et al. 2008). Alternatively, we can curtail the risk formative measurement poses to model estimation by following certain guidelines. These can be divided into the pre-data analysis phase (Figure 3) and the data analysis phase (Figure 4). The pre-data analysis phase focuses on the development of a theoretically valid formative measurement model. The key activity in the data analysis phase is to examine the formative measurement model in terms of the potential risks associated with interpretational compounding and external consistency.

Pre-Data Analysis Phase

In the early stage of this phase, researchers choose constructs to launch the target study. In deciding constructs, some studies support the notion that a construct is inherently either formative or reflective, and therefore should be specified and measured as such (e.g., Hardin et al. 2008; Jarvis et al. 2003; Podsakoff et al. 2003).² However, there is evidence that even constructs generally specified by formative indicators can be transformed into reflective ones adequately. For example, Kluegel et al. (1977) showed the alteration in which formative measures of the socioeconomic status (SES) construct were converted to reflective ones. Borsboom et al. (2004, p. 1069) note that, "One may also imagine that there could be procedures to measure constructs like SES reflectively—for example, through a series of questions like *How high are you up the social ladder?*" In the IS field, Petter et al. (2007) offered an example in which the formative measures of operational excellence (Gattiker and Goodhue 2005) were

transformed into reflective ones. Therefore, researchers can remain open-minded on the measurement approach.

Once constructs of interest are decided, the researcher needs to develop an auxiliary theory. The auxiliary theory is a measurement-model-related theory that includes the conceptualization of the nature and direction of the relationship between a construct and its measures (Diamantopoulos and Siguaw 2006; Edwards and Bagozzi 2000). In general, an auxiliary theory should include (1) a theoretical definition of a construct and its content domains and (2) criteria (antecedents and consequences associated with a construct) (for more details, see Netemeyer et al. 2003). Generally, IS researchers have paid more attention to the nomological theory that explains the relationship between constructs than the auxiliary theory. However, a flawed auxiliary theory and subsequently the weakened measurement model can seriously threaten the estimation integrity of the structural model, and therefore researchers should be careful in developing a solid auxiliary theory.

In defining the theoretical meaning of a construct and its content domain, a researcher first needs to consider the directionality between a construct and its indicators. Current studies make it clear that the choice of the indicator type (reflective versus formative) should be based on theoretical considerations (or auxiliary theory) regarding the direction of linkage between a construct and its indicators (Diamantopoulos and Siguaw 2006; Edwards and Bagozzi 2000). In some cases, this choice is straightforward as there is a clear causal priority. For example, if a researcher chooses the SES (socio-economic status) construct and decides on education, income, and occupation as the content domain, the formative relationship from the indicators to SES can easily be decided.

For many constructs, however, the choice between the reflective and the formative measurement becomes somewhat arbitrary as the directionality of the relationship is far from obvious (Bollen and Ting 2000; Diamantopoulos and Siguaw 2006; Edwards and Bagozzi 2000). We believe that one of the reasons for this obscurity is that the theoretical definitions of most constructs fail to provide a clear guidance of the indicator type. For example, the construct *coercive power* introduced by French and Raven (1959) is defined as "the target's belief that the agent has the ability to punish him or her," without referring to relevant indicators. Subsequently, the construct has been subject to both formative and reflective measurement (Diamantopoulos and Winklhofer 2001). This means that in many situations where there is much ambiguity with the theoretical definition, researchers need to use their own discretion to determine the theoretical definition and subsequently the indicator type. It is, therefore, suggested that,

²There is lack of agreement on this point. Others, such as Howell et al (2007) and Bollen (2007), argue that constructs are neither formative nor reflective but it is the indicators that are formative or reflective.

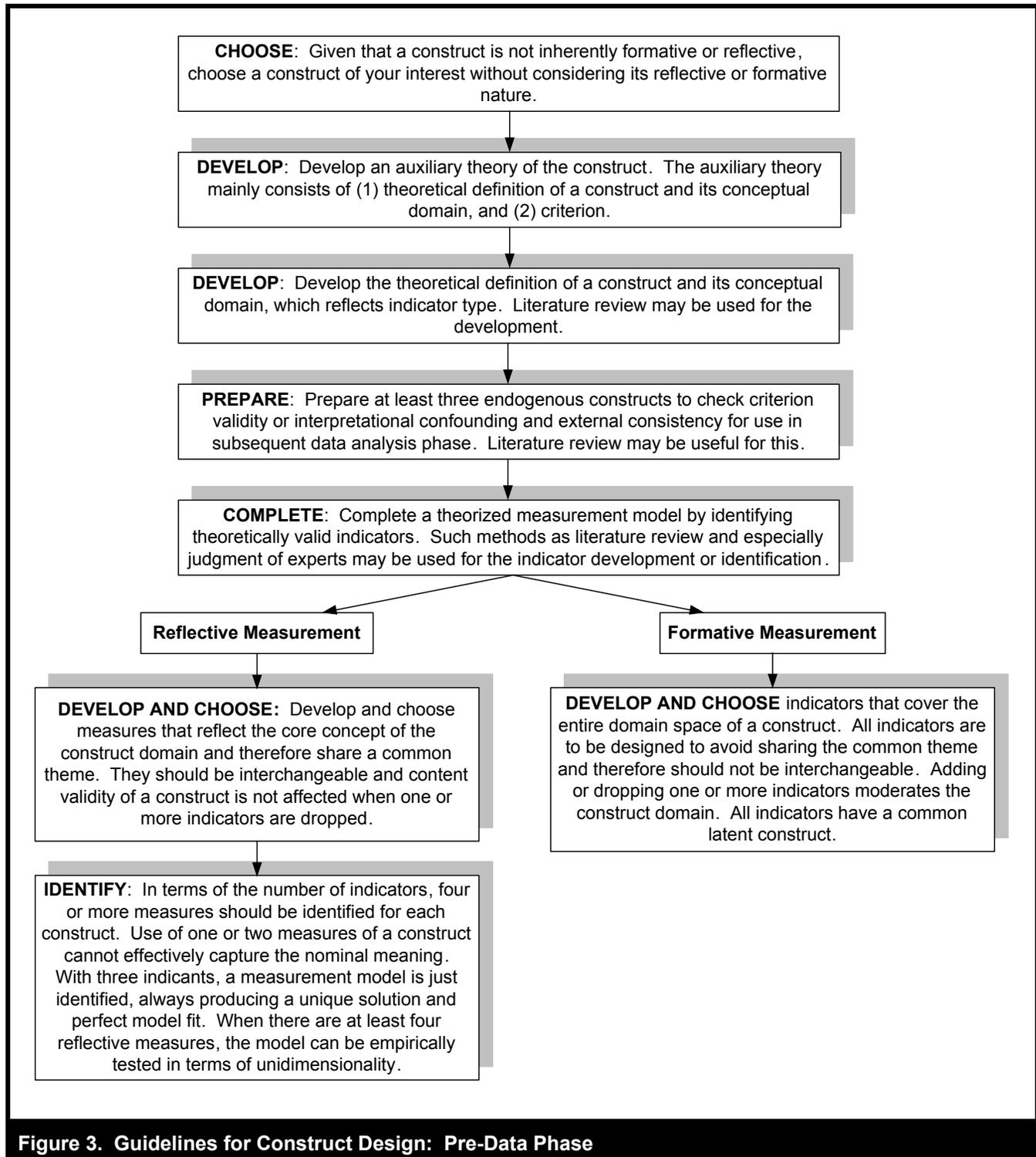


Figure 3. Guidelines for Construct Design: Pre-Data Phase

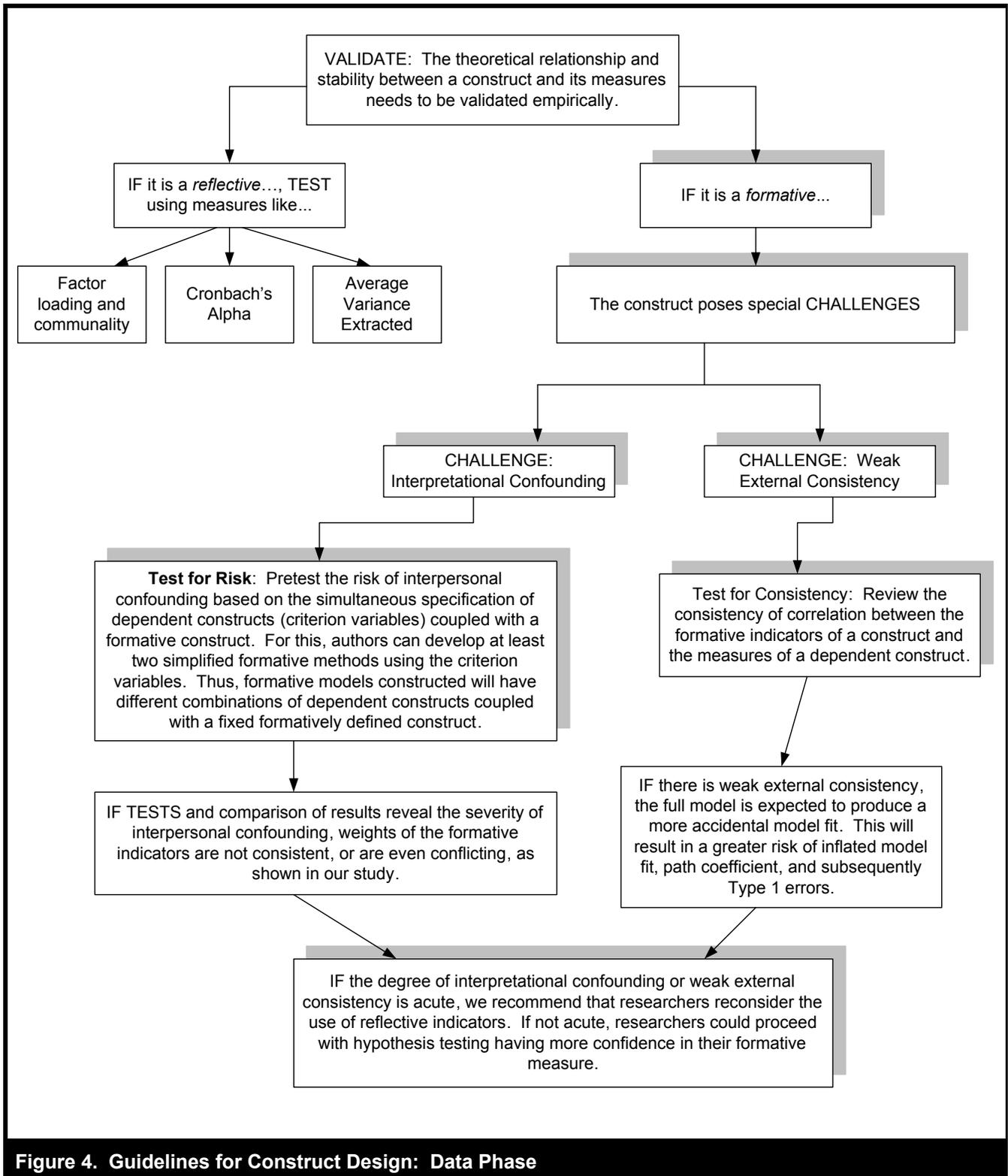


Figure 4. Guidelines for Construct Design: Data Phase

if the theoretical definition of a construct signifying the indicator type is not readily available, a researcher develops it by identifying adequate content domains via literature review. This part of the auxiliary theory should contribute to a correctly specified measurement model.

When the first part of the auxiliary theory is completed, then the researcher reviews literature and identifies major criterion variables (antecedents or consequences of a construct) evaluating the situation or the boundary in which the construct is used. Using criterion variables, a researcher who chooses the formative indicator type can assess the problem associated with interpretational confounding and external consistency at the data analysis phase. To examine the severity of interpretational confounding, at least one endogenous construct needs to be different in two different formatively defined structural models. This assumes that each model uses at least two endogenous constructs to avoid the under-identified model and the limitations of scaling (Diamantopoulos et al. 2008; Franke et al. 2008; Jarvis et al. 2003; Petter et al. 2007). Thus, the researcher should prepare at least three endogenous criterion variables (latent constructs) theoretically related to a formatively defined exogenous construct.

Once the auxiliary theory is thus completed, the next phase is to construct a theoretical measurement model by identifying theoretically valid indicators grounded on the auxiliary theory. As frequently pointed out by extant studies (e.g., Coltman et al. 2008; Diamantopoulos et al. 2008), theoretically valid formative indicators should cover the entire domain of a construct, avoid sharing a common theme to make them non-interchangeable, and result in the moderation of the construct domain if one or more indicators are added or dropped. Also, as pointed out by Bollen (2007), all formative indicators should have a causal priority toward the common latent construct. In contrast, theoretically valid reflective indicators manifest the core concept of the construct domain, share a common theme, are interchangeable, and do not affect the content validity of the construct when one or more indicators are dropped or added. It was pointed out that the identification of theoretically valid indicators requires systematic literature review and/or consultation with domain experts (DeVellis 2003).

A researcher should also consider the number of indicators for a construct. For the formative measurement, the number of indicators changes according to the breadth and complexity of the construct domain (Netemeyer et al. 2003). As for the reflective measurement, four or more measures should be used for a construct although many previous studies have relied on less than four indicators. Having one or two measures cannot effectively capture the nominal meaning of a

construct. With three indicators, a measurement model is just identified, always producing a unique solution and a perfect model fit. When there are at least four reflective measures, the model can be empirically tested in terms of unidimensionality. With theoretically valid indicators, a researcher completes the development of a theory-driven measurement model, and it becomes the basis of the correctly specified measurement model.

Data Analysis Phase

At this phase, due to the potential estimation bias of formative measurement, we suggest that researchers pre-examine data, giving consideration to interpretational confounding and external consistency. In other words, the stability between a formatively designed construct and its measures needs to be empirically validated based on the theoretical measurement model derived at the pre-data analysis stage. The validation examines the consistency between the theoretical measurement model introduced and its empirical measurement model. The method utilized by our study can be applied for the validation. To examine the existence of interpretational confounding, the researcher needs to form two different formative models, each of which contains two dependent constructs (criterion variables) coupled with the formatively defined construct. The two models are over-identified and free from structural misspecification (Bollen 2007), and overcome the limitations of conventional scaling approaches (Diamantopoulos et al. 2008; Franke et al. 2008). On comparing the two models, interpretational confounding is present if the weights of formative indicators are not consistent or are even conflicting, as shown in our study.

The problem associated with external consistency can be investigated by estimating the structural model based on the correlations in which the consistency between the formatively designed construct's indicators and the two endogenous constructs' indicators is weakened. Here, the same formative models used for the interpretational confounding test become the baseline models. If the models' fit improves despite the weaker external consistency, this is an indication of *accidental* model fit with inflated path coefficients and subsequent type I errors.

If the pretest does not reveal substantial problems associated with interpretational confounding and weak external consistency, the researcher can have confidence in the formative indicators and may continue the analysis to understand the relationship among studied constructs for hypothesis testing. If the problem of interpretational confounding or weak external consistency is severe, however, we recommend that the researcher consider the use of reflective indicators.

Conclusion

The growing use of formative indicators in IS research has been largely framed as a positive trend that enriches our ability to measure constructs more effectively. Recently, academic debate has ensued between proponents and opponents of formative measurement on the cause of recognized arbitrariness in formative model estimation, and subsequently its viability as a reliable tool. Given that much empirical effort is underway in IS research, there is danger that uncritical acceptance of the formative paradigm might lead to erroneous conclusions regarding hypotheses testing.

The central objective of this article is to empirically evaluate whether the alleged arbitrariness of model estimation is triggered by measurement model misspecification (proponent's perspective) or it simply represents the shortcoming of formative models (opponent's perspective). For this, we utilized real survey data and examined the opposing explanations of both camps from the theoretical perspectives of interpretational confounding and external consistency of SEM. We found the existence of arbitrariness in estimating measure weights, path coefficients, and model fit, and it was triggered by the use of formative measurement itself. The findings were consistent. First, the formative approach resulted in a considerable interpretational confounding in which standardized weights of the measures of a formatively designed construct widely fluctuated with the change of dependent constructs. Second, the formative approach inflated path coefficients and model fit indices, increasing the chance of type 1 error. Third, weakened external consistency of a formative model contributed to the unpredictability of model fit, highlighting its risk as a point variable. The instability of estimation indicates that the alleged problems associated with the use of formative measurement could be rather fundamental, making them difficult to avoid or remedy. The usage of formative measurement, therefore, may have negative repercussions on the quality of IS research, especially given its increasing incidence in leading IS journals. Construct reusability could especially be jeopardized, leading to inconsistency in theory testing and barriers to building cumulative knowledge. The evidence underscores that caution is necessary in interpreting studies conducted on the formative paradigm. To reduce the risks emanating from its usage, we offered simple and practical guidelines for researchers on the planning and usage of formative measurement.

We hope that the consistent results found in this study spawn further empirical efforts on the efficacy of formative measures. The IS field is at an important juncture regarding formative measures, and there is potential for the views and subsequent mechanics to become institutionalized without

complete debate. We invite such debate—starting with comments on the work presented in this research note.

Acknowledgments

The authors would like to thank the senior editor, the associate editor, and the three anonymous referees for offering valuable comments that made this manuscript possible.

References

- Anderson, J. C., and Gerbing D. W. 1982. "Some Methods for Re-Specifying Measurement Models to Obtain Unidimensional Construct Measurement," *Journal of Marketing Research* (19:4), pp. 453-460.
- Bagozzi, R. P. 2007. "On the Meaning Formative Measurement and How It Differs From Reflective Measurement: Comment on Howell, Breivik, and Wilcox (2007)," *Psychological Methods* (12:2), pp. 229-237.
- Bagozzi, R. P., Yi, Y., and Philips, L. W. 1991. "Assessing Construct Validity in Organization Research," *Administrative Science Quarterly* (36:3), pp. 421-458.
- Boar, B. 1996. *Cost Effective Strategies for Client/Server Systems*, New York: John Wiley & Sons.
- Bollen K. A. 2007. "Interpretational Confounding Is Due to Misspecification, Not to Type of Indicator: Comment on Howell, Breivik, and Wilcox (2007)," *Psychological Methods* (12:2), pp. 219-228.
- Bollen, K., and Lennox, R. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2), pp. 305-314.
- Bollen K. A., and Ting, K. 2000. "A Tetrad Test for Causal Indicators," *Psychological Methods* (5:1), pp. 3-22.
- Borsboom, D., Mellenbergh, G. J., and van Heerden, J. 2004. "The Concept of Validity," *Psychological Review* (111:4), pp. 1061-1071.
- Boynton, A., Zmud, R. W. and Jacobs, G. C. 1994. "The Influence of IT Management Practice on IT Use in Large Organizations," *MIS Quarterly* (18:3), pp. 299-318.
- Broadbent, M., Weill, P., O'Brien, T., and Neo, B. S. 1996. "Firm Context and Patterns of IT Infrastructure Capability," in *Proceedings of the 14th International Conference on Information Systems*, J. I. DeGross, S. Jarvenpaa, and A. Srinivasan (eds.), Cleveland, OH, pp. 174-194.
- Burt, R. S. 1976. "Interpretational Confounding of Unobserved Variables in Structural Equation Models," *Sociological Methods and Research* (5:1), pp. 3-52.
- Byrd, T. A., and Turner, D. E. 2000. "Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a Construct," *Journal of Management Information Systems* (17:1), pp. 167-208.
- Carver, C. S. 1989. "How Should Multifaceted Personality Constructs be Tested? Issues Illustrated by Self-Monitoring, Attribu-

- tional Style, and Hardiness,” *Journal of Personality and Social Psychology* (56:4), pp. 577-585.
- Coltman, T., Devinney, T. M., Midgley, D. F. and Venaik, S. 2008. “Formative Versus Reflective Measurement Models: Two Applications of Formative Measurement,” *Journal of Business Research* (61:12), pp. 1250-1262.
- Davenport, T., and Linder, J. 1994. “Information Management Infrastructure: The New Competitive Weapon?,” in *Proceedings of the 27th Hawaii International Conference on System Sciences*, Wailea, HI, January 4-7, pp. 885-896.
- Davenport, T. H., and Short, J. E. 1990. “The New Industrial Engineering: Information Technology and Business Process Redesign,” *Sloan Management Review* (31:4), pp. 11-28.
- DeSanctis, G., and Jackson, B. M. 1994. “Coordination of Information Technology Management: Team-Based Structures and Computer-Based Communication Systems,” *Journal of Management Information Systems* (10:4), pp. 85-110.
- DeVellis R. F. 2003. *Scale Development: Theories and Applications* (2nd ed.), Thousand Oaks, CA: Sage Publications.
- Diamantopoulos A., Riefler, P., and Roth, K. P. 2008. “Advancing Formative Measurement Models,” *Journal of Business Research* (61:12), pp. 1203-1218.
- Diamantopoulos, A., and Sigauw, J. A. 2006. “Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration,” *British Journal of Management* (17), pp. 263-282.
- Diamantopoulos, A., and Winklhofer, H. M. 2001. “Index Construction with Formative Indicators: An Alternative to Scale Development,” *Journal of Marketing Research* (38:2), pp. 259-277.
- Duncan, N. B. 1995. “Capturing Flexibility of Information Technology Infrastructure: A Study of Resource Characteristics and Their Measure,” *Journal of Management Information Systems* (12:2), pp. 37-57.
- Edwards, J. R., and Bagozzi, R. P. 2000. “On the Nature and Direction of Relationships between Constructs,” *Psychological Methods* (5:2), pp. 155-174.
- Franke, G., Preacher, C., and Rigdon, E. 2008. “The Proportional Structural Effects of Formative Indicators,” *Journal of Business Research* (61:12), pp. 1129-1237.
- French, Jr., J. R., and Raven, B. 1959. “The Bases of Social Power,” in *Studies in Social Power*, D. Cartwright (ed.), Ann Arbor, MI: Institute for Social Research, pp. 150-167.
- Gattiker, T. F., and Goodhue, D. L. 2005. “What Happens After ERP Implementation: Understanding the Impact of Interdependence and Differentiation on Plant-Level Outcomes,” *MIS Quarterly* (29:3), pp. 559-585.
- Gibson, R. 1993. “Global Information Technology Architecture,” *Journal of Global Information Management* (4:1), pp. 28-38.
- Hardin, A. M., Change, J. C., and Fuller, M. A. 2008. “Formative vs. Reflective Measurement: Comment on Marakas, Johnson, and Clay (2007),” *Journal of Association for Information Systems* (9:9), pp. 519-534.
- Howell, R. D., Breivik, E., and Wilcox, J. B. 2007. “Reconsidering Formative Measurement,” *Psychological Methods* (12:2), pp. 205-218.
- Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. “A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research,” *Journal of Consumer Research* (30:2), pp. 199-218.
- Jiang, J. J., Klein, G., Slyke, G. V. and Cheney, P. 2003. “A Note on Interpersonal and Communication Skills for IS Professionals: Evidence of Positive Influence,” *Decision Sciences* (34:4), pp. 799-812.
- Karimi, J., Somers, T. M. and Gupta, Y. P. 2001. “Impact of Information Technology Management Practices on Customer Service,” *Journal of Management Information Systems* (17:4), pp. 125-158.
- Kluegel, J. R., Singleton, R., and Starnes, C. E. 1977. “Subjective Class Identification: A Multiple Indicators Approach,” *American Sociological Review* (42:4), pp. 599-611.
- Lee, D. M. S., Trauth, E., and Farwell, D. 1995. “Critical Skills and Knowledge Requirements of IS Professionals: A Joint Academic/Industry Investigation,” *MIS Quarterly* (19:3), pp. 313-340.
- Li, E. Y., Jiang, J. J., and Klein, G. 2003. “The Impact of Organizational Coordination and Climate on Marketing Executives’ Satisfaction with Information Systems Services,” *Journal of the Association for Information Systems* (4), pp. 99-115.
- Mackenzie, S. B., Podsakoff, P. M., and Jarvis, C. B. 2005. “The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions,” *Journal of Applied Psychology* (90:4), pp.710-730.
- Nelson, R. 1991. “Educational Needs as Perceived by IS and End User Personnel: A Survey of Knowledge and Skill Requirement,” *MIS Quarterly* (15:4), pp. 503-525.
- Netemeyer, R. G., Bearden, W. O., and Sharma, S. 2003. *Scaling Procedures*, Thousand Oaks, CA: Sage Publications.
- Pavlou, P. A., and Gefen, D. 2004. “Building Effective Online Marketplaces with Institution-Based Trust,” *Information Systems Research* 15(1), pp.37-59.
- Petter, S., Straub, D., and Rai, A. 2007. “Specifying Formative Constructs in Information Systems Research,” *MIS Quarterly* (31:4), pp. 623-656.
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., and Lee, J. Y. 2003. “Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies,” *Journal of Applied Psychology* (88:5), pp. 879-903.
- Powell, T. C., and Dent-Micallef, A. 1997. “Information Technology as Competitive Advantage: The Role of Human, Business, and Technology Resources,” *Strategic Management Journal* (18:5), pp.375-405.
- Sabherwal, R. 1999. “The Relationship between Information System Planning Sophistication and Information System Success: An Empirical Assessment,” *Decision Sciences* (30:1), pp. 137-167.
- Segars, A. H., and Grover, V. 1999. “Profiles of Strategic Information Systems Planning,” *Information Systems Research* (10:3), pp. 199-232.
- Tesch, D., Jiang, J. J., and Klein, G. 2003. “The Impact of Information System Personnel Skill Discrepancies on Stakeholder Satisfaction,” *Decision Sciences* (34:1), pp. 107-127.

Wilcox, J. B., Howell, R. D., and Breivik, E. 2008. "Questions About Formative Measurement," *Journal of Business Research* (61:12), pp. 1219-1228.

About the Authors

Gimun Kim is an assistant professor of Information Systems at the College of Business Administration, Konyang University in Korea. He received his Ph.D. from Yonsei University, Korea, and M.S from Georgia State University. His research interests are in the business value of information technology capabilities, user behavior in electronic commerce, and research methodology related to structural equation modeling. He has published in such journals as *Information Systems Journal*, and *Information & Management*.

Bongsik Shin is a professor at the Department of Information and Decision Systems at San Diego State University. He earned a Ph.D. from the University of Arizona and has taught computer networking, electronic commerce, IT management/strategy, data warehousing, and statistics. His research interests include IT-driven business

models, electronic and mobile commerce, and research methodology. His work has been published in such journals as *Communications of the ACM*, *IEEE Transactions on Systems, Man, and Cybernetics*, *IEEE Transaction on Engineering Management*, *Journal of Association for Information Systems*, *Information Systems Journal*, *Journal of Management Information Systems*, *Information & Management*, *Journal of Organizational Computing and Electronic Commerce*, and *Decision Support Systems*.

Varun Grover is the William S. Lee (Duke Energy) Distinguished Professor of Information Systems at Clemson University. He has published extensively in the information systems field, with over 200 publications in refereed journals. Seven recent articles have ranked him among the top five researchers based on publications and citation impact in the top Information Systems journals over the past decade. In addition, he has received numerous awards for his research and teaching from professional and academic organizations. Dr. Grover is a senior editor for the *Journal of the AIS* and *Database*, and senior editor (emeritus) for *MIS Quarterly*. He is currently working in the areas of IT value and impact, and recently released his third book (with M. Lynne Markus) on process change.

Appendix A

Description of Research Method

Survey Design and Validation

Initially candidate survey items were compiled from existing literature and presented on a seven-point Likert scale. Then the survey items were filtered through several iterations of pretests. First, these items were examined by a focus group. This group consisted of three colleagues who are knowledgeable about the research subject as well as the measurement theory, and five senior IT managers with practical knowledge in IT infrastructure. This group of people met three times to examine the content validity of the research instrument during the two week period. Each time they met, the participants gradually reduced the items through intensive discussions. Second, the revised questionnaire was pretested with 20 senior IT managers. Each participant was asked to complete the questionnaire and, during the debriefing period, to offer any suggestions for improvement. Again, from this process, four survey items were dropped and several minor refinements on the remaining items were made. Third, the revised questionnaires were mailed to another group of 20 senior managers for a pilot test. Follow-up interviews with 20 respondents did not indicate any need for substantive changes to the questionnaire. This seemed to indicate strong preliminary evidence of construct validity. Survey items used for this research and their sources are listed next.

Survey Constructs, Measures, and Sources

IT Performance (Source: Powell and Dent-Micallef 1997)

1. New information technologies have dramatically increased our productivity.
2. New information technologies have improved our competitive position.
3. New information technologies have dramatically increased our sales.
4. New information technologies have dramatically increased our profitability.

Financial Performance (Source: Powell and Dent-Micallef 1997)

1. Over the past 3 years, our financial performance has been outstanding.
2. Over the past 3 years, our financial performance has exceeded our competitors'.
3. Over the past 3 years, our sales growth has been outstanding.
4. Over the past 3 years, we have been more profitable than our competitors.

Business Process Performance (Source: Davenport and Short 1990)

1. Our company' is better than competitors in connecting (e.g., communication and information sharing) parties within a business process.
2. Our company' is better than competitors in reducing cost and human labor within a business process.
3. Our company' is better than competitors in bringing complex analytical methods to bear on a business process.
4. Our company' is better than competitors in bringing detailed information into a business process.

IT Planning (Sources: Boynton et al. 1994; Karimi et al. 2001; Sabherwal 1999; Segars and Grover 1999)

1. We continuously examine the innovative opportunities for the strategic use of IT.
2. We enforce adequate plans for the introduction and utilization of IT.
3. We perform IT planning processes in systematic and formalized ways.
4. We frequently adjust IT plans to better adapt to changing conditions.

IT Coordination (Sources: Boynton et al. 1994; DeSanctis and Jackson 1994; Karimi et al. 2001; Li et al. 2003)

1. In our organization, IS and line people meet frequently to discuss important issues both formally and informally.
2. In our organization, IS people and line people from various departments frequently attend cross-functional meetings.
3. In our organization, IS and line people coordinate their efforts harmoniously.
4. In our organization, information is widely shared between IS and line people so that those who make decisions or perform jobs have access to all available know-how.

Relational Knowledge (Sources: Boar 1996; Broadbent et al. 1996; Byrd and Turner 2000; Duncan 1995; Jiang et al. 2003; Lee et al. 1995; Nelson 1991; Tesch et al. 2003)

1. Our IT personnel are very capable in terms of planning, organizing, and leading projects.
2. Our IT personnel are very capable in terms of planning and executing work in a collective environment.
3. Our IT personnel are very capable in terms of teaching others.
4. Our IT personnel work closely with customers and maintain productive user/client relationships.

IT Infrastructure Flexibility (Derived sources: Broadbent et al. 1996; Byrd and Turner 2000; Davenport and Linder 1994; Duncan 1995; Gibson 1993)

Our company' IT infrastructure is better than competitors' in

1. linking to any other components inside and outside the organizational environment.
2. sharing any type of information across any technology components.
3. adding, modifying, and removing the modules of business applications with little or no widespread effects on the applications collectively.
4. retrieving and flowing data between authorized personnel in an organization or between organizations regardless of the underlying database standards utilized.

Data Collection

Data were collected through a field survey. The firms in the DART System (an electronic system for public announcement) supervised by the Financial Supervisory Service of the Korean government was adopted as a sampling frame. This system includes a mailing list of 1,835 firms, comprising 629 firms listed on the Korea Stock Exchange, 857 firms listed on the KOSDAQ, and 349 unlisted firms. From this sampling frame, a random sample of 800 firms was chosen. Using a key informant approach (Bagozzi et al. 1991), the survey was targeted at senior IS executives such as CIOs, directors, and senior managers of the IT department, as they are most likely to be informed about strategic issues pertaining to IT decisions and practices in their organizations.

Senior IT executives were contacted via various channels (e.g., e-mail, phone) to solicit participation in this study. A total of 655 executives (81.9%) agreed to participate, indicating that their organizations have a formal and sizable IT function. Four weeks after the initial mailing, a follow-up survey was sent to those individuals who did not return a completed questionnaire. Finally, a total of 251 responses were received. Of those, eight responses were discarded due to incompleteness. With the final sample consisting of 243 responses (103 firms listed on the Korea Stock Exchange, 85 firms listed on the KOSDAQ, and 55 unlisted firms), the actual response rate was 37.1 percent. To check for non-response bias, respondents were compared with those in the mailing list in terms of organization size and industry. The results of the chi-square analyses revealed no systematic bias. Chi-square tests comparing early and late respondents on organization size and industry also revealed no significant response bias.

A significant portion (47.7 percent) of the respondents were either chief information officers or vice presidents of the IS department. The job titles of the other respondents (senior vice president, vice president of technology, assistant vice president, director of information technology) also indicate that they were senior IS executives. All the respondents indicated that they were within two levels from the highest position in their organizational hierarchy.

Appendix B

Correlation Matrices (for *IT Infrastructure Flexibility*)

IT Infrastructure Flexibility (x1–x4), Financial Performance (y1–y4), and IT Performance (y5–y8)												
Input	y1	y2	y3	y4	y5	y6	y7	y8	x1	x2	x3	x4
y1	1.00											
y2	0.82	1.00										
y3	0.68	0.81	1.00									
y4	0.75	0.86	0.83	1.00								
y5	0.28	0.29	0.31	0.30	1.00							
y6	0.33	0.37	0.35	0.36	0.74	1.00						
y7	0.27	0.29	0.30	0.30	0.58	0.68	1.00					
y8	0.30	0.33	0.35	0.34	0.72	0.77	0.76	1.00				
x1	0.19	0.18	0.10	0.15	0.40	0.37	0.23	0.34	1.00			
x2	0.18	0.17	0.12	0.17	0.31	0.27	0.27	0.30	0.24	1.00		
x3	0.24	0.25	0.19	0.24	0.37	0.36	0.33	0.35	0.36	0.44	1.00	
x4	0.15	0.15	0.14	0.13	0.26	0.33	0.29	0.28	0.23	0.38	0.36	1.00

IT Infrastructure Flexibility (x1–x4), Financial Performance (y1–y4), and Process Performance (y5–y8)												
Input	y1	y2	y3	y4	y5	y6	y7	y8	x1	x2	x3	x4
y1	1.00											
y2	0.82	1.00										
y3	0.68	0.81	1.00									
y4	0.75	0.86	0.83	1.00								
y5	0.31	0.27	0.21	0.27	1.00							
y6	0.30	0.29	0.24	0.37	0.70	1.00						
y7	0.30	0.29	0.23	0.31	0.61	0.71	1.00					
y8	0.24	0.24	0.20	0.25	0.61	0.67	0.69	1.00				
x1	0.19	0.18	0.10	0.15	0.31	0.23	0.25	0.31	1.00			
x2	0.18	0.17	0.12	0.17	0.21	0.17	0.15	0.29	0.24	1.00		
x3	0.24	0.25	0.19	0.24	0.29	0.31	0.33	0.38	0.36	0.44	1.00	
x4	0.15	0.15	0.14	0.13	0.18	0.21	0.16	0.25	0.23	0.38	0.36	1.00

Appendix C

Correlation Matrices (for *Relational Knowledge*)

Relational Knowledge (x1–x4), Financial Performance (y1–y4), and IT Performance (y5–y8)												
Input	y1	y2	y3	y4	y5	y6	y7	y8	x1	x2	x3	x4
y1	1.00											
y2	0.82	1.00										
y3	0.68	0.81	1.00									
y4	0.75	0.86	0.83	1.00								
y5	0.28	0.29	0.31	0.30	1.00							
y6	0.33	0.37	0.35	0.36	0.74	1.00						
y7	0.27	0.29	0.30	0.30	0.58	0.68	1.00					
y8	0.30	0.33	0.35	0.34	0.72	0.77	0.76	1.00				
x1	0.16	0.13	0.05	0.12	0.33	0.37	0.33	0.36	1.00			
x2	0.19	0.16	0.08	0.16	0.39	0.40	0.34	0.37	0.83	1.00		
x3	0.17	0.14	0.04	0.13	0.37	0.38	0.32	0.36	0.77	0.83	1.00	
x4	0.17	0.15	0.06	0.14	0.44	0.44	0.38	0.43	0.79	0.84	0.88	1.00

Relational Knowledge (x1–x4), IT Planning (y1–y4), and IT Coordination (y5–y8)												
Input	y1	y2	y3	y4	y5	y6	y7	y8	x1	x2	x3	x4
y1	1.00											
y2	0.73	1.00										
y3	0.67	0.82	1.00									
y4	0.60	0.71	0.68	1.00								
y5	0.46	0.47	0.48	0.47	1.00							
y6	0.41	0.43	0.45	0.45	0.62	1.00						
y7	0.52	0.53	0.52	0.45	0.72	0.69	1.00					
y8	0.43	0.47	0.45	0.42	0.62	0.62	0.69	1.00				
x1	0.52	0.58	0.57	0.48	0.31	0.35	0.45	0.36	1.00			
x2	0.57	0.57	0.56	0.47	0.30	0.36	0.47	0.36	0.83	1.00		
x3	0.51	0.48	0.47	0.40	0.32	0.36	0.46	0.32	0.77	0.83	1.00	
x4	0.59	0.55	0.54	0.46	0.35	0.35	0.48	0.35	0.79	0.84	0.88	1.00