Journal of Business Research 69 (2016) 3177–3185

Contents lists available at ScienceDirect

Journal of Business Research

Comparing reflective and formative measures: New insights from relevant simulations☆

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

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form 15 December 2014
Accepted 31 December 2014
Available online 30 January 2016

Keywords:
Construct specification
Formative measurement
Reflective measurement
Monte Carlo simulation

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 mis-specified 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.

> "*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 are 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 overlook 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

☆ The authors are listed in alphabetical order.
* Corresponding author. Tel.: +1 309 438 8429; fax: +1 309 438 3508.
E-mail 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.

http://dx.doi.org/10.1016/j.jbusres.2015.12.006
0148-2963/

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