

Reconsidering Formative Measurement

Roy D. Howell
Texas Tech University

Einar Breivik
Norwegian School of Economics and Business Administration

James B. Wilcox
Texas Tech University

The relationship between observable responses and the latent constructs they are purported to measure has received considerable attention recently, with particular focus on what has become known as formative measurement. This alternative to reflective measurement in the area of theory-testing research is examined in the context of the potential for interpretational confounding and a construct's ability to function as a point variable within a larger model. Although these issues have been addressed in the traditional reflective measurement context, the authors suggest that they are particularly relevant in evaluating formative measurement models. On the basis of this analysis, the authors conclude that formative measurement is not an equally attractive alternative to reflective measurement and that whenever possible, in developing new measures or choosing among alternative existing measures, researchers should opt for reflective measurement. In addition, the authors provide guidelines for researchers dealing with existing formative measures.

Keywords: formative, reflective, measurement, structural equation modeling

The examination of social science theories has traditionally emphasized specifying and testing relationships among latent variables or theoretic concepts. Recently, more attention has been focused on the nature of relationships between constructs and their measures. Because these relationships constitute an auxiliary theory that links abstract or latent constructs with observable and measurable empirical phenomena (Meehl, 1990), this work is important. Citing Blalock (1971), Edwards and Bagozzi (2000) noted that, "Without this auxiliary theory, the mapping of theoretic constructs onto empirical phenomena is ambiguous, and theories cannot be empirically tested" (p. 155).

The most common auxiliary measurement theory underlying measurement in the social sciences has its basis in classical test theory and the factor analytic perspective, wherein observable indicators are reflective effects of latent

constructs. As noted by Bollen (2002), "Nearly all measurement in psychology and the other social sciences assumes effect indicators" (p. 616). An alternative conceptualization wherein observable indicators are modeled as the cause of latent constructs has also been offered and investigated (Blalock, 1964; Bollen & Lennox, 1991; Cohen, Cohen, Teresi, Marchi, & Velez, 1990; Edwards & Bagozzi, 2000; Fornell & Bookstein, 1982; Heise, 1972; Howell, 1987; Law, Wong, & Mobley, 1998; MacCallum & Browne, 1993; Podsakoff, MacKenzie, Podsakoff, & Lee, 2003). We refer to these alternative views of the direction of the relationship between latent constructs and their associated observables as reflective and formative measurement, respectively.¹

The focus of discussions regarding reflective and formative measurement has been primarily on identification and estimation issues (Blalock, 1971; Bollen & Lennox, 1991; Heise, 1972). For example, MacCallum and Browne (1993)

Roy D. Howell and James B. Wilcox, Marketing Area, Rawls College of Business, Texas Tech University; Einar Breivik, Department of Strategy Management, Norwegian School of Economics and Business Administration, Bergen, Norway.

Roy D. Howell was supported by the J. B. Hoskins Professorship in Marketing, and James B. Wilcox was supported by the Alumni Professorship in Marketing.

Correspondence concerning this article should be addressed to Roy D. Howell, Marketing Area, Rawls College of Business, Texas Tech University, Lubbock, TX 79409-2101. E-mail: roy.howell@ttu.edu

¹ Blalock (1964) referred to reflective measures as "effect indicators" and formative measures as "cause indicators." Bollen and Lennox (1991) followed in that tradition. Cohen et al. (1990) used the term "latent variable" for the reflective model and the term "emergent variable" for constructs based on the formative model. MacCallum and Browne (1993) labeled constructs based on the formative model as "composite" constructs with "formative measures," reserving the term "indicators" for observables in a reflective model.

showed how formative indicators can be used in structural equation modeling. Edwards and Bagozzi (2000, p. 156) addressed the conditions under which measures should be modeled as formative or reflective and provided guidelines drawn from the literature on causality for “specifying the relationship between a construct and a measure in terms of (a) direction (i.e., whether a construct causes or is caused by its measures) and (b) structure (i.e., whether the relationship is direct, indirect, spurious or unanalyzed).” Bollen and Ting (2000) proposed an empirical test based on model-implied vanishing tetrads to determine whether indicators are more likely to be formative or reflective.

Taken as a group, the works of Bollen and Lennox (1991), MacCallum and Browne (1993), Edwards and Bagozzi (2000), and Bollen and Ting (2000) have given researchers conceptual and methodological tools for dealing with formative observable indicators. This body of research has shown that formative measurement can be dealt with. It does not directly address a perhaps more important question, however: Should formative measurement be used?

Researchers in a number of disciplines are now opting to depart from the dominant reflective measurement tradition in the social sciences, choosing instead to develop and use formative measures. For example, Diamantopoulos and Winklhofer (2001) presented guidelines for developing a formative measure that parallel DeVellis’s (1991) paradigm for scale development under the reflective model. Similarly, Jarvis, MacKenzie, and Podsakoff (2003) suggested guidelines for the development of formative measures. Law et al. (1998) discussed the development and analysis of constructs under what they term the “aggregate model,” which corresponds to our conceptualization of formative measurement. Diamantopoulos and Winklhofer seemed to implicitly assume that a construct is either inherently formative or reflective, and, in the case that the construct is formative, they outlined a set of procedures for determining and refining its indicators. Similarly, Jarvis et al. and Podsakoff et al. (2003) classified constructs as either formative or reflective. As we discuss below, we believe that, in general, constructs can be measured either formatively or reflectively. Prior to designing a measurement instrument, the researcher may have a choice as to which type of measurement model to use, either in the measure development process or in the choice among existing measures, some of which may be either formative or reflective (see, e.g., Law et al., 1998). Our first question is as follows: Given a choice between developing a formative or reflective measure or between an existing formative or reflective measure, which should be preferred?

To address this question, we briefly discuss two forms of formative measurement models in contrast to the traditional reflective model, followed by a discussion of the philosophy underlying constructs and measures. We then examine the issues of interpretational confounding and the ability of a

construct to function as a point variable or unified entity (Burt, 1976) in the context of formative measurement. These points are further addressed by means of an example, using MacCallum and Browne’s (1993) specification for the inclusion of formative measurement models in the context of structural equation models. This is followed by a discussion of the claim that constructs are inherently formative or reflective. Finally, suggestions for dealing with formative measures are provided.

Measurement Models

The traditional reflective measurement model is

$$x_i = \lambda_i \eta + \varepsilon_i, \quad (1)$$

where i subscripts the indicators, λ_i refers to the loading of the i th indicator on the latent trait η , and ε_i represents the uniqueness and random error for the i th indicator. The observable variables are conceptualized as a function of a latent variable. Variation in the latent variable conceptually precedes variation in its indicants. This general conceptualization is consistent with the true score in classical test theory (Lord & Novick, 1968), the latent θ in item response theory (e.g., Samejima, 1969), common factor analysis (Spearman, 1904; Thurstone, 1947), and confirmatory factor analysis (CFA; Jöreskog & Sörbom, 1993). In the context of CFA, if there are three x_i , then the λ_i can be estimated and the measurement model in Equation 1 is just identified, and if there are four or more indicators, then the fit of Equation 1 to the data can also be assessed. We refer to this model as a purely epistemic baseline measurement model because it can be estimated from a set of observables for a single latent variable without reference to other variables or constructs.

In contrast, one conceptualization of the formative measurement model is

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_n x_n, \quad (2)$$

where γ_i is a parameter reflecting the contribution of x_i to the latent construct η . Here, the latent construct is viewed as a function of the observables. In contrast to λ_i in Equation 1, γ_i in Equation 2 cannot be estimated without reference to some other variable(s) or construct(s).² Several approaches can and have been used to estimate the γ_i in Equation 2, including regressing a dependent variable y on the x_i (ordi-

² Principal components (PC) analysis estimates the γ_i such that the variance of η is maximized. Thus, the weights in PC analysis can be estimated without reference to a dependent variable (as is the case for the common factor model in Equation 1 applied to a single construct). However, for a single (first) PC to account for the x_i they must be intercorrelated, which is not a requirement for formative indicators.

nary least squares or OLS) wherein η is equivalent to the predicted value of y , partial least squares (PLS), canonical correlation (CC), and redundancy analysis (RA; Van den Wollenberg, 1977). Although this is a measurement model in the context of formative measurement, it by necessity requires a dependent variable or construct for estimation. Equations in which a set of variables predict a variable or construct are conventionally referred to as structural or as part of a structural model (Anderson & Gerbing, 1988). Consistent with this usage and to distinguish this model from the purely epistemic model in Equation 1, we refer to this model and its parameters as structural.

For a given set of indicators, the weights in OLS, PLS, CC, and RA as opposed to being epistemic (measurement) are entirely structural. That is, they depend on the context in which they are estimated—the dependent variable(s) in the model. As recognized by Heise (1972), a construct measured formatively is not just a composite of its measures; rather, “it is the composite that best predicts the dependent variable in the analysis. . . . Thus the meaning of the latent construct is as much a function of the dependent variable as it is a function of its indicators” (p. 160).

Another formative specification advocated by Bollen and Lennox (1991), Diamantopoulos and Winklhofer (2001), Jarvis et al. (2003), and Law et al. (1998) and explored by MacCallum and Browne (1993) includes an error term, ζ , at the construct level:

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_n x_n + \zeta. \quad (3)$$

As in Equation 2, taken alone the parameters in Equation 3 are not identified. For Equation 3 to be identified, however, η must relate to at least two endogenous variables or constructs with no direct path between those endogenous constructs and uncorrelated error terms (MacCallum & Browne, 1993). This puts the model in Equation 3 in the same class as OLS, PLS, CC, and RA—the γ_i , and thus η , are dependent on the endogenous variables included in the model.

Constructs and Their Measurement

We consider a measure to be an observed score (observable) generated via self-report, observation, or some other means. It is quantified datum, taken as an empirical analog to a construct (Edwards & Bagozzi, 2000, p. 156; Lord & Novick, 1968). We define a *construct* as a conceptual term used to describe a phenomenon of theoretical interest (Cronbach & Meehl, 1955; Edwards & Bagozzi, 2000). Although constructs refer to reality, they remain “elements of scientific discourse that serve as verbal surrogates for phenomena of interest” (Edwards & Bagozzi, 2000, p. 157). Structural equation models, as tools in theory testing, are specified such that the theoretic constructs of interest have substan-

tive interpretations prior to and apart from the estimation of parameters linking them to their respective measures. Responses to the indicators comprising the x_i in Equation 1 are thought to vary as a function of the latent variable. This implies that the latent variable exists apart from its measurement. Although the researcher may have conceptualized a formatively measured latent construct on the basis of its theoretic definition and chosen formative measures accordingly, Equations 2 and 3 make no direct claim that the construct or latent variable exists independent of the measurements (Borsboom, Mellenbergh, & van Heerden, 2003).

Thus, the models in Equations 2 and 3 differ from the reflective model in Equation 1, both conceptually and psychometrically. As noted by Borsboom et al. (2003), “The realist interpretation of a latent variable implies a reflective model, whereas constructivist, operationalist, or instrumentalist interpretations are more compatible with a formative model” (p. 209).

Interpretational Confounding

Do the differences between formative and reflective models with respect to formulation and ontology matter? In the context of reflective measurement, Burt (1976), following Hempel (1970, pp. 654–666), distinguished between the nominal meaning and the empirical meaning of a construct. A construct’s nominal meaning is that meaning assigned without reference to empirical information. That is, it is the inherent definitional nature of the construct that forms the basis for hypothesizing linkages with other constructs, developing observable indicators, and so forth. A construct’s empirical meaning derives from its relations to one or more observed variables. These may be measures of the construct itself (epistemic) or relationships to observable measures of other constructs in a model (structural). Burt discussed the problem he referred to as “interpretational confounding:”

The problem, here defined as “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. Inferences based on the unobserved variable then become ambiguous and need not be consistent across separate models. (Burt, 1976, p. 4)

That is, to the extent that the nominal and empirical meanings of a construct diverge, there is an issue of interpretational confounding. One would expect the observables chosen to measure a construct to be relatively congruent with the nominal meaning or definition of the construct, such that after estimating epistemic relationships the empirical meaning of the construct remains close to its nominal meaning. Burt discussed this problem in the context of reflective models, suggesting that when there are few reflective indicators and/or the epistemic relationships of those indicators to their associated construct are weak relative to structural

parameters in a structural equation model, the meaning of the construct can be as much (or more) a function of its relationship(s) to other constructs in the model as a function of its measures. This is problematic. The name of the construct may remain the same (consistent with its nominal meaning and hopefully consistent with the content of its indicants), but the empirical meaning of the construct (within the context of its structural relationships within the larger model) has changed.

If this is a problem with reflective measurement models, then what of formative models? Equations 2 and 3 have no strictly epistemic relationships. The nature (empirical meaning) of the latent construct depends on the dependent variables or constructs included in the model. Formative measurement models are subject to interpretational confounding (as are reflective measurement models), but because in formative measurement there are no epistemic relationships available to estimate the measurement equations without reference to structural relationships (i.e., a dependent variable in Equation 2 or at least two dependent variables in Equation 3), the problem is difficult to avoid, apart from predetermining the γ_i in some fashion.

For example, Heise (1972) argued that the socioeconomic status (SES) construct is caused by education, income, and occupational prestige. If in a given model (based on the constructs SES predicts in that particular model) SES is almost entirely a function of income with little if any contribution from education and occupation, then interpretational confounding has become a problem. The construct labeled SES in the model is really just income, not the SES defined without reference to parameter estimates. This interpretation comes from the relative magnitude of the paths from the formative indicators to the formatively measured latent construct (the γ coefficients in Equations 2 or 3). Now suppose different dependent constructs are modeled that are closely associated with education but weakly, if at all, related to income. The empirical realization of SES is still inconsistent with its nominal definition, and, in addition, SES in the second model is not the same as SES in the first model.

Reflective measurement models as shown in Equation 1 are also model dependent, in that they depend on other variables—constructs included in the model. For example, indicator-specific variance in Equation 1 for a given indicator could be shared variance with an indicator of a different construct in the context of a larger model, resulting in correlated measurement error terms. Similarly, a reflective indicator may cross-load on another factor when included in a larger model. As we discuss below in more detail, in a structural equation model constructs dependent on a reflectively measured construct are mathematically equivalent to additional (perhaps second-order) measures of the construct. No, reflective measures are not immune from con-

tamination and potential interpretational confounding when estimated in a larger structural equation model.

However, reflective models have epistemic relationships that exist independently from structural relationships. Burt (1976) suggested estimating the epistemic relationships in a reflective model either construct by construct or within what we would today call a “measurement model” (Anderson & Gerbing, 1988) within blocks of substantively dissimilar constructs and subsequently fixing them to those values when estimating the structural parameters in the larger model. To the degree that measurement parameters for a given construct freely estimated in an unrestricted larger model differ substantially from measurement parameters fixed to the parameters estimated in the within-blocks or construct-by-construct measurement model, interpretational confounding might be a problem. Neither set of measurement coefficients is necessarily correct in any absolute sense, and the estimates from the within-block measurement models are indeed limited information. Which to choose? Burt suggested a solution:

First, if a set of “core” concepts are to be compared across multiple alternative structural equation models involving substantively different concepts, then it makes sense to hold the interpretations of those core concepts constant across models being compared. The usual comparisons of alternative restrictions on structural parameters and error parameters can be pursued, however, they are pursued subject to the core concepts having their interpretations fixed. Second, if the point variability of one or more unobserved variables is suspect, then it makes sense to have statistical tests available which demonstrate the degree to which unobserved variables in each substantively distinct block operate as point variables within the context of a structural equations model. (Burt, 1976, p. 20)

This is arguably an extreme position. Anderson and Gerbing (1988) have argued for estimating the measurement model separately from the (more restrictive) structural model (without fixing the parameters), whereas Fornell and Yi (1992) have pointed out that one of the advantages of structural equation modeling is the simultaneous estimation of measurement and structural parameters. The debate continues (see, e.g., *Structural Equation Modeling*, 2000). Although we are not inclined to enter this debate, its relevance is apparent. With reflective measures, one can examine changes to the measurement parameters for a given construct (Equation 1) as other constructs are added to the model and thus assess the degree of potential interpretational confounding. This is not an option for constructs measured formatively with estimated parameters, because there is no strictly measurement baseline from which to depart. In the context of formative measurement, it is up to the researcher to compare the empirical realization of the construct with its nominal definition, assessing the degree to which the construct as estimated corresponds to the construct as defined.

To the first point in Burt’s (1976, p. 20) quote above (a set

of core concepts compared across models involving other substantively different concepts), we would add not only the comparison of alternative models but also the cumulative nature of research across studies. If a construct named “A” in one study is substantively different from a construct named “A” in another study, then accumulation of knowledge regarding a construct is rendered meaningless or impossible because the construct in one study is incommensurable with a different construct, but with the same name, in another study. This is a point made clear by Blalock (1982) in his chapter entitled “The Comparability of Measurement.” Blalock observed that,

Whenever measurement comparability is in doubt, so is the issue of the generalizability of the underlying theory. . . . If the theory succeeds in one setting but fails in another, and if measurement comparability is in doubt, one will be in the unfortunate position of not knowing whether the theory needs to be modified, whether the reason for the differences lies in the measurement-conceptualization process, or both. (Blalock, 1982, p. 30)

As we discuss later, Bradley and Corwyn (2002) alluded to this problem in attempting to review the impact of SES on child development.

Although both CFA and item response theory approaches have been developed to assess measurement equivalence (cf. Raju, Laffitte, & Byrne, 2002), they deal more specifically with epistemic relationships in the reflective model across groups of subjects rather than with the interpretational confounding of structural relationships from model to model. We note that reflective measures, conceived as outcomes of a construct that exists apart from its measurement, are conceptually interchangeable and replaceable. Dropping or replacing a reflective observable does not change the nature of the concept (Bollen & Lennox, 1991). Under the formative model, the nature of the construct must change as its indicators, dependent variables, and their relationships change. Although it is beyond the scope of our discussion to deal directly with the assessment of measurement comparability across studies, we do believe it is a topic worthy of more consideration than it has gained to date.

Point Variables

The second point in Burt’s (1976) solution deals with the notion of a point variable. In a structural equation model, the covariances of indicators of η in a reflective model (the x_i in Equation 1) with other latent variables or their indicators should be proportional to their epistemic correlations (λ) with η . That is, following Burt, if we designate Q as the set of indicants of η (as in Equation 1) and let R represent the set of indicators of other constructs in the model, then η operates as a point variable if and only if $\sigma_{ik}/\lambda_{i\eta} = \sigma_{jk}/\lambda_{j\eta}$ for all i, j in set Q and all k in set R . Reduced to the indicator level, this implies the rule of external consistency discussed

by Anderson and Gerbing (1982): $r_{x_1y_j}/r_{x_2y_j} = r_{x_1y_k}/r_{x_2y_k}$, where r is a correlation, x_1 and x_2 are two measures of a given construct, and y_j and y_k are measures of another construct. That is, a variable functions as a unitary entity or point variable in a model when the indicants correlate with other constructs in proportion to their (epistemic) correlation with their own construct. In a structural equation model, the concept of point variability—in particular the external consistency requirement—applies to both reflectively and formatively measured constructs. Because reflective measures are conceptualized as having a common cause, they can be expected to intercorrelate and there is reason to believe that they might therefore relate similarly to other constructs. Internal consistency among formative indicators is not applicable, however, so external consistency becomes the sole condition for assessing the degree to which a formatively measured construct functions as a unitary entity. However, previous researchers have noted that formative indicators need not covary (Bollen & Lennox, 1991; Jarvis et al., 2003), need not have the same nomologic net (Jarvis et al., 2003), and hence “are not required to have the same antecedents and consequences” (Jarvis et al., 2003, p. 203). This seems directly at odds with the concepts of point variability and external consistency. We illustrate these points with a few simple examples.

Examples

Consider the model depicted in Figure 1. One could interpret this model as a multiple indicator–multiple cause model (Bagozzi, 1980; Jöreskog & Goldberger, 1975). That is, η_2 and η_3 are interpreted as two second-order indicants of η_1 and the x_i are taken as direct causes. The model in Figure 1 could also be interpreted as a single construct with two second-order reflective indicants and four formative indicators or as a formatively measured construct that influences two different constructs (in which ζ_1 is interpreted as a measurement error term). These interpretations are empirically indistinguishable. They all produce identical parameter estimates. Just what, then, is η_1 (and thus ζ_1) in Figure 1? Should it be named on the basis of the content of η_2 and η_3 interpreted as (observable variable or first-order factor) reflective indicants or should it be named on the basis of the content of its formative indicants? In either case, η_1 is indeed a function of the covariance of (the common factor underlying) η_2 and η_3 , and the γ s are estimated to explain this common variance. In formative measurement, then, η_1 is named on the basis of the content of the x_i but the γ s relating x_i to η_1 are determined by the common content of the η_2 and η_3 . Similarly, ζ_1 is interpreted as measurement error in the formative construct (interpreted in terms of the x_i) but is determined by the ability of the x_i to explain the common variance in the predicted η s. In this case, the γ s represent the supposed epistemic (measurement) relation-

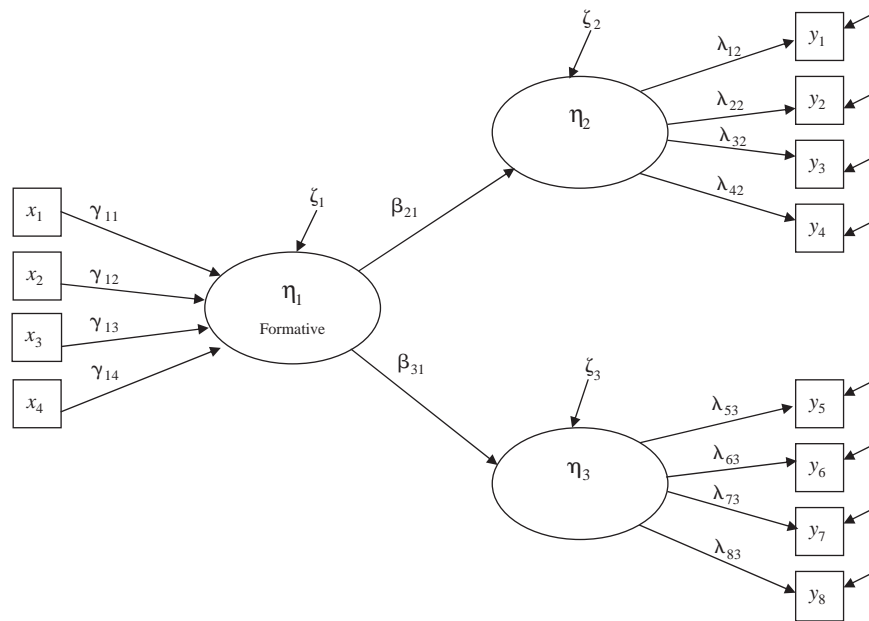


Figure 1. A typical formative measurement model, with the formatively measured construct η_1 measured with four indicants (x_1 – x_4) and including what is commonly construed as a measurement error term at the construct level (ζ_1). This corresponds with Equation 3 in the text. The formatively measured construct affects two endogenous constructs (η_2 and η_3), as is necessary for identification. η_2 and η_3 are reflectively measured with four indicants each. The coefficients in the model correspond to the rows in Table 1 for three hypothetical examples. Error terms for the reflective indicants are omitted for clarity of presentation. x_1 – x_4 may be correlated.

ships consistent with the nominal meaning of the construct, whereas its empirical meaning is derived from the (apparently) structural relationships—a case of potential interpretational confounding.

To the extent that η_2 and η_3 are content valid, reliable, unidimensional reflective measures consistent with the nominal definition of η_1 , thinking of the model in Figure 1 as a single construct with two reflective and three formative indicators may be appropriate and the naming fallacy described by Cliff (1983) and discussed by Bollen and Ting (2000) is not a problem. In the case in which the higher order factor underlying the endogenous constructs has conceptual meaning (note that the model implies that η_2 and η_3 are unrelated given η_1 , such that any covariance between them is due to their common cause), perhaps renaming η_1 in Figure 1 to reflect this meaning and treating the x_i as its exogenous predictors would be appropriate.

However, what of the situation in which η_2 and η_3 are substantively different and largely unrelated alternatives (apart from their having a common predictor) that might be predicted by η_1 ? As noted previously, formative indicants need not covary, need not have the same nomologic net, and hence “are not required to have the same antecedents and consequences” (Jarvis et al., 2003, p. 203). Indeed, referring to indicators used to (formatively) measure charismatic

leadership, Podsakoff et al. (2003) suggested that, “Moreover, the antecedents and consequences of these diverse forms of leader behavior would not necessarily be expected to be the same” (p. 650). If this is the case, then how can they be summarized as a single, meaningful construct with common antecedents and consequences? The model in Figure 1 implies that the effects of x_i are completely mediated (Baron & Kenny, 1986) by η_1 . The x_i as formative indicators are not required to have the same consequence, and thus x_1 may relate strongly to η_2 and weakly to η_3 , whereas x_3 , for example, may relate strongly to η_3 and weakly or negatively with η_2 —yet their effects are hypothesized to flow through a single construct (η_1) with some unitary interpretability. That is, η_1 is “something” and it is supposed to be the same “something” in its relationship with both η_2 and η_3 , yet each x is connected to η_1 through only one γ . In the case in which the x_i relate differently to the included η_i , (a) there will be substantial lack of fit in the model and (b) it will be difficult to interpret the meaning of η_1 either in terms of the (potentially nonexistent) covariance of the η s or the formative indicators.

To illustrate these points, we consider three hypothetical examples consistent with Figure 1. Just as Heise (1972) argued that the SES construct is caused by education, income, and occupational prestige, assume that η_1 (named

“Formative”) in Figure 1, per its nominal definition, is caused by x_1 – x_4 .

Example 1

Following MacCallum and Browne’s (1993) formulation, and using the correlations in Table 2, we obtain the parameter estimates and model fit shown in the first column of Table 1. The model fits well, and indeed “Formative” is consistent with its nominal meaning as being caused by x_1 – x_4 . The γ coefficients are significant, and each of the x_i contributes to the construct. The error term associated with the formative construct is .37 (standardized), consistent with a squared multiple correlation for η_1 of .63. (The unstandardized coefficients are scaled relative to the contribution of the reference indicator fixed to 1.0 to define the metric of the formatively measured construct. Changing the indicator associated with the fixed γ does not change the relative contribution of each indicator, and the standardized solution shows the same relative influence and is unaffected by the choice of indicator used.) This example (lack of interpretational confounding, good model fit) results from a fortuitous choice of dependent constructs rather than any inherent property of the formative measure, as we illustrate below.

Table 1
Estimation Results Corresponding to Examples 1–3

Parameter	Example 1	Example 2	Example 3
γ_{11}	0.91 (.28)	9.31 (.57)	1.00 (.36)
γ_{12}	1.01 (.31)	8.71 (.53)	1.00 (.36)
γ_{13}	1.12 (.35)	1.00 (.06) ^a	1.00 (.36)
γ_{14}^b	1.00 (.31)	1.00 (.06)	1.00 (.36)
β_{21}	0.24 (.81)	0.05 (.92)	0.18 (.53)
β_{31}	0.21 (.70)	0.04 (.69)	0.18 (.53)
ζ_1	3.81 (.37)	54.8 (.21) ^a	1.21 (.16) ^a
ζ_2	0.31 (.34)	0.13 (.15)	0.65 (.72)
ζ_3	0.46 (.51)	0.48 (.53)	0.65 (.72)
λ_{12} – λ_{83}^c	1.00 (.95)	1.00 (.95)	1.00 (.95)
$\chi^2(46)$	52.86 ($p = .23$)	55.25 ($p = .16$)	147.84 ($p < .00$)
RMSEA	.014	.017	.063

Note. Unstandardized coefficients are followed by standardized coefficients in parentheses. The coefficients correspond to Figure 1, and the input matrices are shown in Tables 2–4. All coefficients are significant ($p < .05$) unless otherwise indicated. The sample size is assumed at 432. Interpretational confounding is suggested by a comparison of the γ coefficients relating the formative indicators to the formatively measured construct in Examples 1 and 2. Failure of the formatively measured construct to function as a point variable (as indicated by lack of fit) is illustrated in Example 3. RMSEA = root-mean-square error of approximation.

^a Not significant ($p > .05$). ^b Coefficient fixed to 1.0 in the unstandardized solution for scaling; the other three γ coefficients are scaled in proportion to the reference indicator’s impact on the formatively measured construct (η_1). The standardized solution does not change when other reference indicators are used. ^c λ_{12} and λ_{53} fixed to 1.0 in the unstandardized solution.

Example 2—Interpretational Confounding

What happens if instead of the correlations in Table 2, one observes the correlations in Table 3, with different dependent constructs (either within a given data set or in a different study using the same formatively measured construct)? With these data, x_1 and x_2 are more substantially related to the reflectively measured endogenous constructs than are x_3 and x_4 . This is not inconsistent with the nature of formative measures, which need not have the same consequences. Although model fit remains good, as can be seen from the second column in Table 1, the empirical realization of “Formative” is strongly determined by its indicators’ relationships to the dependent constructs. Standardized estimates of γ_{11} and γ_{12} are .57 and .53, respectively, whereas the estimates of γ_{13} and γ_{14} are each .06 in the standardized metric. The formatively measured construct is empirically realized as a strong function of x_1 and x_2 and weakly (and insignificantly) formed by x_3 and x_4 , whereas its nominal meaning remains consistent with Example 1—as being a function of all four x_i . The construct has changed, although the definition of the construct has not. Whatever its nominal definition, “Formative” in Example 1 is not “Formative” in Example 2. This seems completely consistent with Burt’s (1976, p. 4) description of interpretational confounding as a phenomenon that “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.” Changing dependent constructs changes the formative construct.

Example 3—Failure to Function as a Point Variable

Now assume that the researcher encounters relatively unrelated endogenous constructs that relate differently to different indicators of the formatively measured construct. The correlations in Table 4 present such a situation, where x_1 and x_2 correlate strongly with y_1 – y_4 and weakly with y_5 – y_8 , whereas x_3 and x_4 relate strongly to y_5 – y_8 and weakly to y_1 – y_4 . In this situation (again, not unrealistic given the nature of formative measurement), η_1 cannot completely mediate the relationships between the formative indicators

Table 2
Correlations Corresponding to Example 1

Input	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
x_1	.44	.44	.44	.44	.21	.21	.21	.21
x_2	.43	.43	.43	.43	.26	.26	.26	.26
x_3	.35	.35	.35	.35	.45	.45	.45	.45
x_4	.35	.35	.35	.35	.39	.39	.39	.39

Note. y_1 – y_4 are measures of η_2 , y_5 – y_8 are measures of η_3 , and x_1 – x_4 are formative measures of η_1 , all corresponding to Figure 1. Intercorrelations among y_1 – $y_4 = .90$ and among y_5 – $y_8 = .90$. Correlations of y_1 – y_4 with y_5 – $y_8 = .51$. Intercorrelations among x_1 – $x_4 = .20$.

Table 3
Correlations Corresponding to Example 2

Input	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
x_1	.62	.62	.62	.62	.43	.43	.43	.43
x_2	.59	.59	.59	.59	.43	.43	.43	.43
x_3	.22	.22	.22	.22	.35	.35	.35	.35
x_4	.22	.22	.22	.22	.35	.35	.35	.35

Note. y_1 – y_4 are measures of η_2 , y_5 – y_8 are measures of η_3 , and x_1 – x_4 are formative measures of η_1 , all corresponding to Figure 1. Intercorrelations among y_1 – y_4 = .90 and among y_5 – y_8 = .90. Correlations of y_1 – y_4 with y_5 – y_8 = .57. Intercorrelations among x_1 – x_4 = .20.

and the endogenous constructs—that is, it fails to function as a point variable, lacking external consistency. As expected, this situation is immediately recognizable in the substantial lack of fit of the model, as indicated by both the chi-square and root-mean-square error of approximation values in the third column of Table 1. The relatively low and equal γ coefficients (standardized estimates of .36) reflect the compromise necessary in attempting to find a single set of coefficients emanating from η_1 to reproduce the disparate correlations between the formative indicators and η_2 and η_3 . To fit, η_1 would have to be two separate constructs.

This situation is not unique to formative measurement. Even a reflective measure with strong interindicant correlations can fail to exhibit external consistency. However, when the correlation among reflective indicants is high, there is reason to believe that they might relate similarly to other constructs. Given the lack of any necessary correlation among formative indicators, there is in general no reason to expect that they should not relate differently with different constructs, suggesting that finding a model that fits with a formatively measured construct may indeed be fortuitous.

For the sake of discussion, let us assume that the formatively measured construct in our example is SES. If SES can differ from model to model within a study and the models suffer from empirically indistinguishable alternative interpretations, then it can be expected to be even worse across studies. Although failure to meet the external consistency criterion is clearly reflected by the lack of fit in this example, if η_2 were to be a key endogenous construct in one SES study and η_3 a key outcome of SES in another, then lack of fit would not necessarily indicate the problem. That is, SES would change empirical interpretation but not nominal meaning. The accumulation of knowledge about formatively measured constructs and their consequences would seem to be virtually impossible across models with estimated model-dependent formative measurement parameters that may differ widely. This problem is apparent in reviews of the SES literature, such as that undertaken by Bradley and Corwyn (2002), who noted that the predictive value of specific composites of SES yields inconsistent results across research contexts. They observed that at times components

of SES are similar in their correlations with outcome measures, whereas “[A]t other times they appear to be tapping into different underlying phenomena and seem to be connected to different paths of influence” (p. 373).

Discussion

It seems clear that constructs with formative indicators are subject to interpretational confounding. Further, it seems clear that if measures of formative constructs need not have the same nomologic net and need not have the same antecedents and consequences, then forcing them into a unitary composite may be ill advised. Why, then, has the estimation and identification of formatively measured constructs received so much attention?

We are certainly not the first to notice the susceptibility of formatively measured constructs to interpretational confounding. Edwards and Bagozzi (2000, pp. 158–159) noted that,

Ideally, the association between a construct and its measures should remain stable regardless of the larger causal model in which the constructs and measures are embedded. If the association varies, then no unique meaning can be assigned to the constructs (Burt, 1976), and claims of association between the construct and measures are tenuous. The association between a construct and its measures is generally stable for reflective measures that serve as alternative (i.e., substitute) indicators of a construct, provided these measures correlate more highly with one another than with measures of other constructs. In contrast, associations of formative measures with their construct are determined primarily by the relationship between these measures and measures that are dependent on the construct of interest. This point was made by Heise (1972), who noted that a construct measured formatively is not just a composite of its measures; rather, “it is the composite that best predicts the dependent variable in the analysis. . . . Thus the meaning of the latent construct is as much a function of the dependent variable as it is a function of its indicators” (p. 160). Unstable associations between formative measures and constructs not only create difficulties for establishing causality but also obscure the meaning of these constructs because their interpretation depends on the dependent variables included in a given model.

Despite this clear realization of the interpretational confounding problem (but perhaps not the issue of point vari-

Table 4
Correlations Corresponding to Example 3

Input	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
x_1	.62	.62	.62	.62	.21	.21	.21	.21
x_2	.59	.59	.59	.59	.26	.26	.26	.26
x_3	.19	.19	.19	.19	.45	.45	.45	.45
x_4	.09	.09	.09	.09	.39	.39	.39	.39

Note. y_1 – y_4 are measures of η_2 , y_5 – y_8 are measures of η_3 , and x_1 – x_4 are formative measures of η_1 , all corresponding to Figure 1. Intercorrelations among y_1 – y_4 = .90 and among y_5 – y_8 = .90. Correlations of y_1 – y_4 with y_5 – y_8 = .25. Intercorrelations among x_1 – x_4 = .20.

ability), Heise (1972) continued with an innovative discussion of the treatment of formatively measured constructs and Edwards and Bagozzi (2000) continued with a thoughtful and comprehensive classification of the nature of relationships between constructs and indicators. Why, after clearly acknowledging one of the problems, do these scholars continue to develop methods for identifying and modeling formatively measured constructs? We can only speculate, but we assume that the prevalence of existing measures that have been considered as formative and the abundance of existing data containing formative measures motivate a desire for providing researchers with a methodology for appropriately dealing with them, despite their inherent limitations. This does not mean, however, that today's researchers, in designing future theory-testing studies, would be well served to opt for formative measurement in situations where a choice is possible.

Are Constructs Inherently Formative or Reflective?

According to Podsakoff et al. (2003), "Some constructs . . . are fundamentally formative in nature and should not be modeled reflectively" (p. 650). If constructs, per their nominal meaning, are inherently either formative or reflective, then the researcher would be obliged to measure them accordingly. It seems, however, that this would not be the case under the realist perspective (Borsboom et al., 2003) underlying our conceptualization of the nature of constructs. For example, Heise (1972, p. 153) suggested that, "SES is a construct induced from observable variations in income, education, and occupational prestige, and so on; yet it has no measurable reality apart from these variables which are conceived to be its determinants." However, Kluegel, Singleton, and Starnes (1977) developed subjective SES measures that function acceptably as reflective indicators.³ In questioning the applicability of their conceptualization of construct validity to formatively measured constructs, Borsboom, Mellenbergh, and van Heerden (2004) noted that,

One might 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?* Thus, that attributes like SES are typically addressed with formative models does not mean that they could not be assessed reflectively. (p. 1069)

It would also seem possible to assess an individual's SES through multiple informant reports, network analyses, and so forth.

Jarvis et al. (2003, p. 214) referred to a model similar to our Figure 1 but with five X s, where η_2 and η_3 are interpreted as, "content-valid measures tapping the overall level of the construct (e.g., overall satisfaction, overall assessment of perceived risk, overall trust, etc.), and X_1 – X_5 are measures of the key conceptual components of the construct (e.g., facets of satisfaction, risk, or trust)." They suggested

(Jarvis et al., 2003, p. 214) that the facets—such as "the consumer's satisfaction with distinct attributes or aspects of the product," which could include such things as satisfaction with price, size, quality, service, and color—and the overall measures (reflective indicants of a consumer's satisfaction with a product such as, "Overall, how satisfied are you with this product?" and "All things considered, I am very pleased with this product") should be considered as a measurement model with five formative and two reflective indicators:

Indeed, in this instance, it would not make sense to interpret this structure as five exogenous variables . . . influencing a separate and distinct endogenous latent construct . . . with two reflective indicators, because the five causes of the construct are all integral aspects of it and *the construct cannot be defined without reference to them*. (Jarvis et al., 2003, p. 214, emphasis added)

Satisfaction can and has been defined without reference to its antecedents, and numerous studies have measured overall satisfaction and its consequences (e.g., loyalty, repurchase intention, likelihood of recommending) without any measurement of its antecedents. Admittedly, understanding the drivers of satisfaction can lend insight and managerial relevance, but they need not be considered definitional. Constructs may be defined by their antecedents, but they need not be. As noted by Cohen et al. (1990, p. 186), "One investigator's measurement model may be another's structural model," but this potentially confusing situation need not be the case.

A similar perusal of the exemplars of constructs with formative indicators presented by Jarvis et al. (2003), the formative examples given by Diamantopoulos and Winklhofer (2001), the constructs considered by Bollen and Lennox (1991), and the cases investigated by Bollen and Ting (2000) suggests few that could not be, at least conceptually, defined without reference to their causes and measured reflectively.

At a conceptual level, most constructs could, at least conceivably, be measured reflectively. Although a given research setting or research tradition may tend to favor reflective or formative measurement of a construct, because we believe that under the realist perspective constructs exist apart from their measurement, we believe that it is not the construct itself that is formative or reflective.

³ Edwards and Bagozzi (2000, p. 159) suggested that SES as discussed here is best represented by an indirect formative model: "[I]f we assume scores on income, education and occupational prestige contain measurement error and occur after the phenomena they represent, . . . then these scores may be viewed as reflective measures of socioeconomic constructs that in turn cause SES." Bradley and Corwyn (2002) were even more explicit, noting that the concepts underlying the notion of SES are access to financial capital, human capital (nonmaterial resources, such as education), and social capital (resources achieved through social connections).

Suggestions for Future Measurement

Citing difficulties arising from identifying formative models with measurement error at the construct level (see MacCallum & Browne, 1993), Jarvis et al. (2003, p. 213) suggested that,

In our view, the best option for resolving the identification problem is to *add two reflective indicators* to the formative construct, when conceptually appropriate. The advantages of doing this are that (a) the formative construct is identified on its own and can go anywhere in the model, (b) one can include it in a confirmatory factor model and evaluate its discriminant validity and measurement properties, and (c) the measurement parameters should be more stable and less sensitive to changes in the structural relationships emanating from the formative construct. (emphasis added)

We completely agree with using reflective indicants. With three indicants, the measurement model for the construct alone is just identified and its parameters can be estimated, whereas with four or (preferably) more reflective indicants, the model is subject to empirical testing (e.g., unidimensionality, vanishing tetrads, and so forth). Our reasoning here is straightforward: With multiple reflective indicants there is a strictly epistemic base from which interpretational confounding can be assessed. Further, as the number of high-quality reflective indicants grows, the situation described by Burt (1976) as most likely to lead to interpretational confounding—when there are few reflective indicators and/or the epistemic relationship of those indicators to their associated construct is weak relative to structural parameters—is replaced by the opposite: several indicators with strong epistemic relationships. When epistemic relationships are strong, the likelihood that the construct will function as a point variable increases.

Diamantopoulos and Winklhofer (2001) similarly used reflective indicators in their construction of a formative measure. If the researcher were to use at least four or more reflective indicants, a potentially stable, testable reflective measure would be at hand. That is, in both instances, by using reflective indicators the formative measurement issue vanishes. If the researcher's goal is to predict the construct in question, adding the formative part of the model is clearly an option, but now the model and its parameters can have a much more straightforward interpretation. The interpretational confounding problem may or may not continue to be an issue, but the presence of interpretational confounding can be assessed and dealt with if found to be a problem.

Dealing With Existing Formative Measures

As is evident in the preceding discussion, there is an abundance of possibly formative measures for constructs of theoretical importance and much existing data with constructs measured formatively (for instance, the General Social Survey [Davis & Smith, 1991] from which Bollen and Ting [2000] drew examples). How should they be modeled?

Cohen et al. (1990) and Jarvis et al. (2003) clearly showed that, especially in the case in which the formative indicators have low intercorrelation, simply treating the indicators as reflective will lead to biased coefficients. There will be what amounts to overdisattenuation for measurement error, inflating those coefficients emanating from the improperly modeled construct, so this is not an automatic option.

Given the problem of interpretational confounding associated with estimating the γ s (along with the difficulty in interpreting the error term ζ_1) in the model depicted in Figure 1 (by necessity having two paths emanating from the formatively measured construct to be identified), it may seem logical to revert to Equation 2, foregoing the estimation of a construct level error term. Because PC analysis is the only method for estimating the γ s in Equation 2 not subject to interpretational confounding, it may seem to be a reasonable option. However, to the extent that the formative indicants are uncorrelated, a single PC will account for very little variance in the set of indicators. Indeed, if the indicators are completely uncorrelated, then there will be as many nontrivial PC as there are indicators. If some of the indicators are intercorrelated and some are not, then the first PC will largely represent the indicators that covary, which is inconsistent with the conceptualization of formative measurement.

Another approach may be to predetermine the γ s in Equation 2, rather than estimating them. This does indeed avoid the problem of interpretational confounding. Weights could be determined on the basis of the theoretic contribution of each of the formative indicators to the composite, or equal weights could be assigned in which the indicators form a summated scale. There are, however, shortcomings with this approach.

First, there is a high potential for loss of information in forming the composite of uncorrelated variables (see the Appendix). For example, if three, say, Likert-type items measured on 5-point scales are considered independently, then there are 125 ($5 \times 5 \times 5$) possible numeric configurations. That is, a subject could have any one of 125 different possible score patterns across the three items. If, however, the three items are summed to form a measure, then the number of values the measure can take on is 13 (3, 4, 5, . . . , 15), representing a potential information loss of 90%. This is not a serious problem when forming a summated scale for reliable indicators because the number of observed configurations will be substantially smaller than the number of possible configurations. That is, when indicators are substantially correlated, many of the configurations will be sparsely represented in one's data (e.g., scores of 5, 1, 5 or 1, 5, 1). Because formative indicators are not necessarily expected to correlate, all possible configurations could be considered equally likely.

Second, and closely related to the above, forming a composite of potentially unrelated indicators ignores the possi-

ble effects of differing configurations that can lead to the same composite score. For example, a three indicator composite with $x_1 = x_2 = x_3 = 20$ results in $\eta = 60$ (given $\gamma_i = 1$). Would one expect this to have an identical impact on endogenous construct(s) as a configuration of $x_1 = 50$, $x_2 = 5$, and $x_3 = 5$, also resulting in $\eta = 60$? Assuming each scale has an integer range of 1 to 50, there are 1,603 different combinations of x_1 , x_2 , and x_3 that yield an identical score of 60. Again, because the indicators may be uncorrelated, one could expect to see any of these configurations. This is, of course, a matter of theory, but it should not be overlooked in forming a composite. Configurations could imply interactions among the predictors in their effects on endogenous constructs. Indeed, Blalock (1982, p. 88) explicitly recommended the consideration of multiplicative models. Such interactions can be modeled when individual formative indicators are used.

Third, we refer to our previous discussion of the degree to which uncorrelated indicators can function as point variables or "unitary attributes" (Nunnally, 1967). As discussed earlier, if formative indicants need not have the same nomologic net and need not have the same antecedents and consequences, where is the logic of forming them into a single composite?

If estimating the measurement parameters in formative models may lead to issues of interpretational confounding and predetermining the weights is problematic with respect to information loss and the point variability issue, then what is one to do with formative measures? A reasonable course of action appears to be, at least initially, to treat the measures as different constructs. For example, Adams, Hurd, McFadden, Merrill, and Ribeiro (2003) considered SES as a vector variable, including measures of wealth, income, education, neighborhood, and dwelling as independent predictors of a variety of health conditions, and found that there are important differences in the relationships among the components of SES and health conditions. Edwards and Parry (1993), in the context of difference scores (scores formed by taking the difference between variables), proposed using variables individually in response surface analysis (polynomial regression), indicating that the use of such methods allows for the examination of the separate and joint influences of the variables. We suggest the same arguments apply to composites (weighted or unweighted) of formative indicators—that the indicators should be used individually.

Using the components of a formative measure as individual constructs allows the differential effects of each aspect of the overall construct to be apparent. Clearly, this approach is at the expense of parsimony and fails to account for measurement error in the indicators—but forming a composite of potentially unrelated or even negatively related indicators similarly does not account for measurement error at the indicator level and does not seem to be a good way to advance understanding. Similar arguments can be

made when the formatively measured latent construct is endogenous. Again, if the endogenous construct's formative indicants differ with respect to their antecedents, then one would be better served by specifying them as different constructs. The lack of parsimony when modeling formative indicators as separate constructs is an issue. However, finding a common label for disparate constructs in the name of parsimony may be part of the problem. When one gives a name to a collection of attributes or characteristics in a common realm for the sake of convenient communication and then treats them as if a corresponding entity exists, "one is reifying terms that have no other function than that of providing a descriptive summary of a set of distinct attributes and processes" (Borsboom et al., 2004, p. 1065). If the indicators do show substantially similar relationships to their antecedents and consequences in a model where they are estimated as separate constructs, then their aggregation may indeed be justified because external consistency or point variability is indicated (i.e., the appropriate tetrads in Bollen & Ting, 2000, vanish), but this is an empirically answerable question. In the event that they do, the logically formative indicators are not unlikely to be positively correlated and using them as separate constructs may result in multicollinearity problems and the associated difficulty in interpreting parameters.

It is interesting to note that Bollen and Ting (2000) discussed the possibility that a set of indicators may be causal with respect to one construct and reflective with respect to another. In their example, a child's viewing of violent television programs, playing violent video games, and listening to music with violent themes may be formative indicators of "exposure to media violence" but reflective indicators of "propensity to seek violent entertainment." As such, a reconceptualization of the construct may be appropriate.

Delimiting Conditions

We want to be clear that our discussion pertains to the measurement of constructs as elements of theory, in the theory evaluation and testing endeavor. Formative composites in the form of indexes (Diamantopoulos & Winklhofer, 2001) are conceptually appropriate. (Of interest, Diamantopoulos & Winklhofer, 2001, begin their discussion in the context of index construction, yet their example refers to constructs in the context of theory testing.) When a particular dependent phenomenon is of interest and no theoretical interpretation is attached to the meaning of the index, PLS, CC, RA, or subjective weighting are all potentially appropriate (see Arnett, Laverie, & Meiers, 2003, for such an example). Similarly, many formative constructs are simply concatenations. Just as it is appropriate to add together the number of times a 12-cm ruler must be laid end to end to measure the length of an object, simply adding together a

homogeneous interchangeable set of occurrences is often justified, although whether such qualifies as formative measurement of a latent variable is another question.

Summary

In summary, we strongly suggest that when designing a study, researchers should attempt to measure their constructs reflectively with at least three, and preferably as many as is (practically) possible, strongly correlated indicators that are unidimensional for the same construct. This is indeed conventional wisdom, but we believe it remains correct. The reflective measurement model is and has been dominant in psychology and other social sciences with good reason, and it should retain its dominance. The degree to which interpretational confounding is a problem with reflective measures in the context of a larger model can and should be assessed. With formative measurement, there are no strictly epistemic relationships to fall back on and the estimated relationships between the construct and its measures must be defined in terms of other constructs in the model. It is then the responsibility of the researcher to compare the formatively measured construct's nominal definition with its empirical realization and, in the event the two differ, to modify the discussion (and perhaps the labeling or naming) of the construct to reflect the difference.

The degree to which a reflective measure operates as a point variable can and should be assessed through an examination of model fit. The same is true for formative measures, but there is greater reason to believe that this may be a problem for indicators that, by definition, need not have the same antecedents and consequences.

We conclude that formative measurement is not an equally attractive alternative. Although we would perhaps not go so far as Ping (2004, p. 133), who stated, "Because formative constructs are not unobserved variables I will not discuss them further," we note that Borsboom et al. (2004) questioned the applicability of their conceptualization of validity in the context of formative measurement. This follows from their straightforward concept of validity (does the attribute exist and do variations in the attribute causally produce variations in the outcomes of the measurement procedure?), which is inconsistent with formative measurement.

In the case of existing data or when reflective measurement is infeasible or impossible, we recommend that formative indicators be modeled as separate constructs, at least until the equivalence of their relationships with antecedents and/or consequences can be established. The choice to use reflective measurement for constructs traditionally measured formatively may indeed change the nature of the construct but perhaps with additional insight. As Blalock (1982) pointed out, "it is again impor-

tant to emphasize the necessity of explicit theoretical formulations that lay open to public inspection exactly what assumptions are being made in each measurement decision" (p. 107). We agree.

References

- Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112, 3–56.
- Anderson, J. C., & Gerbing, D. W. (1982). Some methods for respecifying measurement models to obtain unidimensional construct measurement. *Journal of Marketing Research*, 19, 453–460.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411–423.
- Arnett, D. B., Laverie, D., & Meiers, A. (2003). Developing parsimonious retailer equity indexes using partial least squares analysis: A method and applications. *Journal of Retailing*, 79, 161–170.
- Bagozzi, R. P. (1980). *Causal models in marketing*. New York: Wiley.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182.
- Blalock, H. M. (1964). *Causal inference in nonexperimental research*. New York: Norton.
- Blalock, H. M. (1971). Causal models involving unmeasured variables in stimulus–response situations. In H. M. Blalock (Ed.), *Causal models in the social sciences* (pp. 335–347). Chicago: Aldine.
- Blalock, H. M. (1982). *Conceptualization and measurement in the social sciences*. Beverly Hills: Sage.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605–634.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314.
- Bollen, K. A., & Ting, K. (2000). A tetrad test for causal indicators. *Psychological Methods*, 5, 3–32.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203–219.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399.
- Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research*, 5, 3–52.
- Cliff, N. (1983). Some cautions concerning the application of

- causal modeling methods. *Multivariate Behavioral Research*, 18, 115–126.
- Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement*, 14, 183–196.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Davis, J. A., & Smith, T. W. (1991). *General social surveys, 1972–1991: Cumulative codebook*. Chicago: NORC.
- DeVellis, R. F. (1991). *Scale development: Theories and applications*. Newbury Park, CA: Sage.
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators. *Journal of Marketing Research*, 38, 269–277.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.
- Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, 36, 1577–1613.
- Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit—Voice theory. *Journal of Marketing Research*, 10, 440–452.
- Fornell, C., & Yi, Y. (1992). Assumptions of the two-step approach to latent variable modeling. *Sociological Methods and Research*, 20, 291–320.
- Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analysis. *Sociological Methods and Research*, 1, 147–173.
- Hempel, C. G. (1970). Fundamentals of concept formation in empirical science. In O. Neurath, R. Carnap, & C. Morris (Eds.), *Foundations of the unity of science* (Vol. 2, pp. 653–740). Chicago: University of Chicago Press.
- Howell, R. D. (1987). Covariance structure modeling and measurement issues: A note on “Interrelations among a channel entity’s power sources.” *Journal of Marketing Research*, 14, 119–126.
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30, 199–216.
- Jöreskog, K., & Goldberger, S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70, 6331–6339.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the simplis command language*. Lincolnwood, IL: Scientific Software International.
- Kluegel, J. R., Singleton, R., & Starnes, C. E. (1977). Subjective class identification: A multiple indicators approach. *American Sociological Review*, 42, 599–611.
- Law, K. S., Wong, C., & Mobley, W. H. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, 23, 741–755.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114, 533–541.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, 1, 108–141.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Ping, R. A., Jr. (2004). On assuring valid measurement for theoretical models using survey data. *Journal of Business Research*, 57, 125–141.
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Lee, J. Y. (2003). The mismeasure of man(agement) and its implications for leadership research. *The Leadership Quarterly*, 14, 615–656.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87, 517–529.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 17.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Structural Equation Modeling: A Multidisciplinary Journal*. (2000). 7(1).
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Van den Wollenberg, A. L. (1977). Redundancy analysis—An alternative to canonical correlation analysis. *Psychometrika*, 42, 207–219.

(Appendix follows)

Appendix

Information Loss From Forming a Composite

Number of components	Number of scale points	Information loss
2	5	0.640
3	5	0.896
4	5	0.973
5	5	0.993
2	6	0.694
3	6	0.926
4	6	0.984
5	6	0.997
2	7	0.735
3	7	0.945
4	7	0.990
5	7	0.998

Note. Let NC = number of components in a formative measure and NSP = number of scale points per component. Assuming scale points start at one and run to NSP, the number of unique configurations from using the components is $[\text{NSP}]^{\text{NC}}$. The number of values that can be taken on in a formative scale formed as a composite is $\text{NC}(\text{NSP}) - (\text{NC} - 1)$. Information loss can thus be defined as a percentage of information available (as measured by number of unique configurations): $1 - [\text{NC}(\text{NSP}) - (\text{NC} - 1)] / \text{NSP}^{\text{NC}}$.

Received October 5, 2004

Revision received October 16, 2006

Accepted November 14, 2006 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!