The Fallacy of Formative Measurement

Jeffrey R. Edwards[1]

Abstract

In management research, there is a growing trend toward formative measurement, in which measures are treated as causes of constructs. Formative measurement can be contrasted with reflective measurement, in which constructs are specified as causes of measures. Although recent work seems to suggest that formative measurement is a viable alternative to reflective measurement, the emerging enthusiasm for formative measurement is based on conceptions of constructs, measures, and causality that are difficult to defend. This article critically compares reflective and formative measurement on the basis of dimensionality, internal consistency, identification, measurement error, construct validity, and causality. This comparison leads to the conclusion that the presumed viability of formative measurement is a fallacy, and the objectives of formative measurement can be achieved using alternative models with reflective measures.

Management research places a premium on the development of theory (Bacharach, 1989; Smith & Hitt, 2005; Sutton & Staw, 1995; Whetten, 1989). Theory development usually emphasizes the relationships among constructs, describing the direction, sign, and form of these relationships and explaining why and under what conditions these relationships occur. An equally important aspect of theory development concerns the relationships between constructs and measures (Edwards & Bagozzi, 2000). These relationships constitute an auxiliary theory that connects abstract theoretical constructs to observable phenomena (Blalock, 1968; Costner, 1969), thereby rendering theories amenable to empirical research.

When developing auxiliary theories, perhaps, the most basic consideration involves the direction of the relationship between constructs and measures. One option is to treat constructs as causes of measures, such that measures are reflective manifestations of underlying constructs. Reflective measurement is rooted in the common factor model (Harman, 1976; Kim & Mueller, 1978), which treats measures as outcomes of unobserved latent variables. Another option is to specify measures as causes of constructs, such that measures form or induce an underlying latent variable. Formative measurement is consistent with principal components analysis (Dunteman, 1989; Jolliffe, 2002),

---

[1] University of North Carolina, Chapel Hill, USA

Corresponding Author:
Jeffrey R. Edwards, University of North Carolina, Chapel Hill, Box No 3490, NC 27599, USA
Email: jredwards@unc.edu

in which measures are combined to form weighted linear composites intended to represent theoretically meaningful concepts.

Formative measurement is gaining interest in management research, as evidenced by journal issues devoted to the advancement of formative measurement (Diamantopoulos, Riefler, & Roth, 2008) and prescriptive articles indicating that formative measurement should be more widely adopted in management research (Diamantopoulos & Siguaw, 2006; Law & Wong, 1999; MacKenzie, Podsakoff, & Jarvis, 2005; Podsakoff, Shen, & Podsakoff, 2006). This surge of interest has been fueled by discussions of formative measurement in psychology (Bollen & Lennox, 1991; MacCallum & Brown, 1993), sociology (Blalock, 1964; Bollen, 1984; Heise, 1972), and marketing (Diamantopoulos & Winklhofer, 2001; Fornell & Bookstein, 1982; Jarvis, MacKenzie, & Podsakoff, 2003; Rossiter, 2002), which position formative measurement as a viable alternative to reflective measurement. Moreover, formative measurement might appeal to researchers who want to study constructs that combine multiple dimensions, examine the linkages between specific and general constructs, or summarize relationships involving several conceptually related dimensions in terms of a single parameter. In combination, these forces are likely to attract increasing numbers of management researchers toward formative measurement, and the momentum of this movement shows little sign of abating.

In this article, I argue that the growing enthusiasm surrounding formative measurement is misguided, and justifications given for using formative measures are based on expressed beliefs about constructs, measures, causality, and other measurement issues that are difficult to defend. These beliefs are fodder for methodological urban legends in which reasons for using formative measures are propagated without stopping to question their veracity. As a countermeasure, I develop criticisms of formative measurement, organized in terms of six core themes that address dimensionality, internal consistency, identification, measurement error, construct validity, and causality. These criticisms integrate and extend concerns about formative measurement that have begun to emerge (Bagozzi, 2007; Borsboom, 2005; Borsboom, Mellenbergh, & Van Heerden, 2003, 2004; Howell, Breivik, & Wilcox, 2007a, 2007b; Iacobucci, 2010; Wilcox, Howell, & Breivik, 2008). I further contend that the rationale for formative measurement can be fulfilled using reflective measurement models specified in particular ways, and these models serve the purposes of formative measurement models while avoiding their drawbacks. The ultimate goals of this article are to expose the fallacy of formative measurement and encourage researchers to adopt alternative measurement models that lead to better auxiliary theories relating constructs to measures in management research.

Reflective and Formative Measurement Models

The distinctions between reflective and formative measures can be seen by comparing their respective measurement models. As indicated earlier, reflective measures are treated as outcomes of constructs. A reflective measurement model is shown in Figure 1, in which ξ signifies a latent variable representing the construct of interest and x_1, x_2, and x_3 are reflective measures of the construct. The δ_1, δ_2, and δ_3 are uniqueness associated with the reflective measures and combine item specificity with random measurement error (Bollen, 1989; Long, 1983), and the loadings λ_1, λ_2, and λ_3 capture the magnitude of the effects of ξ on x_1, x_2, and x_3. As a substantive example, ξ could be conceived as the autonomy an employee perceives in his or her job, and x_1, x_2, and x_3 could be scores on the items "Doing my work in my own way," "Determining the way my work is done," and "Making my own decisions" (Cable & Edwards, 2004). The arrows leading from ξ to x_1, x_2, and x_3 capture the premise that perceived autonomy influences scores on the 3 items that reflect the level of the underlying construct. This premise represents a critical realist perspective (Campbell, 1960; Loevinger, 1957; Messick, 1981) in which constructs are considered real entities that influence scores on their associated measures (Borsboom et al., 2003; Edwards & Bagozzi, 2000).
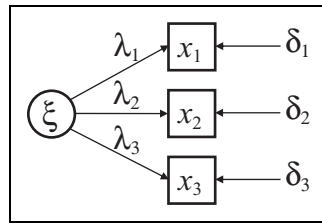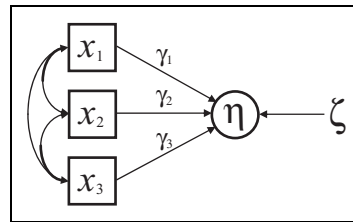
Figure 1. Reflective measurement model

Figure 2. Formative measurement model

A formative measurement model is given in Figure 2, where η depicts the construct of interest and x_1, x_2, and x_3 are formative measures of the construct. The coefficients γ_1, γ_2, and γ_3 indicate the magnitude of the effects of x_1, x_2, and x_3 on η, and the residual ζ is taken to represent aspects of η not explained by x_1, x_2, and x_3. Occasionally, the residual term ζ is excluded from formative measurement models, in which case the latent variable η is an exact weighted linear composite of its measures. Figure 2 shows that x_1, x_2, and x_3 freely correlate, such that the relationships among formative measures are absorbed by their intercorrelations, not by the paths relating the measures to the construct (MacCallum & Browne, 1993). An example of formative measurement is overall job satisfaction measured in terms of satisfaction with specific job facets, such as work, pay, coworkers, supervisor, and promotion opportunities (MacKenzie et al., 2005). Formative measurement is consistent with a constructivist position (Fosnot, 1996; von Glasersfeld, 1995) in which constructs are viewed as elements of language in theoretical discourse and are not ascribed any real existence independent of their measurement (Borsboom et al., 2003; Nunnally, 1978). Formative measurement might also be framed in terms of operationalism or instrumentalism, such that constructs are merely latent variables that serve as analytical devices for combining measures, akin to data reduction in principal components analysis (Borsboom et al., 2003). The philosophical underpinnings of formative and reflective measurement are considered in greater detail later in this article.

Reflective and formative measurement models can also be distinguished according to the equations implied by each model. A reflective measurement model corresponds to the following equation:

$$x_i = \lambda_i \xi + \delta_i \tag{1}$$

where x_i is a reflective measure, ξ is its associated construct, λ_i is the effect of ξ on x_i, and δ_i is the uniqueness of the measure (i ranges from 1 to 3 for the model in Figure 1). In contrast, a formative measurement model is represented by the following equation:

$$\eta = \gamma_i x_i + \zeta \tag{2}$$

where x_i is a formative measure, η is the construct, γ_i is the effect of x_i on η, and ζ is the residual, which is taken as that part of η not explained by the x_i (i again ranges from 1 to 3 for the model in Figure 2). If

the residual is omitted from a formative measurement model, then Equation 2 reduces to:

$$\eta = \gamma_i x_i \tag{3}$$

which specifies η as a weighted linear combination of the x_i. Equations 1–3 can be interpreted as regression equations relating constructs to measures (Bollen & Lennox, 1991). For instance, Equation 1 indicates that x_i is dependent on ξ, and the proportion of variance in x_i explained by ξ is the reliability of x_i. For Equations 2 and 3, η is dependent on x_i, although the proportion of variance in η explained by the x_i does not signify reliability, but instead is the relative amount of η explained by the x_i. In Equation 2, the causes of η beyond the x_i are collapsed into ζ, whereas Equation 3 excludes any causes of η other than the x_i, such that the variance in η is fully explained by the x_i.

Comparing Reflective and Formative Measurement

Discussions of measurement models in the psychological, sociological, and management literatures identify various features that distinguish reflective and formative measures. Some of these features are framed as criteria for deciding whether measures should be specified as reflective or formative, whereas other features are consequences of this specification decision. This section summarizes key features that distinguish reflective and formative measures, with an eye toward evaluating the conditions under which formative measurement constitutes a viable alternative to reflective measurement.

Dimensionality

Reflective and formative measures have been distinguished according to whether the measures are unidimensional or multidimensional. Reflective measures are assumed to represent a single dimension, such that the measures describe the same underlying construct, and each measure is designed to capture the construct in its entirety (Bollen 1984; Bollen & Lennox, 1991; Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005). Because they describe the same dimension, reflective measures are conceptually interchangeable, and removing any one of the measures would not alter the meaning or interpretation of the construct (Bollen & Lennox, 1991; MacKenzie et al., 2005; Podsakoff et al., 2006). When designed properly, reflective measures exhibit what DeVellis (1991) calls useful redundancy, such that the items have the same meaning without relying on the same terminology or grammatical structure.

Unlike reflective measures, formative measures are characterized as describing different dimensions or facets of a construct (Bollen & Lennox, 1991; Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005; Podsakoff et al., 2006). The multidimensionality of formative measures can be attributed to guidelines for formative measurement, which recommend that each measure should describe a distinct aspect of the construct, and redundancy among the measures should be eliminated during the measurement development process (Diamantopoulos & Siguaw, 2006). The idea that formative measures describe distinct aspects of a construct is also manifested by the notion that eliminating a formative measure is tantamount to removing part of the construct (Bollen & Lennox, 1991; Diamantopoulos & Siguaw, 2006; Franke, Preacher, & Rigdon, 2008; MacKenzie et al., 2005).

There is little dispute that formative measures typically describe multiple dimensions, as often demonstrated by examples used to illustrate formative measurement (Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005; Podsakoff et al., 2006). However, there is reason to dispute whether the construct represented by conceptually heterogeneous measures is useful or interpretable. When conceptually distinct measures are channeled into a single construct, the resulting construct is conceptually ambiguous. This ambiguity is analogous to that created by multibarreled items, which ask

respondents to assign one score to a question that describes more than one idea (Converse & Presser, 1986; DeVellis, 1991). For instance, consider an item intended to measure job performance that asks respondents to assign a single rating to task performance, job dedication, and interpersonal facilitation. Scores on this triple-barreled item would confound the three dimensions of job performance, making it unclear whether variation in the scores represented variation in all three dimensions simultaneously, one dimension alone, or some other combination of the dimensions. Moreover, most researchers trained in survey design would identify the triple-barreled item as problematic, perhaps, replacing it with 3 items that describe the dimensions separately. However, if scores on these 3 items were collapsed into a single variable, that variable would suffer from the same type of ambiguity associated with the triple-barreled item. This ambiguity would undermine the interpretation of the variable, regardless of whether it is a summed score or a latent variable with the scores treated as formative measures (MacKenzie et al., 2005).

It might seem that the ambiguity surrounding a latent variable formed from conceptually distinct formative measures can be resolved by knowing the magnitudes of the paths linking the measures to the latent variable. However, the meaning of the latent variable is determined not only by the paths from the measures but also by the variances and covariances of the measures. To illustrate, consider the following equation, which refers to the formative model in Figure 2:

$$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \zeta \tag{4}$$

The variance of the construct η can be written as:

$$\begin{aligned} V(\eta) = {}& \gamma_1^2 V(x_1) + \gamma_2^2 V(x_2) + \gamma_3^2 V(x_3) \\ & + 2[\gamma_1\gamma_2 C(x_1, x_2) + \gamma_1\gamma_3 C(x_1, x_3) + \gamma_2\gamma_3 C(x_2, x_3)] + V(\zeta) \end{aligned} \tag{5}$$

where V(.) and C(.) refer to variance and covariance, respectively. Equation 5 shows that the variance of η is determined not only by the magnitudes of γ_1, γ_2, and γ_3 but also by the variances and covariances of x_1, x_2, and x_3. As the variance of one measure increases relative to the others, the meaning of η will become dominated by that measure. The interpretation of η is further complicated by variation in the magnitudes of γ_1, γ_2, and γ_3, which combine with the variances and covariance of x_1, x_2, and x_3 to determine the variance of η. Thus, the multidimensionality of formative measures should be considered a liability, not a property that formative measurement models accommodate in some useful fashion.

Internal Consistency

Another feature that distinguishes reflective and formative measures concerns the internal consistency of the measures. Internal consistency refers to the extent to which measures are positively correlated, with higher correlations resulting in higher estimates of internal consistency reliability, such as Cronbach's alpha (Cronbach, 1951) and coefficient omega (Heise & Bohrnstedt, 1970; Jöreskog, 1971). Reflective measures are expected to correlate positively, given that they are designed as alternative indicators of the same underlying construct (Bollen 1984; Bollen & Lennox, 1991; Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005; Podsakoff et al., 2006). Internal consistency among reflective measures also arises from the structure of reflective measurement models, as shown in Figure 1. As the loadings relating the measures to the construct increase, the correlations among the measures likewise increase, given that the reproduced correlation between any pair of reflective measures is a function of the product of the loadings of the measures (Bollen, 1989).

In contrast to reflective measures, formative measures are not necessarily expected to demonstrate internal consistency. As noted earlier, formative measures are intended to represent different facets of a construct (Diamantopoulos & Siguaw, 2006), and generally speaking, there is no necessary reason to

expect the facets represented by formative measures to correlate with one another (Bedeiań, Day, & Kelloway, 1997; Bollen, 1984; Bollen & Lennox, 1991; MacCallum & Browne, 1993; MacKenzie et al., 2005; Nunnally & Bernstein, 1994; Podsakoff et al., 2006). Moreover, prescriptions for developing formative measures often treat high correlations among the measures as a problem, akin to multicollinearity in multiple regression analysis (Bollen & Lennox, 1991; Diamantopoulos et al., 2008; Diamantopoulos & Winklhofer, 2001; MacKenzie et al., 2005). As the correlations among formative measures increase, the loadings relating the measures to the construct become unstable and tend to exhibit large standard errors, which create difficulties for estimation and interpretation. Furthermore, low correlations among formative measures imply that each measure represents a unique facet of the construct, which is considered desirable for formative measurement.

The lack of internal consistency expected for formative measures has planted the seeds for various misconceptions. For instance, in response to evidence of low internal consistency, some researchers have concluded that their measures are formative rather than reflective. This conclusion is unjustified, because low internal consistency could merely constitute evidence of poorly designed reflective measures (Edwards & Bagozzi, 2000; Diamantopoulos & Siguaw, 2006). Furthermore, low internal consistency is neither necessary nor sufficient to conclude that measures are formative. As shown by the models in Figure 2, formative measures are exogenous variables and, as such, their covariances are not explained by the model but instead are taken as given. Consequently, there is no basis to expect covariances among formative measures to be any particular size or follow any type of pattern. In contrast, if a reflective measurement model is correctly specified, the covariances among the measures should follow predictable patterns (Anderson & Gerbing, 1988; Bollen & Ting, 2000). Nonetheless, the occurrence of these patterns does not provide evidence that measures are reflective, because the patterns expected for reflective measures could emerge for formative measures by happenstance. Thus, the internal consistency among a set of measures has no bearing on whether the measures should be treated as reflective or formative. Internal consistency is relevant when researchers choose reflective over formative measurement, but it is not justification for making that choice.

Identification

Reflective and formative measurement models also differ in terms of identification, which concerns whether unique values can be obtained for model parameters. In general, a reflective measurement model is identified provided that the model has at least three measures, the uniquenesses are independent, and a scale is set for the latent variable (Bollen, 1989; Davis, 1993; Reilly, 1995).[1] These conditions are met for the model in Figure 1 assuming a scale is set for ξ, which is usually accomplished by fixing its variance to unity or by fixing a loading to some constant, typically unity. Specified in this manner, the model in Figure 1 is just identified and will fit the data perfectly. With four or more measures, the model is overidentified, and the fit of the model to the data can be tested. With two measures, the model is identified if additional constraints are imposed (e.g., the loadings are set equal to one another) or the latent variable is allowed to covary with another latent variable that has at least two reflective measures (Bollen, 1989; Costner, 1969). With one measure, the model is identified if the loading and the variance of the uniqueness are both fixed.

In contrast to reflective measurement models, formative measurement models such as that shown in Figure 2 are not identified, regardless of the number of measures used. To achieve identification, the model must be supplemented by at least two reflective measures that are caused directly or indirectly by the latent variable (Bollen & Davis, 2009; MacCallum & Browne, 1993). Examples of such models are shown in Figures 3 and 4. The model in Figure 3 adds two reflective measures directly to the construct η, producing a multiple indicator multiple cause (MIMIC) model (Hauser & Goldberger, 1971). Reflective measures used in this manner should describe the construct in its entirety, such that the measures and construct are at the same level of abstraction. For instance, if
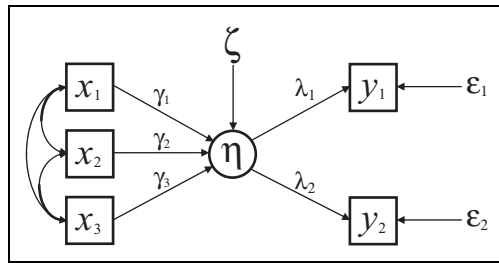
Figure 3. Multiple indicator multiple cause (MIMIC) model

x_1, x_2, and x_3 describe different facets of job performance, y_1 and y_2 could describe job performance in general terms (e.g., "overall, this employee performs the job well" and "this employee fulfills the requirements of the job"). The model in Figure 4 incorporates two constructs as outcomes of the focal construct, with each outcome having two reflective measures. If again the focal construct is job performance, these additional constructs could be outcomes such as supervisor recognition and career advancement.

The complexities surrounding the identification of formative measurement models might seem benign, given that identification rules for such models are available (Bollen & Davis, 2009; MacCallum & Browne, 1993), and in principle, reflective measures needed for identification can be found for most models. However, these identification rules require researchers to add reflective measures to their models regardless of whether these measures and the outcomes they represent are conceptually justified or relevant to the goals of the study. For instance, consider a model that specifies stress as a cause of health measured with an index of symptoms treated as formative measures (Fayers & Hand, 1997). To achieve identification, the model would have to include outcomes of health even if the theory guiding the study treats health as a criterion with no outcomes of its own (e.g., Edwards, Caplan, & Harrison, 1998). More problematic is the fact that the reflective measures used for identification have substantial effects on the loadings relating the measures to the construct, which in turn affect the meaning of the construct itself (Edwards & Bagozzi, 2000; Franke et al., 2008; Heise, 1972; Howell et al., 2007b). As explained by Heise (1972), the construct induced by formative measures "is not just a composite formed from its indicators; 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, and the results of any single analysis cannot be used to create a generalized scale" (p. 160). This property of formative measurement models renders the interpretation of the construct unstable, because choosing different outcomes can alter the meaning of the construct, perhaps to a substantial degree (Diamantopoulos, 2006; Howell et al., 2007b). This problem is a manifestation of interpretational confounding (Edwards & Bagozzi, 2000; Howell et al., 2007b; Wilcox et al., 2008), which occurs when "the assignment of empirical meaning to an unobserved variable . . . 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). Thus, when specifying formative measurement models, what might seem like an innocuous choice of outcomes to achieve identification can have fundamental ramifications for the meaning of the formative construct and the inferences based on it.

Measurement Error

Reflective and formative measurement models have been distinguished in terms of how they accommodate measurement error. Reflective measurement models incorporate error for each measure as
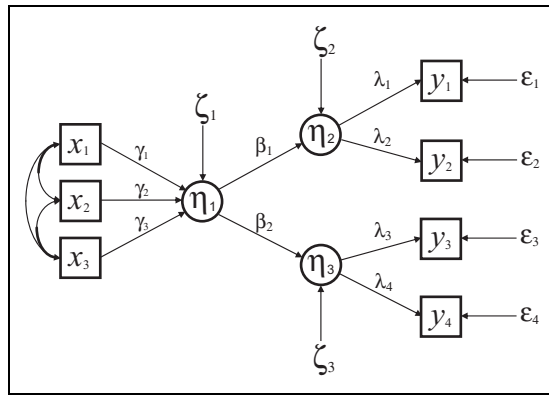
Figure 4. Formative measurement model with two outcome constructs

part of the uniqueness terms δ_i in Equation 1 and δ_1, δ_2, and δ_3 in Figure 1. As noted earlier, these terms combine item specificity with random measurement error, consistent with the common factor model on which reflective measurement is based (Harman, 1976; Kim & Mueller, 1978). Typically, the uniqueness terms are specified as independent of one another, although covariances are sometimes introduced to represent minor factors shared by subsets of measures (Gerbing & Anderson, 1984) or methodological factors common across measures (Cole, Ciesla, & Steiger, 2007; Marsh, 1989). When specified as shown in Figure 1, reflective measurement models correct relationships among latent variables for measurement error in a manner analogous to the correction for attenuation underlying classical measurement theory (Bollen, 1989; Cohen, Cohen, Teresi, Marchi, & Velez, 1990; DeShon, 1998).

Formative measurement models do not incorporate measurement error, as evidenced by the absence of uniqueness terms assigned to the formative measures in Figure 2. Rather, these models assign an error term to the construct, represented by the residual ζ in Figure 2. As noted earlier, this error term should be interpreted not as measurement error, but instead as aspects of the construct η not associated with its measures (Bollen, 2007; Diamantopoulos, 2006; Franke et al., 2008; MacKenzie et al., 2005; Podsakoff et al., 2006). In this sense, formative measurement rests on the assumption that measures are error-free indicators of the facets they are intended to represent (Diamantopoulos, 2006; Iacobucci, 2010). For instance, if x_1, x_2, and x_3 in Figure 2 are scores on parental income, education, and occupation, such that η refers to socioeconomic status, we must assume that these scores have no error, such that the levels of income, education, and occupation recorded in the scores match the levels of these facets as they actually exist.

The assumption that formative measures contain no error is difficult to reconcile with the basic premise that measures are nothing more than scores collected using methods such as self-report, interview, or observation (Edwards & Bagozzi, 2000; Lord & Novick, 1968; Messick, 1995). The information yielded by these methods is not perfect, but instead is flawed due to errors in recall, idiosyncratic interpretation of items, transient distractions, coding mistakes, technical problems, and other vagaries of the measurement process. Returning to the case of socioeconomic status, it would be truly remarkable if scores on a variable as seemingly objective as parental income exactly match the income that each respondent's parents actually earned (Borsboom, 2008). Consequently, researchers should allow for measurement error when interpreting and analyzing the scores that constitute measures, whether those measures are viewed as reflective or formative. Treating formative measures as if they contain no error yields biased estimates of the loadings relating the measures to the construct (i.e., the γ_i in Figure 2) for much the same reason that measurement error introduces bias into coefficient estimates in multiple regression analysis (Cohen, Cohen, West, & Aiken, 2003;

Pedhazur, 1997). Because formative measurement models do not account for measurement error, they fail to capitalize on one of the key advantages of structural equation modeling (Bollen, 1989).

Construct Validity

The criteria used to evaluate construct validity differ for reflective and formative measures. With reflective measures, construct validation centers on the extent to which the measures represent the construct of interest. This perspective is consistent with the definition of construct validity as the correspondence between a construct and a measure treated as an indicator of the construct (Cronbach & Meehl, 1955; Edwards, 2003; Nunnally, 1978; Schwab, 1980). From a conceptual standpoint, the correspondence between a construct and its measures depends on whether the definition of the construct is embodied in the measures. This aspect of construct validity relies on the judgment of the researcher to evaluate the extent to which measures can be interpreted in a manner consistent with the meaning of the construct (Cronbach & Meehl, 1955; Edwards, 2003; Schwab, 1980). Empirically, the correspondence between a construct and its measures is manifested by the magnitudes of the loadings relating the measures to the construct, as represented by the λ_i in Figure 1. For reflective measures, the magnitudes of these loadings are largely determined by the covariances among the measures, with higher covariances indicating stronger correspondence between each measure and the construct.

The construct validity of formative measures has been addressed in various ways. In some cases, construct validity is viewed as the strength of the relationships between the construct and its measures, as evidenced by the magnitudes of the γ_i in Figure 2 (Bollen, 1989; Diamantopoulos et al., 2008; MacKenzie et al., 2005). In other cases, construct validity is evaluated according to the proportion of variance in the construct attributed to the residual ζ, such that the construct validity of the measures is considered greater when the variance due to the residual is small (Diamantopoulos, 2006; Diamantopoulos et al., 2008; MacKenzie et al., 2005; Podsakoff et al., 2006). Construct validity has also been framed in terms of the relationship between the construct and other variables, drawing from principles of nomological validity and criterion-oriented validity (Bollen & Lennox, 1991; Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005; Podsakoff et al., 2006). Other approaches examine the indirect effects of formative measures on the outcomes of the construct, treating the construct as a mediator of these effects (Bollen & Davis, 2009; Diamantopoulos et al., 2008; Franke et al., 2008; MacKenzie et al., 2005).

The approaches used to evaluate the construct validity of formative measures suffer from several important drawbacks. In particular, estimates of the parameters used as evidence for construct validity are sensitive to the outcomes used to identify the model (Edwards & Bagozzi, 2000; Franke et al., 2008; Heise, 1972; Howell et al., 2007b). One set of outcomes might indicate that the formative measures are strongly related to the construct and explain much of its variance, whereas another set of outcomes could show that the measures have weak relationships with the construct, and much of its variance is explained by the residual ζ. These differences would occur despite the fact that the formative measures are the same in both instances and should therefore demonstrate a consistent degree of construct validity.

Another drawback results from the premise that, because formative measures are conceptually heterogeneous, they are not expected to exhibit the same relationships with outcomes of the formative construct or share the same nomological networks (Franke et al., 2008; MacKenzie et al., 2005; Podsakoff et al., 2006). This heterogeneity undermines the use of parameters linking the construct to its outcomes as evidence for the nomological and criterion-oriented validity of the formative measures. Returning to the model in Figure 4, let us assume previous research shows that x_2 should relate to η_2 but not η_3, whereas x_3 should relate to η_3 but not η_2. Unfortunately, this pattern of relationships cannot be disentangled using the parameter estimates from the model. To elaborate, the effect

of x_2 and η_2 is the product of the parameters relating x_2 to η_1 and η_1 to η_2, or $\gamma_2\beta_1$. However, this product is not uniquely determined by the relationship between x_2 and η_2, because β_1 is also influenced by the relationships of x_1 and x_3 with η_2. Likewise, the effect of x_3 on η_3 is the product $\gamma_3\beta_2$, but β_2 is influenced not only by the relationship between x_3 on η_3 but also by the relationships of x_1 and x_2 with η_3.

The conceptual heterogeneity of formative measures also obscures the interpretation of the construct itself within its hypothesized nomological network. Returning to the model in Figure 4, if x_2 should relate uniquely to η_2 and x_3 should relate uniquely to η_3, then what is the interpretation of the construct η_1 through which these unique relationships are channeled? As observed by Wilcox et al. (2008, pp. 1225-1226) in reference to the unique effects of the x_i on the η_i in a model analogous to that in Figure 4:

> The x_i as formative indicators are not required to have the same consequences, and thus x_1 may relate strongly to η_2 and weakly to η_3, while 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 where the x_i relates differently to the included η_i, (a) substantial lack of fit in the model will be evident, and (b) it will be difficult to interpret the meaning of η_1 either in terms of the (potentially nonexistent) covariance of the η_2 and η_3 or the formative indicators.

As Wilcox et al. (2008) emphasize, a construct that presumably carries relationships linking heterogeneous measures to distinct outcomes becomes a conceptual polyglot with no clear interpretation of its own. Consequently, there is little justification for using relationships associated with such constructs to evaluate the construct validity, nomological validity, or criterion-oriented validity of the construct or its measures.

Causality

Perhaps the most fundamental distinction between reflective and formative measurement concerns the causal relationships between the construct and its measures. As illustrated in Figure 1, reflective measurement models specify constructs as causes of measures. Consistent with this specification, discussions of reflective measurement models often describe the paths relating the construct to its measures in causal terms (Bedeian et al., 1997; Blalock, 1964; Bollen, 1984; Borsboom et al., 2003, 2004; Diamantopoulos et al., 2008; Jarvis et al., 2003; MacCallum & Browne, 1993; Podsakoff et al., 2006). For instance, Podsakoff et al. (2006, p. 210) state that "the defining characteristic of reflective measurement models is that the construct underlies the measures, and changes in the construct are expected to *cause* changes in the measures" (emphasis added). Causality in reflective measurement models is also implied when researchers say the construct determines its measures (Bollen & Lennox, 1991; MacKenzie et al., 2005) or explains the variances and covariances of the measures (MacKenzie et al., 2005; Podsakoff et al., 2006).

In contrast to reflective measurement models, formative measurement models treat measures as causes of constructs, as depicted by the model in Figure 2. Accordingly, most discussions of formative measurement use causal language when describing the relationship between the measures and the construct. In many cases, researchers explicitly state that the measures cause the construct (Bedeian et al., 1997; Blalock, 1964; Bollen, 1984; Cohen et al., 1990; Diamantopoulos et al., 2008; Diamantopoulos & Siguaw, 2006; Jarvis et al., 2003; MacCallum & Browne, 1993). For example, Jarvis et al. (2003, p. 201) assert that, in formative measurement models, "the direction of causality flows from the

measures to the construct." In other cases, researchers indicate that the measures determine the construct (Fornell & Bookstein, 1982; Bollen & Lennox, 1991; Heise, 1972; Podsakoff et al., 2006), which might be intended to convey a softer stance toward causality. For instance, when describing the relationship between formative measures and constructs, Bollen and Lennox (1991, p. 306) explain that they "do not attribute any special significance to the term *cause* other than the fact that the indicators determine the latent variable" (emphasis in original). From this, it seems that *determine* is meant to imply something other than *cause*, but the distinction between these terms is not elaborated, and other researchers have used these terms interchangeably (Diamantopoulos & Siguaw, 2006; MacKenzie et al., 2005). Thus, the relationship between formative measures and constructs is often described as causal, albeit some researchers hesitate to make explicit causal claims.

Although discussions of reflective and formative measurement models both characterize the relationships between constructs and measures as causal, these discussions implicitly rely on different conceptualizations of constructs, measures, and causality. For reflective measurement models, constructs can be meaningfully viewed from a critical realist perspective (Borsboom et al., 2003; Campbell, 1960; Edwards & Bagozzi, 2000; Loevinger, 1957; Messick, 1981), such that constructs refer to real entities that are assessed imperfectly by their measures. The entities that constructs describe are real in the sense that they have the capacity to influence one another, as explained by theoretical models that guide research (Bacharach, 1989; Dubin, 1976; Smith & Hitt, 2005). Measures refer to scores obtained by self-report, interview, observation, or some other method (Edwards & Bagozzi, 2000; Lord & Novick, 1968; Messick, 1995). As noted earlier, the scores that constitute measures contain errors that arise in the measurement process, and multiple measures can be used to offset the imperfections of each individual measure (Campbell & Fiske, 1959; Schwab, 1980).

When constructs and measures are defined in this manner, it is reasonable to conceive of constructs as causes of measures, consistent with reflective measurement models (Borsboom et al., 2003; Edwards & Bagozzi, 2000; Howell et al., 2007b). That is, constructs refer to entities that exist in the real world, independent of attempts by the researcher to measure them. When the researcher arrives on the scene, he or she uses various methods to obtain scores that serve as proxies for the construct. The status of the construct causes certain scores to be realized, and the researcher collects these scores, uses them to form measures, and subjects the measures to some type of analysis. At the time of analysis, the measures are inert, given that they are empirical traces of phenomena that previously occurred. In this sense, causation happened when the measures were collected, at which time the entities referenced by the constructs caused the measures to take on the values obtained by the researcher.

The foregoing account of the causal relationship between constructs and measures seems readily applicable when the construct is a psychological state and measures are collected through self-report, in which case it is reasonable to assume that the psychological state caused the scores reported by the respondent (Blalock, 1964; Borsboom et al., 2003, 2004; Edwards & Bagozzi, 2000; Howell et al., 2007a, 2007b; for an alternative position, see Bagozzi, 2007). Although perhaps less obvious, this account also applies to constructs such as age, education, and income, which are not psychological states in the same sense as perceptions, beliefs, and attitudes. For example, when a researcher measures income, he or she asks respondents to report their income or consult archival records maintained by some agency or organization. The scores obtained by the researcher are the result of the income actually earned by the respondent. As noted earlier, reported income can contain measurement error, such that the scores obtained by the researcher are imperfect representations of income actually earned (Borsboom, 2008; Edwards & Bagozzi, 2000). This type of reasoning can be applied to numerous constructs and measures, provided we firmly keep in mind the meaning of constructs and measures under the critical realist perspective.

With formative measurement, constructs and measures take on different meanings. In many cases, discussions of formative measurement describe the construct not as an entity that exists separately from its measures, but instead as a composite of its measures (Bollen & Lennox, 1991;

Fornell & Bookstein, 1982; Heise, 1972; Jarvis et al., 2003; MacCallum & Browne, 1993; MacKenzie et al., 2005; Podsakoff et al., 2006). For instance, Fornell and Bookstein (1982) indicate that formative measurement is appropriate "when constructs are conceived as explanatory *combinations* of indicators" (p. 442, emphasis in original). Some researchers take this notion a step further, saying that formative measures define the construct (Franke et al., 2008; MacKenzie et al., 2005). For example, MacKenzie et al. (2005, p. 713) describe formative measures as "defining characteristics that collectively explain the meaning of the construct." When viewed in these terms, the constructs involved in formative measurement are not real entities that exist separately from their measures, as would be the case under the critical realist perspective. As explained by MacCallum and Browne (1993, p. 534), "when a construct is defined as having only causal indicators, that construct is not a latent variable in the traditional sense. Rather, it is a linear combination of its observed causal indicators, plus a disturbance term." In light of this, MacCallum and Browne (1993) characterize the construct associated with formative measures not as a latent variable, but instead as a *composite variable*, given that "these causally indicated constructs are, in fact, linear composites of their indicators, plus a disturbance term" (p. 534). Treating constructs as composites of measures represents an operationalist or instrumentalist orientation, whereby constructs are mechanisms for combining measures rather than real entities that exist separately from their measures (Borsboom et al., 2003). When such constructs are ascribed theoretical meaning, the resulting perspective can be characterized as constructivist, whereby the construct is treated as a conceptual entity that has no reality beyond the language used to describe what the measures collectively represent (Borsboom et al., 2003; Howell et al., 2007a).

Discussions of formative measurement also treat measures in terms other than scores collected by the researcher. Instead, formative measures are imbued with causal potency, as if they themselves were constructs that influence other variables, including the construct that serves as the target of the formative measures. For instance, when discussing socioeconomic status as an example of formative measurement, Nunnally and Bernstein (1994, p. 449) say that "People have high socioeconomic status because they are wealthy and/or well-educated; they do not become wealthy or well-educated because they have high socioeconomic status." This reasoning is reflected in other discussions of formative measurement that use socioeconomic status as an illustrative example (Bollen, 1984; Bollen & Lennox, 1991; Heise, 1972; MacCallum & Browne, 1993). Elsewhere, examples of formative measures often describe specific facets that underlie general concepts, such as leadership, trust, organizational justice, rivalry, interorganizational coordination, job enrichment, role stress, organizational commitment, job satisfaction, and job performance (Diamantopoulos & Siguaw, 2006; Jarvis et al., 2003; Law & Wong, 1999; MacKenzie et al., 2005; Podsakoff et al., 2006). In each case, the measures are effectively equated with theoretical facets or dimensions that are linked to the general construct.

The characterizations of constructs and measures underlying formative measurement are problematic. In particular, if a construct is *composed* of its measures, such that the measures are considered parts of the construct, then it makes little sense to treat the relationships between the construct and measures as causal. A prerequisite for causality is that the variables involved refer to distinct entities (Edwards & Bagozzi, 2000), and if one variable is part of another, then their association is a type of part–whole correspondence, not a causal relationship between distinct entities. If a construct is *defined* by its measures, then an implicit form of operationalism is invoked in which the measures and construct are considered one and the same (Bridgman, 1927; Campbell, 1960). This correspondence is inexact when formative measurement models include a residual that is freely estimated (Bollen, 2007), but the residual represents unknown aspects of the construct that are distinct from its measures (Diamantopoulos, 2006) and therefore does not contribute to the meaning of the construct from a conceptual standpoint.

The conceptualization of formative measures as facets that cause general constructs is equally difficult to defend. Returning to the example of socioeconomic status, *measures* of education and

income are not themselves the causal forces that influence socioeconomic status (Borsboom, 2008; Edwards & Bagozzi, 2000). Rather, education is what students experience in schools and other learning venues, and income is what people earn from their employment, investments, and other sources. Education and income are not directly observed by the researcher but instead exist in the socioeconomic world that the researcher wishes to study. For this reason, measures researchers obtain of education and income are not themselves education and income as they actually exist in the real world. Rather, these measures are potentially flawed indicators of education and income. When a researcher asks a respondent to report his or her income, the obtained score is not *itself* the respondent's income. Rather, the report is whatever number the respondent claims is his or her income. The score yielded by this report can contain errors due to recall, record-keeping, efforts to maintain privacy, and so forth. However, the key point is that the score is *not* income itself, but instead *is a potentially flawed measure of the actual income of the respondent*.

In summary, the conceptualizations of constructs, measures, and causality underlying reflective measurement are consistent with a critical realist ontology of constructs and the notion that measures are scores that serve as potentially flawed indicators of real phenomena that researchers attempt to study. In contrast, under formative measurement, constructs overlap with their measures, and measures are equated with theoretical dimensions or facets that have causal potency, disregarding any error that almost certainly exists in the measures. As such, formative measurement signifies an ontology of constructs that could be characterized as constructivist, operationalist, or instrumentalist rather than realist (Borsboom et al., 2003). Which perspective is more defensible? Although opinions on this matter might differ, the viewpoint expressed by Borsboom (2005, p. 153) seems eminently reasonable:

> The only thing all measurement procedures have in common is either the implicit or explicit assumption that there is an attribute out there that, somewhere in the long and complicated chain of events leading up to the measurement outcome, is determining what values the measures will take. This is not some complicated and obscure conception but a very simple idea. If one, however, fails to take it into account, one ends up with an exceedingly complex construction of superficial epistemological characteristics that are irrelevant to the validity issue.

The concerns expressed by Borsboom (2005) and echoed by others (Howell et al., 2007a) argue in favor of reflective measurement, which has decided advantages over formative measurement in terms how constructs, measures, and causality are conceptualized.

Alternatives to Formative Measurement

The shortcomings of formative measurement lead to the inexorable conclusion that formative measurement models should be abandoned. This conclusion creates a dilemma for researchers who might be drawn toward formative measurement models to represent constructs that are defined in terms of multiple dimensions, examine the relationships between general and specific constructs, or summarize relationships involving multiple dimensions in terms of single parameters. Fortunately, the objectives of formative measurement models can be served by alternative models that incorporate reflective measurement and, by doing so, avoid the shortcomings of formative measurement.

One alternative model is shown in Figure 5. This model was suggested by Edwards and Bagozzi (2000) as a modification of the formative measurement model in Figure 2 in which the formative measures x_1, x_2, and x_3 are replaced by the constructs ξ_1, ξ_2, and ξ_3, with x_1, x_2, and x_3 respecified as reflective measures of these constructs. The model in Figure 5 incorporates error into the measurement of each facet, as captured by the uniqueness terms δ_1, δ_2, and δ_3. The model also indicates that the constructs ξ_1, ξ_2, and ξ_3, not the measures x_1, x_2, and x_3, are causes of the construct η, thereby avoiding the assumption that measures cause constructs. To achieve identification, the loadings relating ξ_1, ξ_2, and
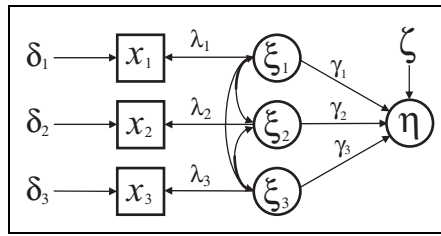
Figure 5. Model that replaces formative measures with facet constructs and single reflective measures

ξ_3 to their measures (i.e., λ_1, λ_2, and λ_3) and the variances of the unique terms δ_1, δ_2, and δ_3 must be fixed to constants, which can be chosen to incorporate measurement error that reflects reliabilities of x_1, x_2, and x_3 deemed appropriate by the researcher (Kline, 2005). Identification also requires that the construct η directly or indirectly causes at least two reflective measures, as is the case for formative measurement models (Bollen & Davis, 2009; Edwards, 2001; MacCallum & Browne, 1993). The model in Figure 5 has been recognized as an alternative to the conventional formative measurement model (Borsboom et al., 2003; Diamantopoulos, 2006; MacKenzie et al., 2005), although it has received little attention in empirical research (for an illustration, see Edwards, 2001).

One shortcoming of the model in Figure 5 is that each of the constructs ξ_1, ξ_2, and ξ_3 has a single indicator, and therefore the loadings and unique variances of these indicators cannot be estimated (Bollen, 1989; Diamantopoulos et al., 2008). This shortcoming is addressed by the model in Figure 6, in which the constructs ξ_1, ξ_2, and ξ_3 are each represented by three reflective measures (Iacobucci, 2010; MacKenzie et al., 2005). With this model, the loadings and unique variances of the measures assigned to ξ_1, ξ_2, and ξ_3 can be estimated, and the drawbacks of representing constructs with single indicators are avoided by using multiple indicators of each ξ_i. Because the model in Figure 6 requires multiple indicators, it is not a feasible alternative to formative measures that have been developed by eliminating redundant items (Diamantopoulos & Siguaw, 2006), given that the indicators assigned to ξ_1, ξ_2, and ξ_3 should be redundant from a conceptual standpoint (DeVellis, 1991). Moreover, for constructs such as education, income, and age, it might not be sensible to use multiple indicators (e.g., asking respondents to repeatedly report their age could be considered needlessly redundant), and a single measure could suffice for constructs that refer to concrete entities (Rossiter, 2002). In such cases, the model in Figure 5 is viable recourse. As with the model in Figure 5, the model in Figure 6 must be supplemented with at least two reflective measures specified as direct or indirect outcomes of the construct η to achieve identification (Bollen & Davis, 2009; Edwards, 2001; MacCallum & Browne, 1993).

When choosing outcome measures to identify models such as those in Figures 5 and 6, there are certain advantages to using direct reflective measures of the construct η rather than measures assigned to outcomes of the construct, such as η_2 and η_3 in Figure 4 (Howell et al., 2007b; Jarvis et al., 2003). For instance, using direct reflective measures allows the construct η to acquire its meaning through measures that describe the construct itself, as opposed to causes or effects of the construct (Edwards & Bagozzi, 2000). Doing so should help avoid ambiguities in the interpretation of the construct and increase the stability of parameters associated with its measurement. Moreover, using measures that refer directly to the construct should help satisfy the proportionality constraints that are effectively imposed on the relationships linking ξ_1, ξ_2, and ξ_3 to the outcomes of η (Franke et al., 2008). Direct reflective measures are also useful when the theory underlying the model treats η as a final criterion with no further outcomes, in which case the model can be identified without adding outcomes that would be considered inappropriate or irrelevant from a conceptual standpoint. Direct reflective measures of η can usually be developed provided that the construct represented by η can be defined in critical realist terms and is conceived separately from the dimensions
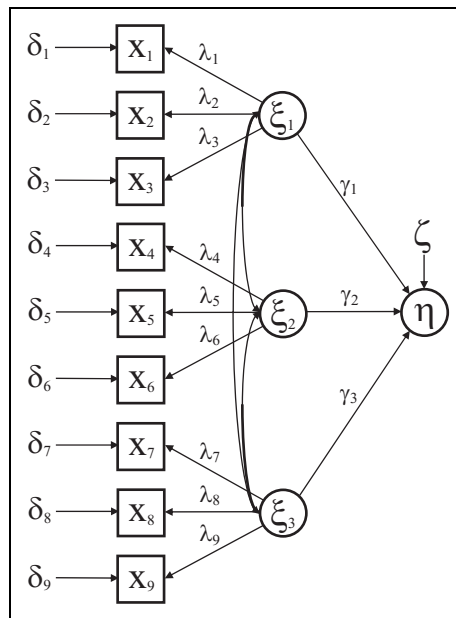
Figure 6. Model that replaces formative measures with facet constructs and multiple reflective measures

represented by the ξ_i. These conditions are satisfied when the construct η is a general concept that is influenced by ξ_i that refer to more specific concepts. For instance, if η is overall job satisfaction and the ξ_i refer to satisfaction with specific job facets, then η can be assessed with general measures such as "In general, I am satisfied with my job" (Edwards & Rothbard, 1999) and the ξ_i can be assessed with measures that describe satisfaction with each job facet. In this manner, the effects of satisfaction with specific job facets on overall job satisfaction could be empirically examined (Ferratt, 1981; Scarpello & Campbell, 1983). This approach might even be applied to socioeconomic status, using measures such as "How high are you up the social ladder?" as reflective measures of η (Borsboom, Mellenbergh, & van Heerden, 2004; Howell et al., 2007b).

If the construct associated with formative measures is defined as nothing more than a combination of its measures, then the construct itself can be eliminated from the model, and the relationships between the ξ_i and other variables in the model can be examined jointly. These relationships can be used to derive multivariate measures of association or interpreted separately using the individual parameters associated with the ξ_i (Edwards, 2001; Heise, 1972). Although this approach abandons the notion that the construct associated with formative measures exists as its own entity, it is reasonable when the construct is composed of or defined by its dimensions (Borsboom et al., 2003; MacCallum & Browne, 1993), in which case the "construct" is nothing more than a label for its dimensions considered collectively (Cohen et al., 1990).

Summary and Conclusion

This article has identified fallacies underlying formative measurement. These fallacies concern the logic and rationale of formative measurement, as evidenced in discussions of such matters as dimensionality, internal consistency, identification, measurement error, construct validity, and causality. By describing these fallacies, I have attempted to demonstrate that formative measurement is not a viable alternative to reflective measurement. As management researchers, we would be better served by carefully developing reflective measures and building models that capture the

multidimensionality of complex concepts (Edwards, 2001; Law, Wong, & Mobley, 1998), as opposed to relying on formative measures with their attendant drawbacks.

Although this article takes a critical stance toward formative measurement models, it is not my intent to denigrate researchers who have written about formative measurement. Indeed, my previous work has treated formative measurement as potentially viable (Edwards & Bagozzi, 2000), and some authors have questioned why my stance toward formative measurement was not more critical (Howell et al., 2007b). These questions are justified, and I have since joined other researchers who recognize fundamental problems with formative measurement (Bagozzi, 2007; Borsboom, 2005; Borsboom et al., 2003; Howell et al., 2007a, 2007b; Iacobucci, 2010; Wilcox et al., 2008). My hope is that researchers who consider formative measurement, including those who frame it as a potential alternative to reflective measurement, will carefully weigh the issues summarized here, as I believe that doing so will lead them to discover the fallacy of formative measurement.

Declaration of Conflicting Interests

Funding

Note

1. This discussion of identification rests on the assumption that the variances and covariances of all measures are nonzero. Otherwise, the models discussed here could suffer from empirical underidentification even if the stated conditions for identification are satisfied.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two step approach. *Psychological Bulletin*, *103*, 411-423.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*, 496-515.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, *12*, 229-237.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement*, *57*, 785-799.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press.

Blalock, H. M. (1968). The measurement problem: A gap between the languages of theory and research. In H. M. Blalock & A. B. Blalock (Eds.), *Methodology in social research* (pp. 5-27). New York: McGraw-Hill.

Bollen, K. A. (1984). Multiple indicators: Internal consistency or no necessary relationship. *Quality and Quantity*, *18*, 377-385.

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

Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12,* 219–228.

Bollen, K. A., & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling*, *16*, 498-522.

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*, 3-22.

Borsboom, D. (2005). *Measuring the mind.* Cambridge, England: Cambridge University Press.

Borsboom, D. (2008). Latent variable theory. *Measurement*, *6*, 25-53.

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.

Bridgman, P. W. (1927). *The logic of modern physics.* New York, NY: Macmillan.

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research*, *5*, 3-52.

Cable, D. M., & Edwards, J. R. (2004). Complementary and supplementary fit: A theoretical and empirical integration. *Journal of Applied Psychology*, *89*, 822-834.

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, *15*, 546-553.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

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.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381-398.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire.* Beverly Hills, CA: Sage.

Costner, H. L. (1969). Theory, deduction, and the rules of correspondence. *American Journal of Sociology*, *75*, 245-263.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Cronbach, L. J., & Meehl, P. C. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Davis, W. R. (1993). The FCI rule of identification for confirmatory factor analysis: A general sufficient condition. *Sociological Methods & Research*, *21*, 403-437.

DeShon, R. (1998). A cautionary note on measurement error corrections in structural equation modeling. *Psychological Methods*, *3*, 412-423.

DeVellis, R. F. (1991). *Scale development: Theories and applications.* Newbury Park, CA: Sage.

Diamantopoulos, A. (2006). The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modelling in Management*, *1*, 7-17.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*, 1203-1218.

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

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

Dubin, R. (1976). Theory building in applied areas. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 17-39). Chicago, IL: Rand McNally.

Dunteman, G. H. (1989). *Principal components analysis.* Newbury Park, CA: Sage.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, *4*, 144-192.

Edwards, J. R. (2003). Construct validation in organizational behavior research. In J. Greenberg (Ed.), *Organizational behavior: The state of the science* (2nd ed., pp. 327-371). Mahwah, NJ: Lawrence Erlbaum.

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., & Rothbard, N. P. (1999). Work and family stress and well-being: An examination of person-environment fit in the work and family domains. *Organizational Behavior and Human Decision Processes*, *77*, 85-129.

Edwards, J. R., Caplan, R. D., & Harrison, R. V. (1998). Person-environment fit theory: Conceptual foundations, empirical evidence, and directions for future research. In C. L. Cooper (Ed.), *Theories of organizational stress* (pp. 28-67). Oxford, UK: Oxford University Press.

Fayers, P. M., & Hand, D. J. (1997). Factor analysis, causal indicators and quality of life. *Quality of Life Research*, *6*, 139-150.

Ferratt, T. W. (1981). Overall job satisfaction: Is it a linear function of facet satisfaction? *Human Relations*, *34*, 463-473.

Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, *19*, 440-452.

Fosnot, C. T. (Ed.). (1996). *Constructivism: Theory, perspectives, and practice.* New York, NY: Teachers College Press.

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

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, *11*, 572-580.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago, IL: University of Chicago Press.

Hauser, R. M., & Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), *Sociological methodology* (pp. 81-117). San Francisco, CA: Jossey-Bass.

Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analysis. *Sociological Methods & Research*, *1*, 147-173.

Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. In E. F. Borgatta, & G. W. Bohrnstedt (Eds.), *Sociological methodology* (pp. 104-129). San Francisco, CA: Jossey-Bass.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods, 12,* 238–245.

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

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*, 90-98.

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*, 199-218.

Jolliffe, I. T. (2002). *Principal components analysis* (2nd ed.). New York, NY: Springer-Verlag.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109-133.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis.* Beverly Hills, CA: Sage.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

Law, K. S., & Wong, C. S. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management*, *25*, 143-160.

Law, K. S., Wong, C. S., & Mobley, W. H. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*, 741-755.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694.

Long, J. S. (1983). *Confirmatory factor analysis: A preface to LISREL.* Beverly Hills, CA: Sage.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

MacCallum, R., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, *114*, 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*, 710-730.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*, 335-361.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, *89*, 575-588.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50*, 741-749.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York, NY: Holt.

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* (Vol. 3, pp. 197-252). Burlington, MA: Elsevier.

Reilly, T. (1995). A necessary and sufficient condition for identification of confirmatory factor analysis models of factoral complexity one. *Sociological Methods & Research*, *23*, 421-441.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, *19*, 305-335.

Scarpello, V., & Campbell, J. P. (1983). Job satisfaction: Are all the parts there? *Personnel Psychology*, *36*, 577-600.

Schwab, D. P. (1980). Construct validity in organizational behavior. In L. L. Cummings, & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 2, pp. 3-43). Greenwich, CT: JAI Press.

Smith, K. G., & Hitt, M. A. (2005). *Great minds in management: The process of theory development.* Oxford, UK: Oxford University Press.

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*, 371-384.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning.* London, England: Falmer.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*, 490-495.

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

Bio

Jeffrey R. Edwards is the Belk Distinguished Professor of Organizational Behavior at the Kenan-Flagler Business School, University of North Carolina. He is past editor of *Organizational Behavior and Human Decision Processes*, past chair of the Research Methods Division of the Academy of Management, and a fellow of the Academy of Management, the American Psychological Association, and the Society for Industrial and Organizational Psychology. His methodological research addresses difference scores, polynomial regression, structural equation modeling, construct validity, and the development and evaluation of theory.