POST-PRINT FINAL AUTHOR COPY
ACCEPTED IN JOURNAL OF BUSINESS RESEARCH

REFLECTIVE ANALYSIS OF FORMATIVE CONSTRUCTS: RELEVANT SIMULATION EVIDENCE

Woojung Chang [a, *], George R. Franke [b,1], Nick Lee [c,2]

[a] College of Business, Illinois State University, 322 State Farm Hall of Business, Campus Box 5590, Normal, IL 61790-5590, USA

[b] Culverhouse College of Commerce, University of Alabama, P.O. Box 870225, Tuscaloosa, AL 35487-0225, USA

[c] School of Business and Economics, Loughborough University, Loughborough LE11 3TU, United Kingdom

Note: The authors are listed in alphabetical order.

* Corresponding author. Tel.: +1 309 438 8429; fax: +1 309 438 3508.

Email addresses: wchang3@ilstu.edu (W. Chang), gfranke@cba.ua.edu (G.R. Franke), N.Lee@lboro.ac.uk (N. Lee).

[1] Tel.: +1 205 348 9435; fax: +1 205 348 6695.

[2] Tel.: +44 1509 228829; fax: +44 1509 223960.

Comparing reflective and formative measures: new insights from relevant simulations

Abstract

Previous simulations comparing formative and reflective models specify formative population models as the only correct model for a given construct, and compare them with various misspecified reflective models. However, this approach does not generalize to situations where both reflective and formative specifications can work well to assess constructs. To address this limitation, this study presents simulations in which both formative and reflective specifications fit the underlying population data equally well. The results show that reflective specifications generate less biased and more powerful results than formative specifications, and make a strong case for considering standardized rather than unstandardized coefficients for both specifications. Therefore, conceptual and empirical consequences of using reflective models for constructs that could also be modeled as formative are less dire than past research has suggested.

Keywords: Construct specification; Formative measurement; Reflective measurement; Monte Carlo simulation.

"Things exist, we don't have to create them; we only have to understand their relationships; and it is the threads of these relationships which form poems and orchestras."

Stéphane Mallarmé

Réponses à des Enquêtes sur l'Evolution Littéraire

1. Introduction

Mallarmé's colorful comment on literature is also relevant to empirical research, because "the sine qua non of measurement is that the numbers assigned to objects reflect the relations among the objects with respect to the aspect being measured" (Pedhazur & Schmelkin, 1991, p. 16). A well-established body of simulation-based research claims that using a reflective measurement model to operationalize attributes that "should have been formatively modeled" has serious consequences in terms of estimating structural relationships between different objects—that is, theoretical constructs (Jarvis, MacKenzie, & Podsakoff, 2003, p. 207; see also MacKenzie, Podsakoff, & Jarvis, 2005; Petter, Straub, & Rai, 2007). For example, MacKenzie et al. (2005, p. 728) assert that "misspecification can inflate unstandardized structural parameter estimates by as much as 400% or deflate them by as much as 80%," with "a substantial probability that measurement model misspecification will not be detected with many of the most commonly used goodness-of-fit indices."

Aguirre-Urreta and Marakas (2012) counter that these claims rely on comparisons of unstandardized coefficients between reflective and formative models, and that using standardized coefficients removes the apparent bias in the structural estimates due to claimed reflective misspecification. In response, Jarvis, MacKenzie, and Podsakoff (2012) and Petter, Rai, and Straub (2012) defend the use of unstandardized coefficients in structural models as a basis for comparing empirical findings across specifications. However, focusing on the relative magnitudes of unstandardized coefficients and rejecting the interpretation of standardized results is contrary to common research practice, would preclude the use of partial least squares modeling (which relies on standardized coefficients) as a formative analysis tool, and overlooks such important criteria as the reliability and significance of measurement indicators and the variance explained in predicting endogenous constructs (Aguirre-Urreta & Marakas, 2012; Hair, Hult, Ringle, & Sarstedt, 2014).

Moreover, the conclusions based on existing simulation research (e.g., Jarvis et al., 2003; MacKenzie et al., 2005; Petter et al., 2007) are problematic in ways that are far more fundamental than those regarding the metric for comparison that Aguirre-Urreta and Marakas (2012) point out. For debate about the empirical comparisons to be productive, simulation designs must be realistic and appropriate. As this paper shows, existing simulation designs do not provide a fair and unbiased comparison of formative and reflective models, meaning that—whether using standardized or unstandardized coefficients—the comparisons between the formative and reflective models in existing simulations are not useful evidence for the strong conclusions drawn in prior work.

Specifically, previous simulations specify formative population models as the only correct model for a given construct, and compare them with various alternative reflective models. This approach has several limitations. One is treating arbitrary scalings of latent variables as the only true population values, whereas an infinite number of alternative scalings (including standardized solutions) would be equally true in the population (Bollen, 1989). Another limitation is confounding lack of fit caused by basic misspecifications in the particular reflective models, with that which may come from using the reflective model as an alternative to the formative model.

The latter issue is particularly important, because existing simulations do not recognize the real possibility that both reflective and formative specifications can work well to conceptualize and assess a construct. Various authors identify constructs in marketing and other fields that could be or arguably should be analyzed as formative, in contrast to the earlier researchers who successfully conceptualized, developed, and analyzed them as reflective (e.g., Diamantopoulos & Winklhofer, 2001; MacKenzie et al., 2005). For example, Jarvis et al. (2003) estimate that previous authors incorrectly modeled 29% of marketing measures as reflective rather than formative. In information systems, Petter et al. (2007) suggest a figure of 31%, and Podsakoff, Shen, and Podsakoff (2006) give a figure of 69% in research on strategy. These claims reject the original measurement conceptualizations and suggest that many papers include important errors in empirical analyses, yet existing simulation studies ensure that the data fit the formative model rather than the reflective model (or both models). Thus the available evidence does not in fact shed light on the effects of formative analysis of (successful) reflective constructs.

A more realistic representation of such situations is to design simulations where formative and reflective specifications both fit perfectly in the population, and compare the results between the alternative models. The present study is the first to design such unbiased simulations, and thus to present comparisons relevant to practicing researchers. Based on simulation results that give a privileged position to formative specifications, Aguirre-Urreta and Marakas (2012, p. 124) present Heresy #1, that “the consequences of misspecification seem to be much less dire than previously thought.” The present paper provides a more extreme Heresy #2, that existing simulation evidence is either misleading or not even relevant to the important question of how alternative measurement specification affects tests of relationships between constructs when both formative and reflective models fit the data. Drawing from the results of these more relevant simulation designs, the paper also presents Heresy #3, that formative models with small sample sizes and inappropriate latent-variable scaling produce considerable bias in cases where both reflective and formative specifications are viable. Thus the paper makes substantial novel contributions to the ongoing conversation on measurement by presenting the first set of simulations that avoid preferential (i.e., biased) treatment of either the formative or reflective model.

This new approach is important because the resulting simulation findings can provide researchers with precise insights into how to operationalize constructs under the possible existence of multiple meaningful specifications, and with implications of successfully treating a measure as reflective when researchers could also have modeled the measure as formative or vice versa. Furthermore, the simulation results can help in evaluating the findings of much extant literature in marketing where researchers disagree about whether formative or reflective models are appropriate. If formative specifications would be consistent with the substantive implications of the (published) reflective specifications, then scholars can interpret the literature in marketing and other fields with more confidence than previous simulation evidence implies. Conversely, if formative specifications are prone to bias with typical research designs, then calls for replacement of reflective with formative models may abate.

As a foundation for the simulation design, the next section focuses on three issues that are critical to proper interpretation of simulation evidence on formative versus reflective specifications. The first issue is whether constructs are inherently formative or reflective; if not, researchers can reasonably evaluate them with alternative models. The second is why researchers can often empirically treat good reflective models as good formative models. The third is the arbitrary scaling of latent variables that underlies the interpretation of previous simulation results. Empirical examples taken from previous studies illustrate all three of these issues. After this foundation, the paper provides the findings of new simulations with multiple correct measurement models by taking random samples from populations where the formative and reflective specifications all fit the same data equally well. The final section discusses how the simulation results provide new perspectives on how to operationalize constructs under the existence of multiple correct models.

2. Conceptual background and illustrations

2.1. Are constructs inherently formative or reflective?

The formative measurement literature is often consistent with a *construct-centric* view, that constructs are inherently either formative or reflective, and thus scholars must model them accordingly (Howell, Breivik, & Wilcox, 2007). For example, Diamantopoulos and Siguaw (2006, p. 270) identify errors that result from using "the *wrong* measurement perspective, given the nature of the construct." Similarly, Petter et al. (2012, p. 148) suggest that researchers can "alter the meaning of a construct…by misspecifying the measurement model" and that "the meaning of the construct changes…based on the measurement specification being formative or reflective." Previous simulation studies comparing formative and reflective models are thus consistent with a construct-centric perspective in asserting that a formative model is true and reflective variants are false.

An alternative to the construct-centric view is the realist ontology of measurement that underlies most contemporary measurement theory in organizational and social science (e.g., Borsboom, 2005). The key tenet of the realist ontology as applied to measurement is that "constructs exist…independent of our attempts to assess them" (Markus & Borsboom, 2013, pp. 10-11). Therefore, "A given research situation or research tradition may favor either formative or reflective measurement, but constructs themselves, posited under a realist philosophy of science as existing apart from their measurement, are neither formative nor reflective" (Wilcox, Howell, & Breivik, 2008, p. 1220). The view that a construct is a composite of its formative indicators inevitably leads to the conclusion that the construct "has no measurable reality apart from those variables which are conceived to be its determinants" (Heise, 1972, p. 153). Many formative measurement theorists reject this operationalist perspective (e.g., Diamantopoulos & Winklhofer, 2001), leading to what may be called an *item-centric* view. This widely-held perspective takes the position that a set of indicators may be formative with respect to one construct but reflective with respect to another (e.g., Bollen & Ting, 2000), so that the measures of a construct do not define the construct, and the construct exists independent of them. Diamantopoulos (2011, p. 336), for example, notes that "there seems to be broad consensus in the literature that constructs themselves are not inherently formative or reflective."

The realist ontological perspective that constructs exist independent of their measurements opens the possibility of multiple viable construct specifications. Therefore, scholars may sometimes appropriately model a set of indicators for a construct both reflectively and formatively. Indeed, many constructs in marketing and business research have advocates for both reflective and formative operationalizations. Two examples (discussed in more detail subsequently) are market orientation and customer relationship management, and Bollen (1989) even shows formative and reflective models for socioeconomic status, which many scholars discuss as an archetypal construct suited to a formative approach. Understanding the empirical implications of competing theoretical perspectives is therefore a worthwhile goal. Well-designed simulation studies should be able to inform researchers on the comparative utility of formative and reflective models on the same data, and also provide guidance to researchers on how to examine their own data. Thus, from a realist, item-centric position, and in contrast to previous simulation studies, the relevant comparison of measurement models requires population data where both formative and reflective models can fit a given set of indicators equally well.

2.2. Formative analysis of reflective models

Prior simulation studies view formative and reflective models as incompatible alternatives. However, because of proportionality constraints on relationships between variables, formative and reflective approaches may often be more empirically compatible than the literature suggests. Formative indicators must have the same relative effect on every outcome that the construct influences as they do on the formatively-measured construct (Franke, Preacher, & Rigdon, 2008; Hayduk, 1987). Reflective indicators must have the same proportional relationships with the antecedents and consequences of the reflectively-measured construct as they have with each other (Anderson, Gerbing, & Hunter, 1987). Therefore, indicators that fit the more restrictive reflective model are often empirically compatible with a formative specification.

Fig. 1. illustrates this pattern. The first model shows a reflectively-measured construct with three indicators, influencing a construct with two indicators. If the model fits the data, the products of the loadings account for the correlations between indicators x1-x3 (e.g., $r_{x1x2} = \lambda_{x1}*\lambda_{x2}$). The second model of Fig. 1. also repeats this relationship, where the correlations between the formative indicators equal the products of the reflective loadings. Researchers conventionally interpret the latent variable in the formative model as the formative construct (e.g., MacCallum & Browne, 1993), though empirically the variable represents whatever the outcome variables have in common, as the reflective model in Fig. 1 indicates (e.g., Howell, 2013). The formative model uses the correlations between the indicators as parameters, whereas

the reflective model accounts for them with indicator loadings. Thus the reflective model usually has more degrees of freedom than the formative model, as well as a higher chi-square value to the extent that indicator correlations do not exactly equal the products of the loadings. Various indicators of model fit may be better or worse for either specification.

Fig. 1 here.

Because the issue of standardized versus unstandardized coefficients is critical to the interpretation of past simulation results, the next section discusses the scaling of latent variables in structural models.

2.3. Latent-variable scaling

Latent variables do not have an intrinsic scale, and researchers must specify constraints on some parameters to identify the models and obtain unique estimates for all the unconstrained parameters (e.g., Aguirre-Urreta & Marakas, 2012). Bollen (1989, p. 152) notes that "Virtually all latent variables have ambiguous scales…When agreement about the measurement unit for [a latent variable] is absent, then the scale choice is largely arbitrary." Two common approaches to setting the scale of latent variables include fixing a selected loading to a nonzero value (normally but not necessarily 1.0, called *unit loading identification*) or fixing the variance of the exogenous latent variable to a nonzero value (also commonly to a value of 1.0, called *unit variance identification*). However, other options include fixing causal paths to endogenous constructs, or the constructs' disturbance terms, to 1.0 or some other value (Kenny, Kashy, & Bolger, 1998). Aguirre-Urreta and Marakas (2012, p. 125) point out that "different approaches to model identification and scale setting, as well as different nonzero values used for this purpose, will result in markedly different unstandardized estimates of the relationship between latent variables, even if the alternative models will fit equally well; however, when standardized, the estimates will all be identical." Though the standardized coefficients will be equal, Gonzalez and Griffin (2001) show that different scaling approaches, including the indicator chosen for unit loading identification, can have a substantial influence on the estimated standard errors. Thus, the true parameter values of the formative population models examined in previous simulations could take on infinitely many other equally-true values, and significance levels for estimated parameters depend in part on the scaling method.

Jarvis et al. (2012) and Petter et al. (2012) argue against comparing standardized coefficients across random samples. However, even though unstandardized effects should be suitable when comparing results across groups, standardized coefficients are quite appropriate when comparing the same samples on different models as in the simulations of the current paper. Previous research also suggests that standardizing formatively-modeled constructs by constraining their structural errors to equal 1.0 minus the variance explained by the formative indicators both increases power and allows an assessment of all the formative indicators, rather than fixing one coefficient to a value of one (Diamantopoulos, 2011; Franke et al., 2008). The following illustrations show these differences.

2.4. Empirical examples

This section uses published correlation matrices to examine three constructs that researchers conceptualize both formatively and reflectively in previous research. The first example is from Chang, Park, and Chaiy (2010), who use a reflective model to analyze customer relationship management (CRM) technology. Other scholars prefer formative specifications for CRM measurement (e.g., Reinartz, Krafft, & Hoyer, 2004). The second example is from Rigopoulou, Theodosiou, Katsikea, and Perdikis (2012), who treat information control as a second-order, more abstract construct that accounts for the covariances between its first-order dimensions. In contrast, Challagalla and Shervani (1997) treat the different dimensions as separate constructs rather than lower-order factors. The third example is from Kara, Spillan, and DeShields (2004), who treat three dimensions of market orientation (MO) as first-order indicators of overall MO. Coltman, Devinney, Midgley, and Venaik (2008, p. 1258) argue for a formative perspective in conceptualizing MO, but note that "from an ontological standpoint, researchers can measure market orientation reflectively (cultural perspective) or formatively (behavioral perspective)."

For each example, Table 1 provides four sets of results. One reflective model, labeled RS in the table, has the exogenous construct scaled with unit variance identification and the endogenous construct scaled with unit loading identification. Table 1 reports the standardized loadings. The second reflective model, RU, uses unit loading identification for both the exogenous and endogenous constructs. Table 1 also provides two formative models for comparison. FS scales the latent variable by constraining the structural error term equal to 1.0 minus the variance explained by the indicators, which standardizes both the formative coefficients and the loadings of the outcome variables. FU scales the construct by setting the coefficient for the most significant formative indicator equal to 1.0. This approach gives more power for testing the other indicators than results from using a poorer indicator to scale the construct. The analyses use LISREL 8.80 to obtain the empirical results. Because the available correlations are for lower-order constructs rather than individual measurement items, structural relationships, fit, and degrees of freedom differ compared to the original publications.

Table 1 here.

The standardized RS and FS results are almost identical for the loadings of the outcome variables (except for an arbitrary change in sign for the information control results, resulting from a negative effect of the controls on the outcomes). Model fits are all good. In every case the direct effect of the exogenous construct is significant, and the reflective models all account for somewhat more variance explained (squared multiple correlations or SMC) than the formative models. For the reflective analyses, the results imply that the higher-order construct influences the outcome variables. Interpretation of the formative models considers the statistical significance of the indicators, which depends on the scaling of the latent construct. Scaling by fixing a formative loading to 1.0 is consistently less powerful than the standardized approach. With standardization, just one formative indicator is significant for CRM technology use; with the other method, none is significant. For information control, either all three or just one of the formative indicators is significant, depending on the scaling approach. For MO, either one or none of the indicators is significant.

Though these examples do not examine the theoretical arguments for the competing interpretations of the constructs considered, or the conceptualization of higher-order reflective models (Edwards & Bagozzi, 2000; Lee & Cadogan, 2013), they do show how formative and reflective approaches can be alternative models for the same construct. Both models fit well and explain similar amounts of variance overall. The biggest empirical difference is that, depending on the scaling of the formatively-measured variables, some or all of the formative indicators are nonsignificant. As such, the comparative findings of Table 1 suggest that researchers avoid scaling formative models with unit loading identification (i.e., the unstandardized approach), and rather use alternatives such as the standardized approach. To compare the effects of modeling decisions more systematically, and to examine more complicated models such as those examined by Aguirre-Urreta and Marakas (2012), Jarvis et al. (2003), and Petter et al. (2007), the next section presents Monte Carlo simulation results.

3. Simulations with multiple correct models

3.1. Simulation design

The present simulation builds on the design of Jarvis et al. (2003), which attempts to show the consequences of treating formative models as reflective. As Fig. 2 shows, Jarvis et al. (2003) create two correctly specified formative models (i.e., population models). One contains an exogenous formatively-measured construct (Fig. 2A) and the other contains an endogenous formatively-measured construct (Fig. 2B)[1]. Both designs have four reflectively-assessed constructs in addition to the formatively-measured construct. Jarvis et al. (2003) examine the effects of modeling the initial formatively-measured construct as reflective in the alternative model (Fig. 2C) in terms of magnitudes of unstandardized parameter estimates and model fit. Subsequent simulations replicate and extend this design (Aguirre-Urreta & Marakas, 2012; Petter et al., 2007). Therefore, the simulation of the present study builds on this design to allow comparisons with previous findings.

Fig. 2 here.

Except for arbitrary differences between unstandardized structural coefficients, the three sets of simulations reveal substantial similarities between the formative and reflective specifications. With a few exceptions, most reflective models show good power and fit the data on average according to commonly-recommended fit statistics. The simulation designs underlie the exceptions. When Jarvis et al. (2003) analyze the exogenous construct assessed by formative models in the population (Fig. 2A) reflectively (Fig. 2C), the estimated structural coefficients account for measurement error, leading to disattenuated (i.e., overestimated) correlations between other indicators relative to the formative model and thus reducing model fit. This pattern is especially evident when the indicator correlations are unrealistically low for reflective models ($r = 0.10$ or $r = 0.40$), producing the lowest reliability and the greatest degree of disattenuation. Conversely, when analyzing the endogenous construct treated as formative in the population (Fig. 2B) reflectively (Fig. 2C), the reduction in fit is highest for the highest indicator correlations ($r = 0.70$). This result is a consequence of indicators that are uncorrelated in the formative model (Y1-Y4 with Y9-Y12) being forced to correlate in the reflective model.

The new simulations in this paper avoid these design effects by making several substantial changes to Jarvis et al.'s (2003) specification. Crucially, in contrast to their assumption that researchers can correctly measure a construct only with one measurement specification, the new simulation allows the possibility that both reflective and formative measurement specifications can work for a given set of indicators to measure a construct. Thus, the population model avoids preferential treatment for one specification over the other such that the three models in Fig. 2 perfectly fit the same population covariance matrices. The new simulation design also specifies that the reflective indicators in all the models correlate 0.50 and thus have loadings of 0.71, rather than the less realistic correlations of 0.10 (with loadings of 0.32) or 0.40 (with loadings of 0.63). Furthermore, the simulations examine three different sample sizes of 250, 500, and 800 to examine how they affect model fit, convergence, and power or bias for the measurement models as well as for the structural models. The simulations include one thousand random samples for each design from the population covariance matrix and analyze them using LISREL 8.80. The population structural paths depend on the measurement model used as shown in Tables 2, 3, and 4 for the three models. The analyses examine how different the values on average obtained from one thousand random samples for each design are from the true values in the population, recognizing that the true values are simply one of infinitely many possible ways of scaling the latent constructs. Tables 2, 3, and 4 summarize the simulation results.

Tables 2, 3, and 4 here.

Given the arbitrariness of unstandardized analyses, and the simpler interpretability and higher power of standardized analyses, comparing the values obtained from random samples with the true population values based on standardized effects is more appropriate than using unstandardized effects. However, for comparison purposes, the tables report results for both standardized and unconstrained analyses. For consistency with the formative models, the completely-reflective analyses scale the exogenous construct with unit loading identification to provide unstandardized effects, and they constrain the variance of the endogenous construct for standardized effects.

3.2. Simulation results

In Table 2, the completely-reflective analyses almost perfectly reproduce the population results across three different sample sizes on average. All 1,000 samples produce acceptable solutions. The average χ^2 value is almost equal to the degrees of freedom as expected; fit statistics are good; and both standardized and unstandardized structural effects are essentially equal to their population values on average. Standardizing the endogenous constructs by constraining their variances increases their t-statistics somewhat. The standardized and unstandardized loadings of the reflective indicators of the exogenous construct, shown in the table for illustrative purposes, represent the results for all reflective indicators both here and in the formative simulations. In every case the loadings of the reflective indicators are almost exactly equal to their population values. The t-values of the reflective indicators are all highly significant and very similar between the standardized and unstandardized solutions.

Table 3 indicates the results when modeling the exogenous construct formatively. Almost 16% of the solutions of 1,000 samples are unacceptable with n = 250, mostly due to SMC values greater than 1. However, as the sample size increases from 250 to 500 and 800, acceptable solutions also increase, and only 2% of the solutions are unacceptable with n = 800. As with the reflective analyses, the acceptable models fit very well. The standardized and unstandardized structural coefficients are almost identical to their population values across three different samples on average. Especially with large samples (n = 800), the standardized and unstandardized structural coefficients equal their population values. The standardized loadings of the formative indicators are similar to their population values. However, when n = 250, the average unstandardized coefficients for the formative indicators are about 33% larger than the population values (1.33 versus 1.00). As the sample size increases, the unstandardized coefficients of the formative indicators get closer to the population values.

The standardized analysis is substantially more powerful than the unstandardized analysis, with greater t-statistics for structural relationships. The tests of formative indicators are also much more significant in the standardized analysis, with t-statistics of 2.93 or above versus just 1.71 for the unstandardized analysis. On average, about 3.33 (of 4) formative indicators are significant in the standardized analysis with n = 250, whereas on average all four formative indicators are significant in the standardized analysis with both n = 500 and n = 800. In contrast, in the unstandardized analysis with n = 250, just 0.91 formative indicators (of 3, since the analysis uses one indicator to set the scale) are significant. As the sample sizes increase, the number of significant formative indicators also increases. When n = 800, on average 2.99 of 3 formative indicators are significant.

Table 4 presents the simulation findings with an endogenous formatively-measured construct. When the formatively-measured construct is endogenous, the results generate a greater percentage of unacceptable solutions. With n = 250, almost 19% of the solutions are unacceptable, whereas about 12% of solutions with n = 500 and 5% with n = 800 are not acceptable, respectively. However, as with the reflective and exogenous formative models, the remaining acceptable solutions fit very well. On average, the standardized structural coefficients are almost identical to the population values across three different sample sizes. While the unstandardized effect of the exogenous construct on the reflectively-measured construct matches the population value on average, the upward bias occurs in the unstandardized influence of the exogenous construct on the formatively-modeled construct. However, the amount of the bias decreases as the sample size increases.

For the formative indicators, the standardized formative coefficients are almost equal to the population values and significant with a t-value of 2.54 or above. In the standardized analysis, on average 2.87 formative indicators of 4 are significant with n = 250, whereas on average 3.99 formative indicators are significant with n = 800. In the unstandardized analysis with n = 250, the formative indicators have an average t-statistic of just 1.48, and only 0.55 (of 3) indicators are significant on average. As the sample size increases, the number of significant formative indicators improves and on average 2.93 formative indicators of 3 are significant when n = 800. When the sample size is small, the unstandardized formative loadings are almost 37% larger than the population values on average (1.37 versus 1.00), whereas with a large sample size, the unstandardized formative loadings become closer to the population values.

3.3. Implications of the simulation results

A key implication of the new simulation results is that when reflective measurement models fit a data set—the situation underlying the motivation but not the design of prior simulation-based studies—the reflective models do not necessarily lead to different substantive interpretations about relationships between constructs from formative models applied to the same data. Contrary to the claims of prior simulations, which are drawn from analyses of misspecified models and unstandardized coefficients, the structural relationships among constructs are quite consistent in formative and reflective approaches. However, the new simulation results demonstrate that the reflective models are more powerful, and produce substantially more acceptable solutions, than formative models when items correlate highly enough to function as acceptable reflective models. Furthermore, the new findings indicate that models with an endogenous formatively-measured construct generate substantial bias in estimating the structural relationships with small sample sizes and unit loading identification.

Another crucial implication of the simulations is the value of standardized analysis that constrains the structural error term compared to unstandardized analysis. The superiority of standardized analysis is especially evident in loadings of formative indicators. When both measurement approaches fit a given set of items, treating them as formative indicators results in many nonsignificant formative indicators due in part to the effects of collinearity. In particular, this pattern is apparent with small samples and unstandardized analyses, which suggests that researchers avoid the use of unstandardized analysis for formatively-measured constructs in cases of small sample sizes. In contrast, standardized approaches almost perfectly reproduce formative indicators coefficients of the population values on average, are more powerful, and provide test results for all indicators. Omitting relevant but nonsignificant indicators misspecifies the formatively-modeled construct and produces biased structural coefficients. Therefore, standardized specifications do better than unstandardized specifications to satisfy criteria for evaluating formative and structural models (Hair et al., 2014). Because researchers using formative models should take steps to maximize the power of their analyses, the higher power of standardized analyses that helps detect significance of formative indicators is critical in formative models with smaller sample sizes. Other methods of scaling latent constructs not examined in the simulations may also be useful, but unit loading identification with a formative indicator is clearly not ideal.

4. Discussion and future research

One common theme in the measurement literature is that theoretical conceptualization of constructs takes precedence over empirical evidence in choosing between formative and reflective specifications (e.g., Diamantopoulos & Siguaw, 2006; Jarvis et al., 2012; Petter et al., 2012). Widespread classifications of measures that supposedly were analyzed with the wrong model, based solely on conceptual interpretations, suggest that some scholars discount the need for empirical evidence in making such decisions. However, under the circumstances in which theory indicates the potential applicability of both reflective and formative perspectives to operationalize constructs, researchers need to know the empirical implications of treating a construct as reflective when they could also have modeled the construct as formative (Fornell, Rhee, & Yi, 1991). To address this question, the current simulations compare the results of reflective, exogenous formative, and endogenous formative models by taking samples from the population in which all the models fit perfectly.

These results are relevant regardless of one’s stance on formative models and the value of empirical tests (Treiblmaier, Bentler, & Mair, 2011). Even scholars who doubt the existence of formative measurement models (e.g., Markus & Borsboom, 2013) may usefully model a given set of observed variables using both formative and reflective approaches. Moreover, the issues this paper addresses are substantively important in the scholarly interpretation of existing work that prior simulation studies criticize as being incorrectly modeled—despite showing empirically successful measurement analysis. Therefore, conceptual arguments and "mental experiments should be used in conjunction with empirical procedures, acknowledging that neither approach provides definitive evidence of association between a construct and a measure" (Edwards & Bagozzi, 2000, p. 158).

The new simulation results make several practical contributions to the measurement literature. First, the findings highlight the strength of reflective analysis when researchers can successfully assess a construct from both reflective and formative approaches. Treating a construct as formative in this situation clearly demonstrates potential bias and limited power in testing formative indicators and structural relationships among constructs, even though the formative model fits in the population. From this perspective, future researchers need to focus on how best to implement reflective analyses of measures that they could also analyze as formative, rather than on the consequences of using one model instead of the other.

The findings also demonstrate drawbacks of unit loading identification with formative indicators in examining formative models. Aguirre-Urreta and Marakas's (2012) simulation results verify that using standardized coefficients to compare formative with reflective analyses produces more consistent structural results than previously inferred from differences between (arbitrary) unstandardized coefficients. Standardizing formative latent variables by constraining their error terms increases statistical power and provides significance tests for all available formative indicators. Although Jarvis et al. (2012) and Petter et al. (2012) consider the focus on standardization to be an important weakness of the Aguirre-Urreta and Marakas (2012) paper, the arbitrariness of unstandardized analyses and the relative power of constrained standardized analyses support the comparison based on standardized effects. In many comparisons, the standardized approach is less biased and more powerful than unstandardized analyses, leading to fewer Type II statistical errors. Unstandardized approaches in conjunction with small samples generate considerable bias in estimating structural relationships among constructs, and to get reasonable results with unstandardized solutions requires larger sample sizes than may be feasible in many research settings.

Unit loading identification of the formative indicators is especially problematic. With smaller samples (n = 250), the unstandardized coefficients of the formative indicators are some 31% - 38% larger than their population value, which researchers might consider a substantial amount of bias in the formative loadings. Furthermore, due to the limited power of the unstandardized analyses, less than one-third of the unstandardized indicators are significant on average in this case. Considering that each formative indicator helps define a construct, the limited power of the unstandardized approaches in testing individual significance of each formative indicator could be a critical drawback, because researchers are likely to drop the nonsignificant indicator from the construct and as a consequence change the meaning of the construct.

5. Conclusion

For tests of theory to be meaningful, “it is essential that researchers correctly specify their measurement models to match their theoretical conceptualizations” (Jarvis et al., 2012, p. 140). In contrast to this principle, much formative literature rejects researchers’ views of their own measures, instead suggesting that only one conceptualization is appropriate for particular constructs—a view that contemporary measurement literature almost uniformly rejects (e.g., Markus & Borsboom, 2013). This perspective is also counterproductive if taken as a license to ignore the conceptual implications of observing that a supposedly formative construct functions effectively when modeled reflectively.

Rather than assuming that alternative specifications are necessarily misspecifications, researchers need to recognize that under certain circumstances scholars can appropriately model constructs using both reflective and formative approaches. In these situations, unit loading identification of formative models with small samples creates considerable bias and loss of power in estimating formative loadings. Reflective analysis often generates less biased and more powerful results, and leads to fewer problems with unacceptable analysis solutions. Because reflective indicators, like formative indicators, should have proportional relationships with other model variables, the good model fit observed with alternative specifications should not be surprising. Thus, in contrast with previous simulation studies, the current analysis shows that when researchers conceptualize their measures as reflective, matching their measurement model to their conceptualization is both appropriate theoretically and beneficial empirically. Indeed, little practical value results from forcing a formative model on existing, empirically successful, reflective measurement models. The simulations presented here show that the detrimental impacts supposedly caused by using reflective models for constructs that could be modeled formatively are virtually nonexistent. Instead, the results presented here suggest that more harm may result from modeling constructs as formative when reflective specifications are plausible alternatives.

References

Aguirre-Urreta, M. I., & Marakas, G. M. (2012). Revisiting bias due to construct misspecification: Different results from considering coefficients in standardized form. *MIS Quarterly*, *36*(1), 123–138.

Anderson, J. C., Gerbing, D. W., & Hunter, J. E. (1987). On the assessment of unidimensional measurement: Internal and external consistency, and overall consistency criteria. *Journal of Marketing Research*, *24*(4), 432–437.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K. A., & Ting, K. (2000). A tetrad test for causal indicators. *Psychological Methods,* *5*(1), 3–22.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.

Cadogan, J. W., & Lee, N. (2013). Improper use of endogenous formative variables. *Journal of Business Research*, *66*(2), 233–241.

Challagalla, G. N., & Shervani, T. A. (1997). A measurement model of the dimensions and types of output and behavior control: An empirical test in a salesforce context. *Journal of Business Research*, *39*(3), 159–172.

Chang, W., Park, J. E., & Chaiy, S. (2010). How does CRM technology transform into organizational performance? A mediating role of marketing capability. *Journal of Business Research*, *63*(8), 849–855.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, *35*(2), 335–358.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*(4), 263–282.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–174.

Fornell, C., Rhee, B. D., & Yi, Y. (1991). Direct regression, reverse regression, and covariance structure analysis. *Marketing Letters*, *2*(3), 309–320.

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research*, *61*(12), 1229–1237.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods, 6*(3), 258–269.

Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks: SAGE.

Hayduk, L. A. (1987). *Structural equation modelling with LISREL: Essentials and advances*. Baltimore: The John Hopkins University Press.

Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analyses. *Sociological Methods and Research, 1*(2), 147–173.

Howell, R. D. (2013). Conceptual clarity in measurement–Constructs, composites, and causes: a commentary on Lee, Cadogan and Chamberlain. *AMS Review*, *3*(1), 18–23.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods, 12*(2), 205–218.

Jarvis, C. B., MacKenzie, S. B., & 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), 199–218.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2012). The negative consequences of measurement model misspecification: A response to Aguirre-Urreta and Marakas. *MIS Quarterly*, *36*(1), 139–146.

Kara, A., Spillan, J. E., & DeShields, Jr., O. W. (2004). An empirical investigation of the link between market orientation and business performance in nonprofit service providers. *Journal of Marketing Theory and Practice*, *12*(2), 59–72.

Kenny, D. A., Kashy, D., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 233–265). New York: McGraw-Hill.

Lee, N., & Cadogan, J. W. (2013). Problems with formative and higher-order reflective variables. *Journal of Business Research*, *66*(2), 242–247.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114*(3), 533–541.

MacKenzie, S. B., Podsakoff, P. M., & 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), 710–730.

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning (multivariate applications series)*. New York: Routledge.

Pedhazur, E., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale: Erlbaum.

Petter, S., Rai, A., & Straub, D. (2012). The critical importance of construct measurement specification: A response to Aguirre-Urreta and Marakas. *MIS Quarterly, 36*(1), 147–156.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, *31*(4), 623–656.

Podsakoff, N. P., Shen, W., & Podsakoff, P. M. (2006). The role of formative measurement models in strategic management research: Review, critique, and implications for future research. In D. J. Ketchen, & D. D. Bergh (Eds.), *Research methodology in strategy and management* (pp. 197–252). Oxford: Elsevier.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research*, *41*(3), 293–305.

Rigopoulou, I., Theodosiou, M., Katsikea, E., & Perdikis, N. (2012). Information control, role perceptions, and work outcomes of boundary-spanning frontline managers. *Journal of Business Research*, *65*(5), 626–633.

Temme, D., Diamantopoulos, A., & Pfegfeidel, V. (2014). Specifying formatively-measured constructs in endogenous positions in structural equation models: Caveats and guidelines for researchers. *International Journal of Research in Marketing*, *31*(3), 309–316.

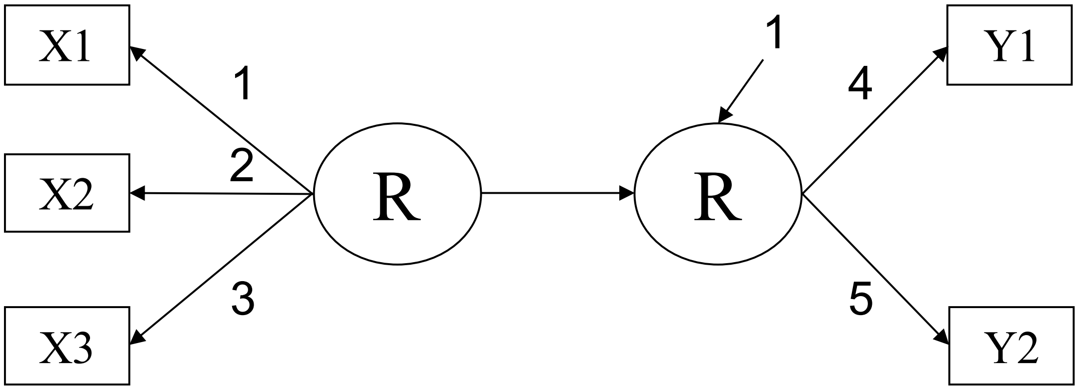
Treiblmaier, H., Bentler, P. M., & Mair, P. (2011). Formative constructs implemented via common factors. *Structural Equation Modeling*, *18*(1), 1–17.

Wilcox, J. B, Howell, R. D, & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, *61*(12), 1219–1228.

Footnote

1. Models containing formatively-measured endogenous variables are problematic (e.g., Cadogan & Lee, 2013; Temme, Diamantopoulos, & Pfegfeidel, 2014). The current research includes them to allow comparisons with prior simulations.

A. Reflective model



B. Formative model with indicator correlations accounted for by reflective loadings

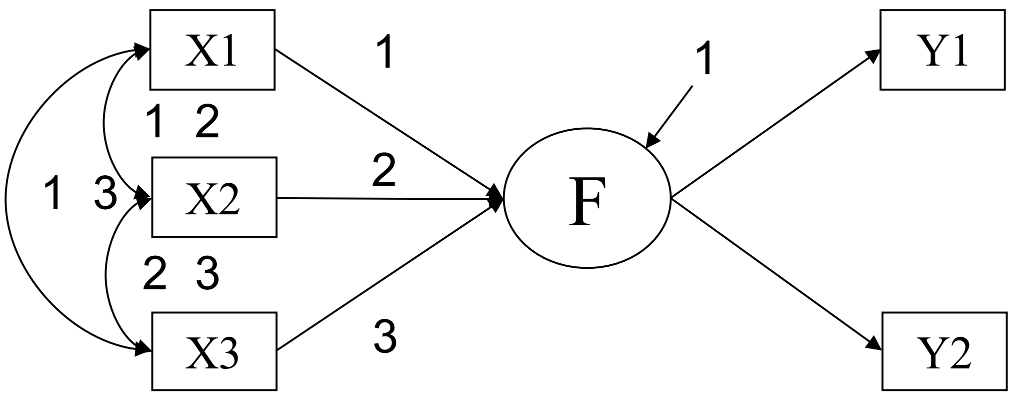


Fig. 1. Reflective and formative models.

Fig. 2A. Model 1: Exogenous Formative Construct

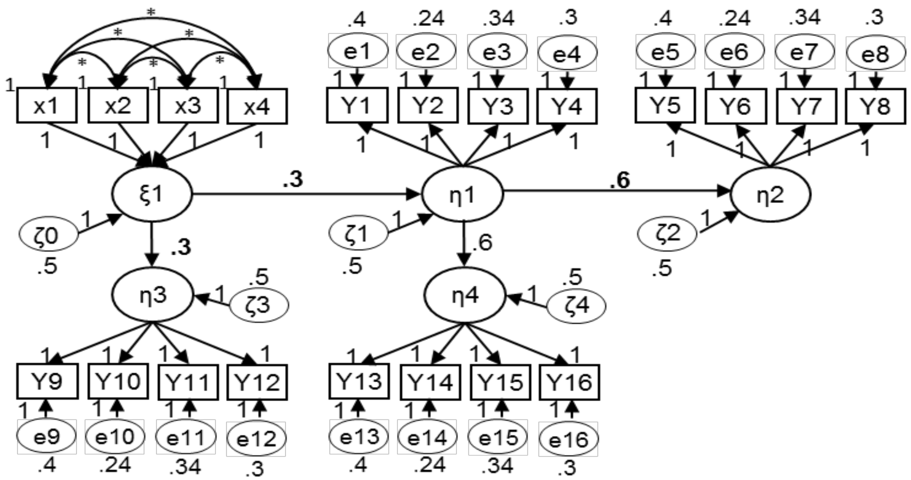


Fig. 2B. Model 2: Endogenous Formative Construct

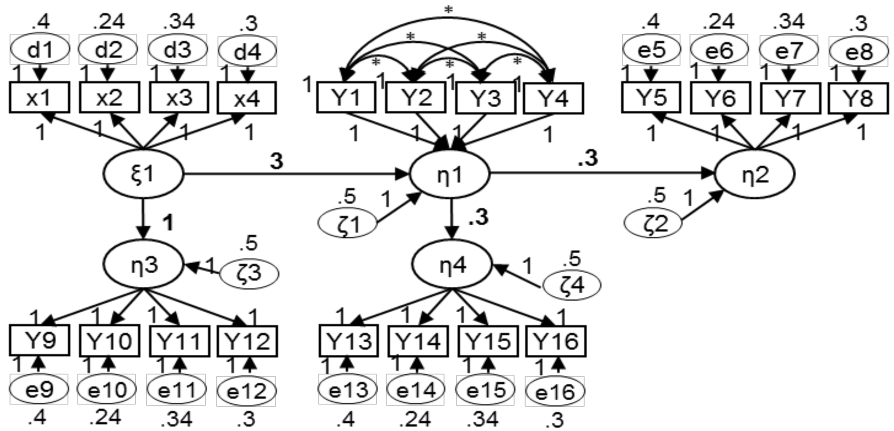


Fig. 2C. Alternative Model: Exogenous and Endogenous Reflective Constructs

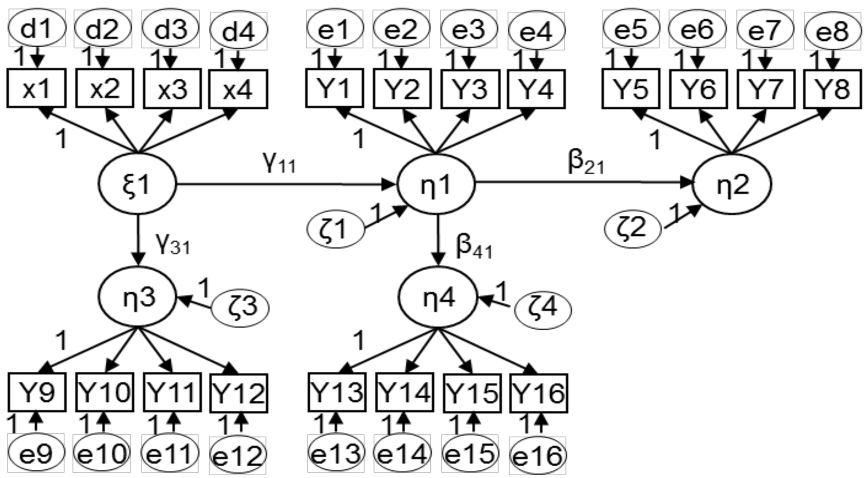


Fig. 2. Jarvis, MacKenzie, and Podsakoff's (2003) simulation study.

Table 1
Example reflective and formative specifications.

CRM Technology Use					Information Control					Market Orientation				
Indicator	RS	RU	FS	FU	Indicator	RS	RU	FS	FU	Indicator	RS	RU	FS	FU
Sales	0.79	1.00	0.25	1.00	Output	0.76	1.00	0.21	0.75	Intelligence	0.64	1.00	-0.05	-0.12
Support	(12.98)	(n.a.)	(2.86)	(n.a.)	Information	(15.17)	(n.a.)	(3.11)	(2.14)	Generation	(7.91)	(n.a.)	(-0.49)	(-0.51)
Service	0.77	0.98	0.13	0.53	Capability	0.84	1.11	0.28	1.00	Intelligence	0.91	1.42	0.42	1.00
Support	(12.69)	(11.56)	(1.49)	(1.27)	Information	(17.46)	(14.92)	(3.73)	(n.a.)	Dissemination	(11.82)	(7.20)	(3.65)	(n.a.)
Analysis	0.82	1.04	0.10	0.40	Activity	0.89	1.17	0.20	0.73	Responsive-	0.72	1.12	0.09	0.23
Support	(13.68)	(12.26)	(1.08)	(0.90)	Information	(18.97)	(15.46)	(2.54)	(1.69)	ness	(9.00)	(7.12)	(0.84)	(0.73)
Data	0.83	1.06	0.15	0.61										
Support	(14.11)	(12.53)	(1.62)	(1.23)										
Outcome					**Outcome**					**Outcome**				
Marketing	0.92	1.00	0.92	0.23	Supervisor	0.94	1.00	-0.95	-0.26	Funding	0.84	1.00	0.84	0.35
Capability 1	(n.a.)	(n.a.)	(15.74)	(2.78)	Ambiguity	(n.a.)	(n.a.)	(-16.06)	(-3.70)	Analysis	(n.a.)	(n.a.)	(9.49)	(3.41)
Marketing	0.96	1.04	0.96	0.24	Job	0.66	0.70	-0.65	-0.18	Funding	0.69	0.82	0.68	0.29
Capability 2	(14.92)	(14.92)	(16.96)	(2.81)	Ambiguity	(9.09)	(9.04)	(-11.36)	(-3.46)	RFPs	(4.87)	(5.92)	(5.55)	(2.97)
										Periodic	0.49	0.58	0.49	0.21
										Fundraising	(4.87)	(4.87)	(5.55)	(2.97)
Direct effect	0.57	0.67				-0.65	-0.80				0.48	0.63		
	(7.75)	(7.48)				(-11.40)	(-10.61)				(4.87)	(4.47)		
SMC	0.33	0.33	0.29	0.29		0.42	0.42	0.37	0.37		0.23	0.23	0.21	0.21
Chi-square	14.92	14.92	6.57	6.57		4.54	4.54	1.41	1.41		5.92	5.92	4.23	4.23
d.f.	8	8	3	3		4	4	2	2		8	8	6	6
NNFI	0.99	0.99	0.98	0.98		1.00	1.00	1.00	1.00		1.01	1.01	1.01	1.01
RMSEA	0.06	0.06	0.08	0.08		0.02	0.02	0.00	0.00		0.00	0.00	0.00	0.00
SRMR	0.02	0.02	0.01	0.01		0.01	0.01	0.01	0.01		0.04	0.04	0.04	0.04

RS and FS: standardized; RU and FU: unstandardized. Values in parentheses are t-values.

Table 2
Simulation results with all reflectively-measured constructs.

	Population	Raw	Std.	Raw	Std.	Raw	Std.
Sample size	–	250	250	500	500	800	800
Usable replications	–	1000	1000	1000	1000	1000	1000
Fit statistics							
Chi-square	0	166.57	166.57	166.10	166.10	167.28	167.28
d.f.	166	166	166	166	166	166	166
NNFI	1	1.00	1.00	1.00	1.00	1.00	1.00
RMSEA	0	0.01	0.01	0.01	0.01	0.01	0.01
SRMR	0	0.04	0.04	0.03	0.03	0.02	0.02
Structural effects	Raw / Std.						
$\xi1 \rightarrow \eta1$	0.60 / 0.60	0.60 (6.83)	0.60 (7.48)	0.60 (9.70)	0.60 (10.62)	0.60 (12.22)	0.60 (13.37)
$\xi1 \rightarrow \eta3$	0.60 / 0.60	0.60 (6.70)	0.60 (7.31)	0.60 (9.56)	0.60 (10.44)	0.60 (12.10)	0.60 (13.21)
$\eta1 \rightarrow \eta2$	0.60 / 0.60	0.60 (6.76)	0.60 (6.76)	0.60 (9.59)	0.60 (9.59)	0.60 (12.17)	0.60 (12.17)
$\eta1 \rightarrow \eta4$	0.60 / 0.60	0.60 (6.77)	0.60 (6.77)	0.60 (9.62)	0.60 (9.62)	0.60 (12.15)	0.60 (12.15)
Reflective loadings	Raw / Std.						
$\xi1 \rightarrow X1$	1 / 0.71	1 (n.a.)	0.71 (9.43)	1 (n.a.)	0.71 (13.31)	1 (n.a.)	0.71 (16.81)
$\xi1 \rightarrow X2$	1 / 0.71	1.01 (9.43)	0.71 (9.43)	1.00 (13.31)	0.71 (13.33)	1.00 (16.81)	0.71 (16.83)
$\xi1 \rightarrow X3$	1 / 0.71	1.01 (9.43)	0.71 (9.44)	1.00 (13.33)	0.71 (13.32)	1.00 (16.83)	0.71 (16.83)
$\xi1 \rightarrow X4$	1 / 0.71	1.01 (9.44)	0.70 (9.36)	1.00 (13.32)	0.71 (13.33)	1.00 (16.83)	0.71 (16.82)
Mean number of significant coefficients	–	3.00	4.00	3.00	4.00	3.00	4.00

Raw: unstandardized; Std.: standardized. Values in parentheses are t-values.

Table 3
Simulation results with one exogenous formatively-measured construct.

	Population	Raw	Std.	Raw	Std.	Raw	Std.
Sample size	–	250	250	500	500	800	800
Usable replications	–	841	841	950	950	980	980
Fit statistics							
Chi-sq	0	160.52	160.52	160.12	160.12	161.25	161.25
d.f.	160	160	160	160	160	160	160
NNFI	1	1.00	1.00	1.00	1.00	1.00	1.00
RMSEA	0	0.01	0.01	0.01	0.01	0.01	0.01
SRMR	0	0.04	0.04	0.03	0.03	0.02	0.02
Structural effects	Raw / Std.						
$\xi1 \rightarrow \eta1$	0.12 / 0.60	0.12 (2.81)	0.61 (6.87)	0.12 (4.04)	0.61 (9.63)	0.12 (5.09)	0.60 (12.10)
$\xi1 \rightarrow \eta3$	0.12 / 0.60	0.12 (2.80)	0.61 (6.78)	0.12 (4.03)	0.60 (9.51)	0.12 (5.08)	0.60 (11.99)
$\eta1 \rightarrow \eta2$	0.60 / 0.60	0.60 (6.74)	0.60 (6.74)	0.60 (9.59)	0.60 (9.59)	0.60 (12.17)	0.60 (12.17)
$\eta1 \rightarrow \eta4$	0.60 / 0.60	0.60 (6.76)	0.60 (6.76)	0.60 (9.62)	0.60 (9.62)	0.60 (12.15)	0.60 (12.15)
Formative coefficients	Raw / Std.						
$X1 \rightarrow \xi1$	1 / 0.28	1 (n.a.)	0.28 (2.95)	1 (n.a.)	0.28 (4.22)	1 (n.a.)	0.28 (5.31)
$X2 \rightarrow \xi1$	1 / 0.28	1.37 (1.73)	0.28 (3.04)	1.10 (2.58)	0.28 (4.26)	1.06 (3.30)	0.28 (5.36)
$X3 \rightarrow \xi1$	1 / 0.28	1.31 (1.71)	0.28 (2.96)	1.09 (2.56)	0.28 (4.22)	1.05 (3.28)	0.28 (5.32)
$X4 \rightarrow \xi1$	1 / 0.28	1.31 (1.71)	0.27 (2.93)	1.08 (2.56)	0.28 (4.20)	1.05 (3.28)	0.28 (5.32)
Mean number of significant coefficients	–	0.91	3.33	2.71	3.95	2.99	4.00

Table 4
Simulation results with one endogenous formatively-measured construct.

	Population	Raw	Std.	Raw	Std.	Raw	Std.
Sample size	–	250	250	500	500	800	800
Usable replications	–	806	806	883	883	951	951
Fit statistics							
Chi-sq	0	156.74	156.74	156.14	156.14	156.92	156.92
d.f.	156	156	156	156	156	156	156
NNFI	1	1.00	1.00	1.00	1.00	1.00	1.00
RMSEA	0	0.01	0.01	0.01	0.01	0.01	0.01
SRMR	0	0.04	0.04	0.04	0.04	0.02	0.02
Structural effects	Raw / Std.						
$\xi 1 \rightarrow \eta 1$	0.94 / 0.17	1.39 (1.08)	0.17 (1.64)	1.09 (1.66)	0.17 (2.32)	1.03 (2.22)	0.17 (3.00)
$\xi 1 \rightarrow \eta 3$	0.60 / 0.60	0.60 (6.70)	0.60 (7.32)	0.60 (9.57)	0.60 (10.44)	0.60 (12.10)	0.61 (13.21)
$\eta 1 \rightarrow \eta 2$	0.11 / 0.60	0.11 (2.52)	0.61 (6.73)	0.11 (3.53)	0.60 (9.36)	0.11 (4.53)	0.60 (11.85)
$\eta 1 \rightarrow \eta 4$	0.11 / 0.60	0.11 (2.52)	0.61 (6.73)	0.11 (3.52)	0.60 (9.39)	0.11 (4.53)	0.60 (11.83)
Formative coefficients	Raw / Std.						
$Y1 \rightarrow \eta 1$	1 / 0.25	1 (n.a.)	0.25 (2.60)	1 (n.a.)	0.25 (3.63)	1 (n.a.)	0.25 (4.68)
$Y2 \rightarrow \eta 1$	1 / 0.25	1.38 (1.48)	0.24 (2.55)	1.14 (2.23)	0.25 (3.68)	1.05 (2.89)	0.25 (4.58)
$Y3 \rightarrow \eta 1$	1 / 0.25	1.37 (1.51)	0.25 (2.60)	1.13 (2.23)	0.25 (3.67)	1.07 (2.91)	0.25 (4.68)
$Y4 \rightarrow \eta 1$	1 / 0.25	1.37 (1.48)	0.24 (2.54)	1.14 (2.24)	0.25 (3.70)	1.07 (2.91)	0.25 (4.68)
Mean number of significant coefficients	–	0.55	2.87	2.20	3.81	2.93	3.99