A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research

CHERYL BURKE JARVIS
SCOTT B. MACKENZIE
PHILIP M. PODSAKOFF*

A review of the literature suggests that few studies use formative indicator measurement models, even though they should. Therefore, the purpose of this research is to (*a*) discuss the distinction between formative and reflective measurement models, (*b*) develop a set of conceptual criteria that can be used to determine whether a construct should be modeled as having formative or reflective indicators, (*c*) review the marketing literature to obtain an estimate of the extent of measurement model misspecification in the field, (*d*) estimate the extent to which measurement model misspecification biases estimates of the relationships between constructs using a Monte Carlo simulation, and (*e*) provide recommendations for modeling formative indicator constructs.

It has been more than two decades since Churchill (1979), Bagozzi (1980), Peter (1981), and Anderson and Gerbing (1982), among others, criticized the field of marketing for failing to pay enough attention to construct validity and associated measurement issues. A good example of this concern was expressed by Peter (1981, p. 133), who noted that "a basic goal of social science is to provide theoretical explanations of behavior. In marketing, this goal includes attempts to explain the behavior of consumers, salespersons, and others involved in discipline-related activities. . . . Because construct validity pertains to the degree of correspondence between constructs and their measures, construct validity is a necessary condition for theory development and testing. Thus, it is enigmatic that marketing researchers have given little explicit attention to construct validation, as is well documented in the marketing literature."

This point was echoed by Anderson and Gerbing (1982, p. 453), who noted that "the reason for drawing a distinction between the measurement model and the structural model is that proper specification of the measurement model is necessary before meaning can be assigned to the analysis of the structural model," and by Bagozzi (1981, p. 376), who argued that "convergence in measurement should be considered a criterion to apply *before* performing the causal analysis because it represents a condition that must be satisfied as a matter of logical necessity."

These criticisms, and those of others, led to increasing attention to construct validity in general and more rigorous assessments of the measurement properties of constructs. Studies now commonly report estimates of internal consistency reliability (e.g., Cronbach's alpha) and conduct factor analysis, and provide some evidence of convergent and discriminant validity. The development of latent variable structural equation modeling (SEM) procedures accelerated this trend for two reasons. First, these procedures drew attention to the distinction between the measurement model, which relates the constructs to their measures, and the structural model, which relates the constructs to each other. Second, they also provided much more rigorous tests of construct reliability, convergent validity, and discriminant validity (e.g., Bagozzi 1980; Fornell and Larcker 1981; Gerbing and Anderson 1988).

However, virtually all of this progress in the assessment of constructs and their measures has been based on classical test theory and the assumptions it makes about the relationships between a construct and its indicators. Classical test theory assumes that the variation in the scores on measures of a construct is a function of the true score, plus error.

*Cheryl Burke Jarvis is an assistant professor of marketing at the W. P. Carey School of Business at Arizona State University, P.O. Box 874106, Tempe, AZ 85287-4106 (Cheryl.Jarvis@asu.edu). Scott B. MacKenzie is a professor and Neal Gilliatt Chair of marketing at the Kelley School of Business, Indiana University, 1309 East Tenth Street, Bloomington, IN 47405-1701 (mackenz@indiana.edu). Philip M. Podsakoff is a professor and John F. Mee Chair of management at the Kelley School of Business, Indiana University, 1309 East Tenth Street, Bloomington, IN 47405-1701 (podsakof@indiana.edu). The authors would like to thank Pamela Paxton for advice on the Monte Carlo simulations.

Thus, the underlying latent construct causes the observed variation in the measures (Bollen 1989; Nunnally 1978). This assumed direction of causality—from the latent variable to its measures—is conceptually appropriate in many instances, but not all. Indeed, it was recognized very early on that, for some constructs, it makes more sense conceptually to view causality flowing from the measures to the construct, rather than vice versa (Bagozzi 1981, 1984; Blalock 1964; Fornell and Bookstein 1982). For example, Fornell and Bookstein (1982, p. 441) noted that "the unobserved constructs can be viewed either as underlying factors or as indices produced by the observable variables. That is, the observed indicators can be treated as reflective or formative. Reflective indicators are typical of classical test theory and factor analysis models; they are invoked in an attempt to account for observed variances or covariances. Formative indicators, in contrast, are not designed to account for observed variables. . . . The choice between formative and reflective models, which substantially affects estimation procedures has hitherto received only sparse attention in the literature."

Bollen and Lennox (1991) extended this line of reasoning by distinguishing between two types of measurement models that assume a direction of causality from the measures to the latent construct. One is a principal component model, in which the construct is a perfect linear combination of its measures. The other they called a composite latent construct model, which posited that the construct is a linear combination of its measures, plus error. Perhaps more important, Bollen and Lennox (1991) called attention to the fact that the traditionally used methods for assessing construct reliability and validity are not appropriate for constructs where the direction of causality is posited to flow from the measures to the constructs. This point has been echoed recently by Diamantopoulos and Winklhofer (2001), who suggested improved procedures for developing measures and evaluating these types of constructs, and by Law and Wong (1999), who provided an empirical example showing that the misspecification of the direction of causality between a construct and its measures can lead to inaccurate conclusions about the structural relationships between constructs.

Taken together, these studies demonstrate that some potentially serious consequences of measurement model misspecification exist, and researchers need to think carefully about the direction of causality between constructs and their measures. However, we do not know how often this type of measurement model misspecification occurs in marketing research or about the specific criteria that should be used to distinguish between formative and reflective indicator constructs. Although Diamantopoulos and Winklhofer (2001) touch on the latter point, they do not attempt to develop a comprehensive set of criteria that can be used to decide how a construct should be modeled. Nor did they systematically review the marketing literature to determine how prevalent measurement model misspecification is in the field. Instead, their objective was to develop guidelines for constructing indices based on formative indicators in much the same way that Churchill (1979) did for reflective indicator constructs.

Therefore, the objectives of this article are to (*a*) develop a set of conceptual criteria that can be used to determine whether a construct should be modeled as having formative or reflective indicators, (*b*) determine the extent of measurement model misspecification by comprehensively reviewing measurement model specifications in four top-tier marketing journals, (*c*) conduct a Monte Carlo simulation designed to examine the severity of the estimation bias due to measurement model misspecification, and (*d*) provide recommendations for modeling formative indicator constructs. We begin with a review of the conceptual distinctions between reflective and formative measurement models. These distinctions then are generalized to multidimensional second-order constructs. Next, we document the use of reflective and formative models in the marketing literature and consider the appropriateness of these specifications for the theoretical constructs represented. We then report the results of the Monte Carlo simulation. Finally, we identify the potential reasons why misspecification occurs and conclude with recommendations for how to specify formative indicator constructs.

CONCEPTUAL DISTINCTIONS BETWEEN TYPES OF MEASUREMENT MODELS

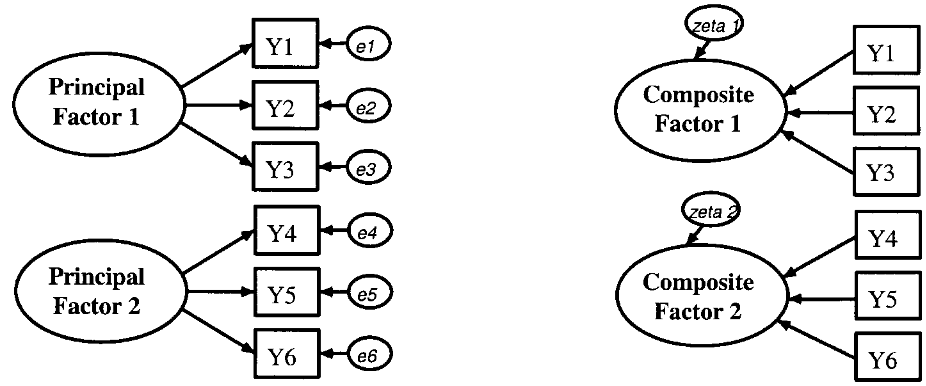
Generally speaking, two different measurement models using multiple indicators of latent constructs have been mentioned in the SEM literature—the principal factor model and the composite latent variable model. Key features of these two models are summarized in figure 1.

The most commonly used latent variable measurement model is the principal factor model, where covariation among the measures is caused by, and therefore reflects, variation in the underlying latent factor. This is indicated in the first column of figure 1, in which each unidimensional construct is represented by a circle with several arrows emanating from it to a set of indicators. The direction of causality is from the construct to the indicators, and changes in the underlying construct are hypothesized to cause changes in the indicators, thus the measures are referred to as reflective (Fornell and Bookstein 1982) or effects (Bollen and Lennox 1991) indicators. In this model, the latent variable influences the indicators, accounting for their intercorrelations. Reflective indicators of a principal factor latent construct should be internally consistent and, because all the measures are assumed to be equally valid indicators of the underlying construct, any two measures that are equally reliable are interchangeable. Thus, although reliability estimates (e.g., Cronbach's alpha) of the set of indicators will be lower if fewer indicators are included in the measurement model, the construct validity is unchanged when a single indicator is removed, because all facets of a unidimensional construct should be adequately represented by the remaining indicators (Bollen and Lennox 1991).

Typical examples of appropriate applications of the reflective indicator model include constructs such as attitudes and

FIGURE 1

SUMMARY OF DIFFERENCES BETWEEN TYPES OF MEASUREMENT MODELS



Principal Factor (Reflective) Model	Composite Latent Variable (Formative) Model
Direction of causality is from construct to measure	Direction of causality is from measure to construct
Measures expected to be correlated (Measures should possess internal consistency reliability)	No reason to expect the measures are correlated (Internal consistency is not implied)
Dropping an indicator from the measurement model does not alter the meaning of the construct	Dropping an indicator from the measurement model may alter the meaning of the construct
Takes measurement error into account at the item level	Takes measurement error into account at the construct level
Construct possesses “surplus” meaning	Construct possesses “surplus” meaning
Scale score does not adequately represent the construct	Scale score does not adequately represent the construct

purchase intention. Attitudes are generally viewed as predispositions to respond in a consistently favorable or unfavorable manner toward an object and are usually measured on multi-item scales with endpoints such as good-bad, like-dislike, and favorable-unfavorable; purchase intentions are typically measured using subjective estimates of how likely-unlikely, probable-improbable, and/or possible-impossible future purchases are perceived to be (e.g., MacKenzie, Lutz, and Belch 1986).

In contrast, in the composite latent variable model, changes in the measures are hypothesized to cause changes in the underlying construct. Thus, this model's measures are referred to as causal (Bollen and Lennox 1991) or formative (Fornell and Bookstein 1982) indicators. Unlike the reflective model, this model does not assume that the measures are all caused by a single underlying construct. Rather, it assumes that the measures all have an impact on (or cause) a single construct. That is, the direction of causality flows from the indicators to the latent construct, and the indicators, as a group, jointly determine the conceptual and empirical meaning of the construct (see col. 2 in fig. 1).

Because some have hypothesized that formative measures influence—rather than are influenced by—the latent construct, they may be correlated, but the model does not assume or require this. Indeed, it would be entirely consistent for formative indicators to be completely uncorrelated. This might be the case where a composite latent construct is represented by mutually exclusive types of behavior. For example, some of the indicators used by Crosby and Stephens (1987) to measure the personal contact of life insurance agents ("I was contacted by my agent who wanted to make changes in this policy to better serve my needs"; "I was contacted by my agent who wanted to sell me more life insurance"; "I was contacted by my agent who wanted to describe new types of policies that had become available"; and "My agent explained why it was a good idea to keep this whole-life policy in force"), may be mutually exclusive. For example, an insurance agent might encourage a customer to keep the current life insurance policy or try to sell the customer a different policy to replace it, but the agent would not do both.

Therefore, internal consistency reliability is not an appropriate standard for evaluating the adequacy of the measures in formative models. Indeed, as noted by Bollen and Lennox (1991, p. 312), "causal indicators are not invalidated by low internal consistency so to assess validity we need to examine other variables that are effects of the latent construct." This would suggest that to assess the validity of composite latent constructs, researchers must pay particular attention to nomological and/or criterion-related validity.

Another implication of the direction of causality in a formative model is that the consequences of dropping one of the indicators are potentially quite serious. It is the opinion of many psychometricians that formative indicators require a census of all concepts that form the construct. Thus, dropping a causal indicator may omit a unique part of the composite latent construct and change the meaning of the variable. Therefore, for formative indicator models, following standard scale development procedures—for example, dropping items that possess low item-to-total correlations—will remove precisely those items that would most alter the empirical meaning of the composite latent construct. Doing so could make the measure deficient by restricting the domain of the construct (Churchill 1979). This is another reason why measures of internal consistency reliability should not be used to evaluate the adequacy of formative indicator models. In addition, multicollinearity among indicators can be a significant problem for measurement model parameter estimates when the indicators are formative, but it is a virtue when the indicators are reflective.

The composite latent variable model includes an error term, as does the principal factor model. However, unlike the principal factor model, error is represented at the construct level rather than at the individual item level. Thus, when using this model, one obtains an estimate of the overall amount of random error in the set of items rather than an estimate attributable to each individual item. This information still permits one to evaluate the reliability of the scale and potentially improve it, but it is somewhat less prescriptive about how the scale can be improved, because the error is associated with the set of items rather than the individual items themselves.

An example of a collection of measured variables used to indicate a composite latent construct might be Singh's (1988) construct called consumer complaint behaviors, which has indicators such as the likelihood of complaining to the store manager, telling friends and relatives about a bad service experience, reporting the company to a consumer agency, or pursuing legal action against the company. In this case, a high likelihood of one particular behavior—say, a complaint to a store manager about poor service—would influence the level of the latent construct, but would not necessarily have an effect on the other measures. One could also conceive of the beliefs construct in the Fishbein and Ajzen (1975) model as having a similar formative measurement structure with each individual belief *x* evaluation component as causing the overall belief construct. The belief construct would be more than simply the sum of the belief by evaluation products because one may not have measured all salient beliefs.

The two types of measurement models have some similarities. Both reflective and formative indicator measurement models possess surplus meaning beyond that captured by the specific items used to measure it. That is to say, "these constructs involve terms which are not wholly reducible to empirical terms; they refer to processes or entities that are not directly observed (although they need not be in principle unobservable); the mathematical expression of them cannot be formed simply by a suitable grouping of terms in a direct empirical equation; and the truth of the empirical laws involved is a necessary but not sufficient condition for the truth of these conceptions" (MacCorquodale and Meehl 1948, p. 104).

Related to this, because composite latent variables and principal factors are more than just a shorthand way of referring to an empirical combination of measures, neither can be adequately represented by a scale score. Using a summed scale score to represent a reflective indicator construct will result in inconsistent structural estimates of the relationships between the construct and other latent constructs because it ignores the effects of measurement error. Using a summed scale score to represent a formative construct will also lead to biased estimates, except in the unlikely event that all of the coefficients relating the measures to the construct are equal to one, and the construct level measurement error is zero.

CRITERIA FOR DISTINGUISHING BETWEEN FORMATIVE AND REFLECTIVE INDICATOR MODELS

Based on the discussion above, it is possible to specify the criteria that researchers might use to distinguish between formative and reflective indicator measurement models. Specifying these criteria is important, because it provides a practical way for researchers to decide on the appropriate measurement model to use in their research. Although there are conceptual discussions of the differences between formative and reflective measurement models (cf. Bollen and Lennox 1991; Diamantopoulos and

Winklhofer 2001; Edwards and Bagozzi 2000), to the best of our knowledge, no comprehensive list of criteria exists to help guide researchers who are struggling with this issue. The criteria are summarized in table 1 in the form of questions that researchers can ask themselves in order to determine what the appropriate relationship is between their measures and their constructs.

Four sets of questions should be used in combination to determine the appropriate measurement model. The first set of questions relate to the direction of causality between the construct and its indicators. For formative measurement models, the direction of causality flows from the measures to the construct, and it flows from the construct to the measures for reflective measurement models. The second set of questions relates to the interchangeability of the indicators. The indicators need not be interchangeable for formative measurement models but should be for reflective measurement models. The third criteria relates to the issue of whether the indicators should covary with each other. Covariation among the indicators is not necessary or implied by formative indicator models, but covariation among the indicators is a necessary condition for reflective indicator models. Finally, the fourth criteria relates to whether all of the measures are required to have the same antecedents and consequences or not. For the reflective indicator model, since all of the indicators reflect the same underlying construct and are assumed to be interchangeable, they should all have the same antecedents and consequences. However, for the formative indicator model, because the measures do not necessarily capture the same aspects of the construct's domain and are therefore not necessarily interchangeable, there is no reason to expect them to have the same antecedents and consequences.

More specifically, a construct should be modeled as having formative indicators if the following conditions prevail: (*a*) the indicators are viewed as defining characteristics of the construct, (*b*) changes in the indicators are expected to cause changes in the construct, (*c*) changes in the construct are not expected to cause changes in the indicators, (*d*) the indicators do not necessarily share a common theme, (*e*) eliminating an indicator may alter the conceptual domain of the construct, (*f*) a change in the value of one of the indicators is not necessarily expected to be associated with a change in all of the other indicators, and (*g*) the indicators are not expected to have the same antecedents and consequences. On the other hand, a construct should be modeled as having reflective indicators if the opposite is true and the conditions shown in the last column in the table are satisfied.

Of course, it is possible that researchers may have difficulty in answering some of the questions, or the answers may be contradictory because the construct has not been adequately defined. In such cases, further refinement of the conceptualization of the construct may be needed. This may require researchers to clarify the construct's domain, evaluate whether all the indicators are within that domain, and consider the measures' relationships to other constructs.

TABLE 1

DECISION RULES FOR DETERMINING WHETHER A CONSTRUCT IS FORMATIVE OR REFLECTIVE

	Formative model	Reflective model
1. Direction of causality from construct to measure implied by the conceptual definition	Direction of causality is from items to construct	Direction of causality is from construct to items
Are the indicators (items) (*a*) defining characteristics or (*b*) manifestations of the construct?	Indicators are defining characteristics of the construct	Indicators are manifestations of the construct
Would changes in the indicators/items cause changes in the construct or not?	Changes in the indicators should cause changes in the construct	Changes in the indicator should not cause changes in the construct
Would changes in the construct cause changes in the indicators?	Changes in the construct do not cause changes in the indicators	Changes in the construct do cause changes in the indicators
2. Interchangeability of the indicators/items	Indicators need not be interchangeable	Indicators should be interchangeable
Should the indicators have the same or similar content? Do the indicators share a common theme?	Indicators need not have the same or similar content/indicators need not share a common theme	Indicators should have the same or similar content/indicators should share a common theme
Would dropping one of the indicators alter the conceptual domain of the construct?	Dropping an indicator may alter the conceptual domain of the construct	Dropping an indicator should not alter the conceptual domain of the construct
3. Covariation among the indicators	Not necessary for indicators to covary with each other	Indicators are expected to covary with each other
Should a change in one of the indicators be associated with changes in the other indicators?	Not necessarily	Yes
4. Nomological net of the construct indicators	Nomological net for the indicators may differ	Nomological net for the indicators should not differ
Are the indicators/items expected to have the same antecedents and consequences?	Indicators are not required to have the same antecedents and consequences	Indicators are required to have the same antecedents and consequences

MULTIDIMENSIONAL FORMATIVE AND REFLECTIVE INDICATOR CONSTRUCTS

The above criteria focus on the relationships between measures and first-order latent constructs. However, it is important to note that conceptual definitions of constructs are often specified at a more abstract level, which sometimes include multiple formative and/or reflective first-order dimensions. For example, a single multidimensional construct might have one type of measurement model relating its measures to its first-order components and a different measurement model relating its components to the underlying second-order factor. Of course, some researchers might argue that a construct must be conceptually and empirically unidimensional to be meaningful. However, such a view is often inconsistent with the way constructs are defined in the field. We would argue that whether a construct is viewed as unidimensional or multidimensional may depend on the level of abstraction used to define the construct. For example, job satisfaction is frequently defined as being composed of several different facets, including satisfaction with one's pay, coworkers, supervisor, opportunities for advancement, and so forth. Although one can look at each facet as being a separate construct, at a more abstract level, they are all integral parts of a person's job satisfaction. Indeed, we think this kind of abstract multidimensional construct definition is quite common in the marketing literature.

Figure 2 illustrates four different possible combinations in second-order factor models. The four main types of second-order models are derived from the fact that (*a*) a first-order construct can have either formative or reflective indicators, and (*b*) those first-order constructs can, themselves, be either formative or reflective indicators of an underlying second-order construct. The combination of these possibilities produces the models shown in figure 2 (Types I–IV). In addition, it is also possible for a model to contain a mixture of formative and reflective indicators. Mixed models could result either because some of the first-order dimensions are formative indicators of the second-order construct and some are reflective indicators of the second-order construct or because some of the first-order dimensions themselves have formative indicators and some have reflective indicators.

Interestingly, the only kinds of second-order factors that have been recognized in the literature appear to be those that have first-order factors as reflective indicators (Type I or Type III). The historical roots of Type I models can be traced back to the work of Bentler and Weeks (1980) and Gerbing and Anderson (1984). This model (shown in fig. 2, upper-left panel) posits a series of first-order latent factors with reflective indicators and also that these first-order factors are themselves reflective indicators of an underlying second-order construct. This type of second-order model has been called a total disaggregation second-order factor model by Bagozzi and Heatherton (1994) in their research on self-esteem.

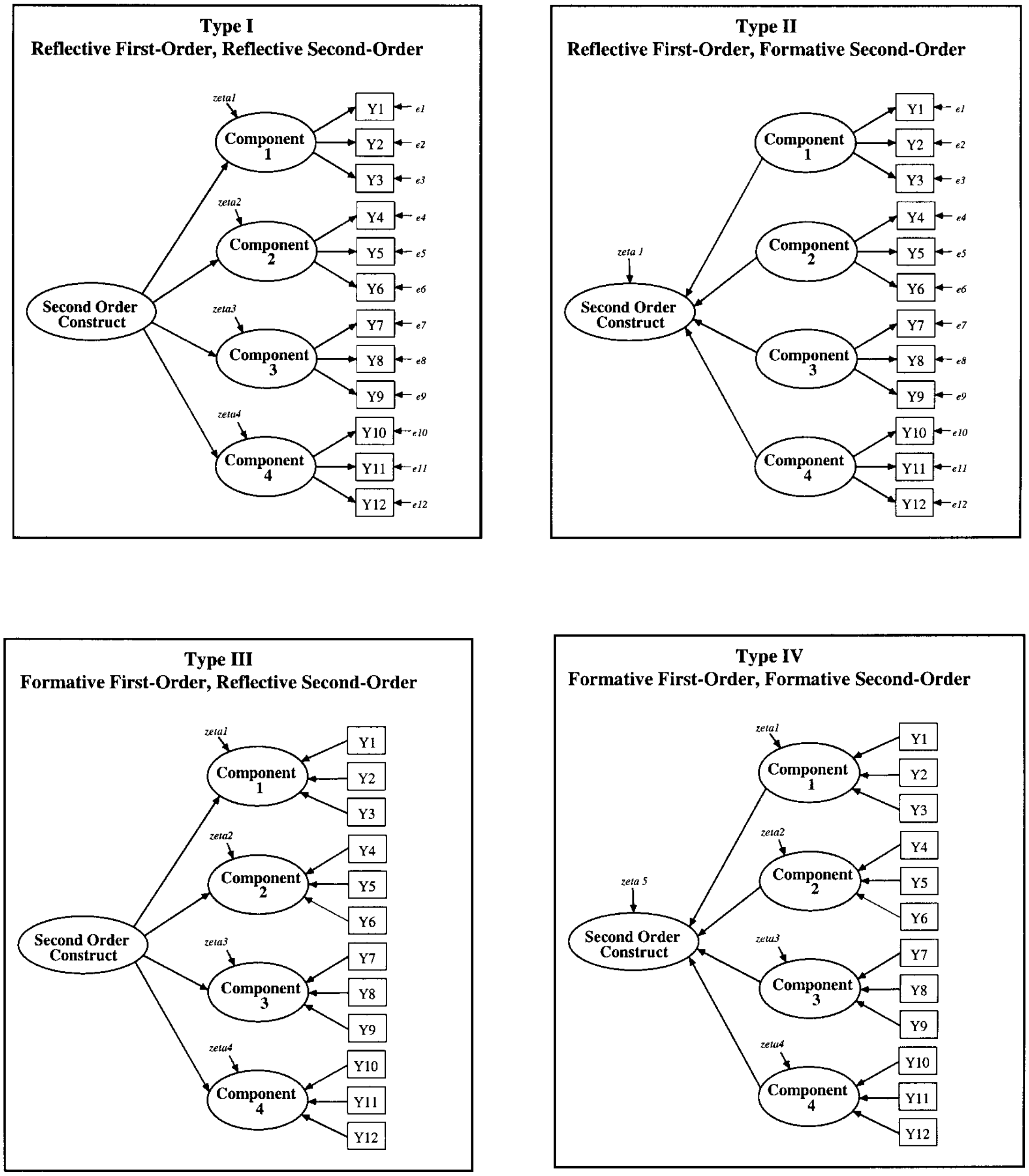
Type III second-order factor models (shown in fig. 2, lower-left panel) have first-order factors as reflective indicators like Type I models, but the first-order dimensions themselves have formative rather than reflective indicators. Although this kind of second-order factor model has not been explicitly recognized in the literature, Reilly's (1982) family social status construct may be an example of this type of model. In Reilly's research, four well-established scales for measuring social status were used as reflective indicators of an underlying factor called family social status. However, each of these social status scales is composed of several formative indicators. For example, one of the reflective indicators of the second-order family social status construct was Warner's Index of Status Characteristics, which is itself a first-order factor composed of ratings of occupation, source of income, dwelling type, and neighborhood quality. Since occupation, source of income, dwelling type, and neighborhood quality are clearly not interchangeable, have potentially different antecedents and consequences, and would not necessarily covary with each other, they should be viewed as formative indicators of status. Thus, consistent with the Type III model shown in the lower-left panel of figure 2, family social status might be thought of as a second-order construct with four first-order factors as reflective indicators, each of which is composed of multiple formative indicators.

Another type of factor model is one where the second-order factor has first-order factors as formative indicators and the first-order factors themselves have reflective indicators (Type II model shown in upper-right panel). Such a model might be appropriate for the multidimensional composite construct of noncontingent influence attributions examined by John (1984). John measured three different types of noncontingent influence (expert, referent, and legitimate) using reflective measures. These three first-order dimensions were then modeled as being related to what is really a second-order noncontingent power construct. Although noncontingent power was modeled as having expert, referent, and legitimate power as reflective indicators, we believe it makes more sense to say that channel members have higher levels of noncontingent influence because they possess expert, referent, and legitimate power than it does to say that they have expert, referent, and legitimate power because they possess noncontingent power. Clearly, expert, referent, and legitimate power are relatively independent sources of influence that, together, all share the characteristic of being noncoercive.

Still another multidimensional factor model is the Type IV second-order construct, which has formative indicators for both the first- and second-order factors. This model might be an appropriate specification for Crosby, Evans, and Cowles's (1990) similarity construct. These authors conceived of overall similarity as being a function of appearance, lifestyle, and status similarity. Each of the three dimensions of overall similarity was measured by a series of formative measures. In our view, these three first-order dimensions of similarity (appearance, status, and lifestyle) are formative indicators of the second-order overall similarity construct, because together they determine the overall level of similarity rather than result from it.

FIGURE 2

ALTERNATIVE SECOND-ORDER FACTOR SPECIFICATIONS



REVIEW OF THE MARKETING LITERATURE

Methodology

Our review up to this point has attempted to clarify the distinctions between formative and reflective indicator measurement models, and the criteria that could be used to decide which model is appropriate to use in a specific instance. Thus, a logical next step would be to review the research literature, apply the criteria, and thereby determine how prevalent measurement model misspecification is in the marketing literature. To our knowledge, no one has ever attempted to systematically evaluate the appropriateness of measurement model specifications used in our field. Such a review would provide insights into not only the extent of

TABLE 2

PERCENTAGE OF CORRECTLY AND INCORRECTLY SPECIFIED CONSTRUCTS BY JOURNAL

	Overall			*JMR*			*JM*		
	Should be reflective	Should be formative	Total	Should be reflective	Should be formative	Total	Should be reflective	Should be formative	Total
Modeled as reflective	810	336	1,146	319	120	439	368	187	555
	(68)	(28)	(96)	(70)	(26)	(96)	(63)	(32)	(95)
Modeled as formative	17[a]	29[b]	46	7[a]	10[b]	17	10[a]	18[b]	28
	(1)	(3)	(4)	(2)	(2)	(4)	(2)	(3)	(5)
Total	827	365	1,192	326	130	456	378	205	583
	(69)	(31)	(100)	(72)	(28)	(100)	(65)	(35)	(100)

	JCR			*MS*		
	Should be reflective	Should be formative	Total	Should be reflective	Should be formative	Total
Modeled as reflective	107	22	129	16	7	23
	(82)	(17)	(99)	(70)	(30)	(100)
Modeled as formative	0[a]	1[b]	1	0[a]	0[b]	0
	(0)	(1)	(1)	(0)	(0)	(0)
Total	107	23	130	16	7	23
	(82)	(18)	(100)	(70)	(30)	(100)

NOTE.—*JMR* = *Journal of Marketing Research*, *JM* = *Journal of Marketing*, *JCR* = *Journal of Consumer Research*, and *MS* = *Marketing Science*. Items shown in parentheses are percentages.

[a]Indicates that although authors correctly identified the construct as reflective, they modeled it using partial least squares (PLS), which assumes a formative measurement model.

[b]Indicates that although authors correctly identified the construct as formative, they modeled it using PLS or scale scores—neither of which estimate construct-level measurement error.

measurement model misspecification but also would reveal which constructs have been most frequently misspecified. Therefore, in this section, we will report the results of a review of measurement model specifications in the top four marketing journals by examining every construct for which a confirmatory factor analysis has been reported.

The *Journal of Consumer Research* (*JCR*), *Journal of Marketing* (*JM*), *Journal of Marketing Research* (*JMR*), and *Marketing Science* (*MS*) were selected as representative of the best journals in the marketing literature. These four journals were searched for the 24-year period from 1977 through 2000 (1982–2000 for *MS*) to identify all empirical applications of latent variable SEM or confirmatory factor analysis. Methodological papers in which only simulated data were analyzed, or actual data were analyzed for illustrative purposes only, were not considered. Similarly, conventional exploratory factor analysis models, path analysis, and other structural models estimated by regression methods (e.g., models estimated by two-stage least squares), nonlinear structural models, and observed variable models were excluded from the analysis. Therefore, the database analyzed here consists of articles incorporating either confirmatory factor models or latent variable SEM. Using these criteria, we identified 178 articles containing 1,192 constructs modeled as latent factors with multiple indicators.

The classification of the constructs followed a multistep procedure. The first step was for each of the three coders (the authors) to independently read the articles, identify those constructs with multiple measures, and determine how their measurement models were specified. Next, using the criteria in table 1, each construct was classified as formative or reflective. A construct was classified as formative if it clearly met the majority of the criteria in the second column of this table and was classified as reflective if it met most of the criteria in the third column of this table. In those cases where all three coders agreed that the construct met the criteria for either a formative or a reflective measurement model, the construct was assigned to that measurement model category. In instances where the coders disagreed about the extent to which the construct met the various criteria, the points of disagreement were discussed until a consensus was reached. In approximately 14% of the cases, it was difficult, if not impossible, to tell whether the construct should have been modeled as a formative or as a reflective construct. In the majority of these instances (12%), this happened because the authors failed to provide a complete set of items or construct definition, although there were a few instances (2%) when the construct of interest met some of the criteria for a formative scale but other criteria for a reflective scale. However, in all of these cases the benefit of the doubt was given to the authors, and the construct was categorized as being correctly modeled as specified by the author. This procedure was adopted with the goal of making the estimates of misspecification as conservative as possible.

Results

Table 2 summarizes our findings across all journals. Several interesting patterns emerge. First, the results indicate that, overall, 71% (68% + 3%) of the latent constructs with multiple measures found in the top-four marketing journals during the past 24 years were correctly modeled, and 29% (28% + 1%) were incorrectly modeled. By far, most of those that were correctly modeled (810 out of a total of 839) were reflective constructs correctly modeled as having reflective measures, while the remainder (29 of 839) were formative constructs correctly modeled as having formative measures. In contrast, the majority of constructs that were incorrectly modeled (336 out of a total of 353) were formative constructs incorrectly modeled as having reflective measures.

Table 2 also reports the results separately for each of the four journals reviewed. The journals differed on the extent to which multiple indicator latent variables were examined. The *JM* had the greatest number of latent variables (583), and *MS* had the least (23). Although this is partially due to the fact that *MS* only began publication in 1982, it cannot account for a discrepancy of this magnitude. It seems more likely that this difference arises more from the methodological focus of *MS*, its relatively higher proportion of purely analytic studies, and the modeling preferences of the journal's most frequent contributors and sponsoring association (Institute for Operations Research and the Management Sciences [INFORMS]). However, even these factors seem insufficient to account for a discrepancy of this magnitude.

In addition, it is clear from table 2 that a substantial amount of measurement model misspecification was present in all four journals. The degree of misspecification ranged from a low of 17% for all constructs in the case of *JCR*, to a high of 34% for all constructs in *JM*, with *JMR* at 28% and *MS* at 30%. Because measurement model misspecification can result in both Type I and Type II errors, these findings suggest that measurement model misspecification is a serious problem, even in the best journals in the field. Finally, table 2 also indicates that the majority of the constructs in the four journals were modeled as having reflective measures and that most of the errors in measurement model specification resulted from the use of a reflective measurement model for constructs that should have been formatively modeled. However, despite these general similarities, there were differences across the journals with respect to the percentage of constructs correctly modeled: 83% for *JCR*, 66% for *JM*, 72% for *JMR*, and 70% for *MS*. This may be a result of differences in the types of constructs typically examined in the journals. For example, *JCR* has a greater proportion of studies investigating psychological constructs than managerial constructs, and the opposite is true for *JM*, *JMR*, and *MS*. An examination of the data suggests that the conceptualizations of managerial constructs in the marketing literature are more frequently formative in nature, often defined as combinations of relatively independent factors that together determine the level of the latent construct. Although this can be true of psychological constructs too, it seems to be more likely to be true of constructs in the managerial domain (e.g., market dynamism, noncoercive power, job performance, strategic performance, unfair trade practices, output controls, competitive and market intelligence) than it is of psychological constructs (e.g., brand attitude, purchase intentions, and ad-evoked feelings).

Exemplars of Constructs with Formative Indicators

Table 3 reports a sample of constructs from our review of the literature that were identified as first-order or second-order constructs with formative indicators. The columns in the table contain the construct names, the nature of the construct (i.e., first-order or second-order), the study in which the construct was examined, and examples of the indicators used to measure the construct. In the case of the second-order constructs, the indicators were typically scale scores representing first-order dimensions of the construct. All of the constructs in this table satisfy the decision rules for formative constructs discussed in table 1.

For illustrative purposes, an attempt was made to select constructs from a variety of different topic areas, including consumer research (e.g., belief structure, emotions, perceived risk, and socioeconomic status); marketing channels (e.g., adaptations made by customer, power source, and satisfaction with channel partner); environmental factors (e.g., supply market dynamism and unfair trade practices); job performance and related behaviors (e.g., sales performance, marketing resources and skills, and marketing orientation); and job attitudes and role perceptions (e.g., job satisfaction, role ambiguity, and role conflict). However, no attempt was made to be exhaustive. Instead, the goal was to demonstrate the appropriateness of a formative indicator model for a broad range of constructs.

HOW BIG A PROBLEM IS MODEL MISSPECIFICATION?

The preceding review of the marketing literature demonstrates that measurement model misspecification is fairly pervasive among published research studies. However, no one has yet demonstrated the extent to which different types of measurement model misspecification influence the estimates of the measurement and structural model parameters. This is important because any bias in the estimates produced by the misspecification could affect the conclusions about the theoretical relationships among the constructs that are drawn from the research. Therefore, a Monte Carlo simulation was conducted to investigate this issue. More specifically, the simulation examined the empirical consequences of inappropriately applying a reflective measurement model to a truly formative construct, since the review of the marketing literature indicates that this is by far the most common type of measurement model specification error.

Figure 3 summarizes the models tested in the Monte Carlo simulation conditions. The simulation conditions varied on whether (1) the measurement model of the focal construct was correctly specified as having formative indicators (e.g., as indicated in correctly specified models 1 and 2 in fig. 3)

TABLE 3

EXEMPLARS OF FIRST-ORDER AND SECOND-ORDER MARKETING CONSTRUCTS WITH FORMATIVE INDICATORS

Construct name	Nature of construct (first or second order)	Studies	Examples of indicators used to measure construct
Adaptations made by the customer	1st	Hallen, Johanson, and Seyed-Mohamed (1991)	Has this customer: Modified his final product in order to suit your product? Adapted certain production procedures as a consequence of using your product? Modified his production schedules in order to meet your delivery capacity?
Anxious emotional involvement	2d	Rose (1999)	Fostering dependence Excluding outside influence Child's independence
Attitudinal orientation	2d	John (1984)	Types of attitudinal orientation: Cognitive Affective Conative
Belief structure	2d	Ryan (1982)	Belief structure: Price Taste Decay Breath, etc.
Beliefs	2d	Shimp and Kavas (1984)	Encumbrances Inconveniences Rewards
Decision-making uncertainty	2d	Achrol and Stern (1988)	Uncertainty of available information Predictability of consequences Degree of confidence
Encumbrances	1st	Shimp and Kavas (1984)	Subscribing to extra media Purchasing nonpreferred brands Shopping at nonfavorite stores
Health behavioral control	1st	Moorman and Matulich (1993)	It's difficult to reduce my sodium intake[a] It's too hard for me to exercise three days a week[a] I find it hard to get enough rest and sleep[a] Going for an annual physical exam is easy for me
Helping behavior	2d	Podsakoff and MacKenzie (1994)	Altruism Courtesy Cheerleading Peace keeping
Imagery quantity/ease	2d	Bone and Ellen (1992)	Imagery ease Imagery quantity
Job performance/sales performance	1st	Cravens et al. (1993)	Producing a high market share for your company Making sales of those products with the highest profit margins Quickly generating sales of new company products
Job satisfaction (facet based)	1st	Hartline and Ferrell (1996)	How satisfied you are with your: Supervisor(s) Organization's policies Organization's customers, etc.
Market orientation	2d	Kohli, Jaworski, and Kumar (1993)	Intelligence generation Intelligence dissemination Responsiveness
Marketing resources and skills	1st	Calantone, Schmidt, and Song (1996)	Our marketing research skills and people were more than adequate Our advertising and promotion resources and skills were more than adequate Our management skills were more than adequate
Negative emotion	2d	Murry and Dacin (1996)	Anger Fear Discouraged
Negative role of price	2d	Lichtenstein, Ridgway, and Netemeyer (1993)	Value consciousness Price consciousness Coupon proneness
Normative belief structure	2d	Ryan (1982)	Normative belief structure: Dentist Children Husband

TABLE 3 (*Continued*)

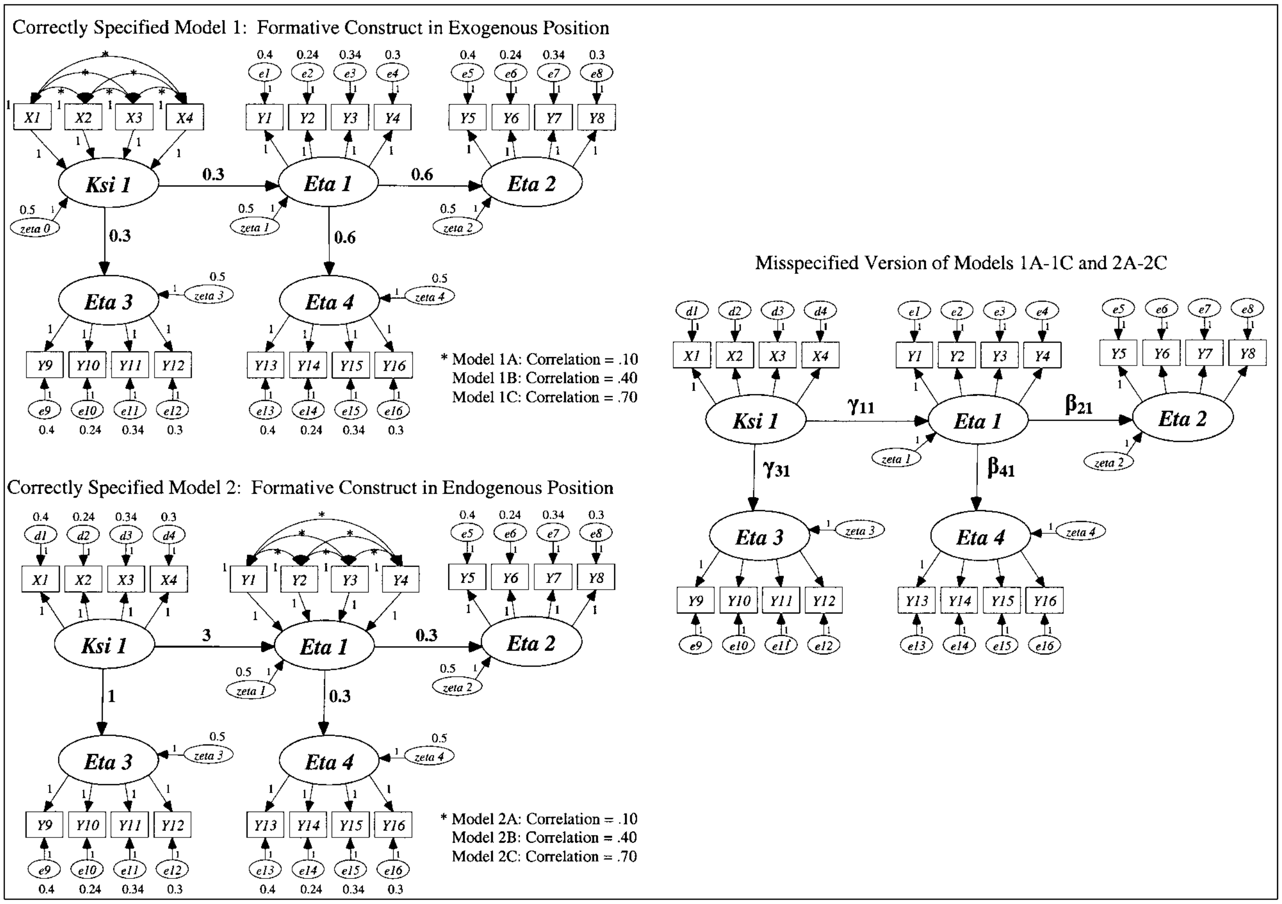
Construct name	Nature of construct (first or second order)	Studies	Examples of indicators used to measure construct
Perceived risk	2d	Srinivasan and Ratchford (1991)	Types of perceived risk: Financial Performance Physical Convenience
Positive emotion	2d	Murry and Dacin (1996)	Contented Happy
Positive experience	2d	Srinivasan and Ratchford (1991)	Experience with previous manufacturer or dealer Experience with previous car
Qualitative power source	2d	Gaski (1986)	Types of qualitative power source: Expert Legitimate Referent
Role ambiguity	1st	Michaels, Day, and Joachimsthaler (1987)	With respect to yourself and your job, please indicate your agreement with each statement listed below: Clear, planned goals and objectives exist for my job[a] I do not know if I utilize my time properly on my job I know what my purchasing responsibilities are[a]
Role conflict	1st	Michaels et al. (1987)	I have to buck rules or policies in order to carry out assignments My purchasing workload seems to be at about the right level[a] I receive incompatible requests from two or more people
Sales organization effectiveness	2d	Cravens et al. (1993)	Financial effectiveness Customer satisfaction
Satisfaction with channel partner	1st	Mohr, Fisher, and Nevin (1996)	How satisfied are you with the following aspects of the relationship with this manufacturer: Cooperative advertising Assistance in managing inventory Profit on sales of manufacturer's product, etc.
Similarity	2d	Crosby et al. (1990)	Types of similarity: Appearance Status Lifestyle
Status similarity	1st	Crosby et al. (1990)	Rating of agent's: Education level Income level Social class
Supply market dynamism	1st	Cannon and Perreault (1999)	How significant are changes in: Pricing Product features and specifications Vendor support services, etc.
Trust	2d	Siguaw, Simpson, and Baker (1998)	Credibility Benevolence
Unfair trade practices	1st	Achrol and Stern (1988)	Bait-and-switch tactics High-pressure sales tactics Price promotions on unavailable items
Utilitarian benefit	2d	Chandon, Wansink, and Laurent (2000)	Savings Quality Convenience Value expression

[a]Reverse coded.

or incorrectly specified as having reflective indicators (as indicated in the misspecified version of models 1A–1C and 2A–2C in fig. 3), (2) the focal construct was an exogenous construct (correctly specified model 1) or an endogenous construct (correctly specified model 2) in the structural model, and (3) the item intercorrelations of the focal construct were weak (.1), moderate (.4), or strong (.7). Manipulating the measurement model specification for the focal construct obviously allowed us to investigate the consequences of misspecification. Manipulating the focal construct's position within the model permitted us to test the effects of misspecification on structural parameters leading to the misspecified construct, as well as those structural paths emanating both directly and indirectly from the misspecified construct. Finally, the manipulation of the magnitude of the correlations among the indicators of the formative construct allowed us to test the significance of the effects of misspecification across a variety of situations, including at least

FIGURE 3

MODELS FOR SIMULATION



one in which the item intercorrelations are high enough (e.g., .7) that it might appear that a reflective indicator model is appropriate.

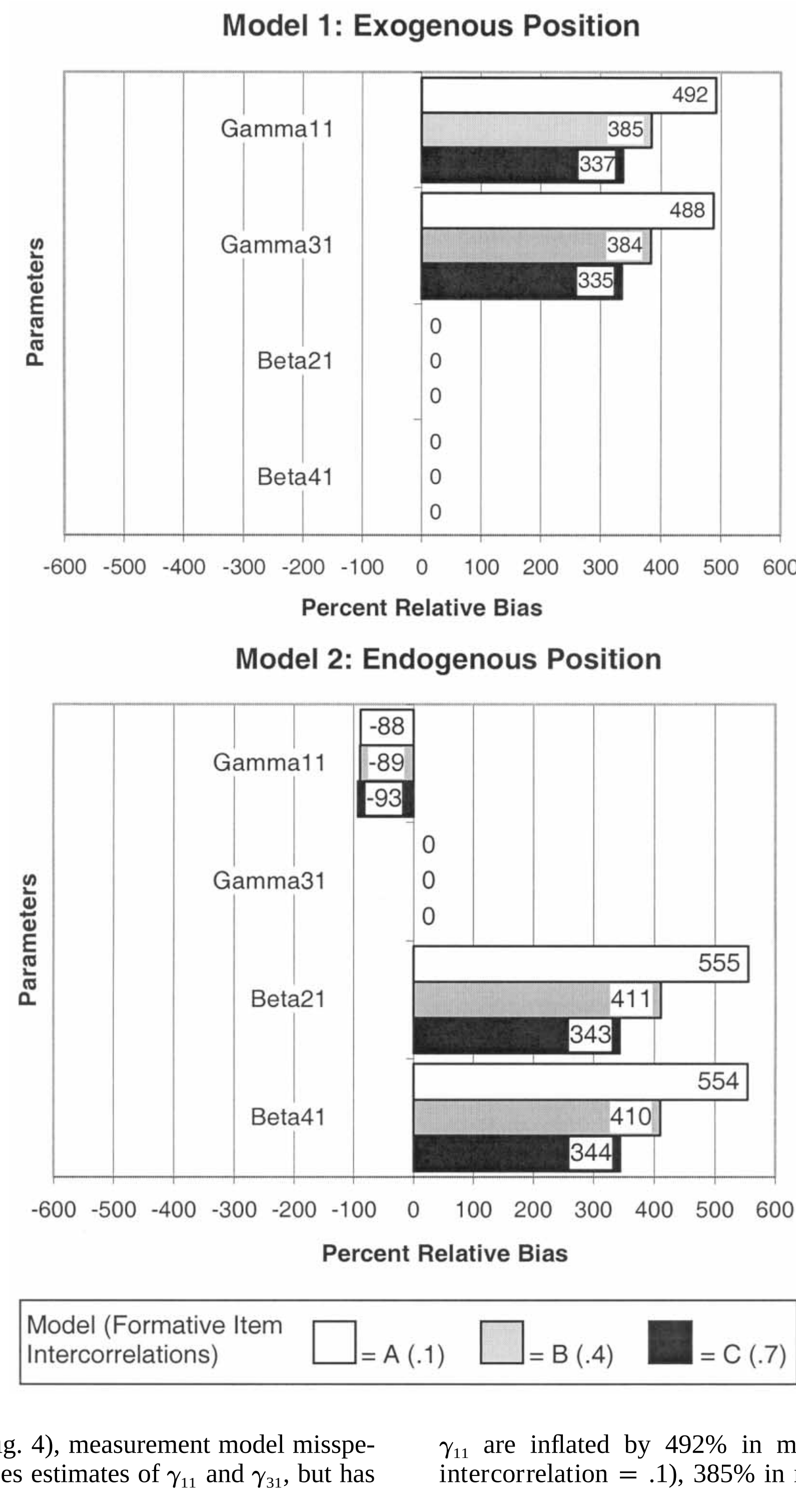
The complexity of the models was chosen to be representative of those typically found in marketing, with five latent variables and four observed items per factor (cf. Baumgartner and Homburg 1996). In addition, the values chosen for the parameters of the correctly specified models were selected to meet several criteria as outlined by Paxton et al. (2001). First, they were chosen to reflect values that are similar in magnitude to those commonly encountered in the marketing literature. Second, the R^2 values produced by the chosen coefficients were also representative of those typically found in consumer behavior and marketing research. Third, the parameters in the model were statistically significant in all populations. Finally, in order to be consistent with the average amount of random and systematic measurement error reported in published marketing studies as reviewed by Cote and Buckley (1987), the item error variances were set to average 32% for each construct. The chosen population values are shown in figure 3.

After the population values were set and the population covariance matrices calculated, a Monte Carlo simulation was conducted using EQS 5.7b. (The model specification and covariance matrices are available from the first author.) The simulation generated raw data sets of a sample size of 500 from each of the two population covariance matrices. These data sets were then fit to both the correctly specified models and the misspecified models, generating parameter estimates and fit statistics for each replication. As recommended by Paxton et al. (2001), only data sets with converged and proper solutions were used in the analysis. A total of 500 data sets meeting this requirement were generated for each of the six population conditions in the simulation (e.g., correctly specified models 1A–1C and 2A–2C).

Figure 4 reports the percent bias in the unstandardized structural parameter estimates for the misspecified model relative to the population value for each of the six correct models. We define percent relative bias as 100 times the difference between the parameter estimate and its population value divided by the population value. When the formatively indicated construct is in the exogenous position in the model

FIGURE 4

PERCENT RELATIVE BIAS IN UNSTANDARDIZED STRUCTURAL PARAMETER ESTIMATES



(model 1, top panel in fig. 4), measurement model misspecification positively biases estimates of γ_{11} and γ_{31}, but has no effect on β_{21} or β_{41}. The degree of bias in the affected parameters is negatively related to the magnitude of the formative item intercorrelations. For example, estimates of γ_{11} are inflated by 492% in model 1A (formative item intercorrelation = .1), 385% in model 1B (formative item intercorrelation = .4), and 337% in model 1C (formative item intercorrelation = .7). Estimates of γ_{31} are positively biased 488% in model 1A, 384% in model 1B, and 335%

in model 1C. Analysis of variance shows that the differences between the average unstandardized parameter estimates and the population values when the focal construct is exogenous are statistically significant for both γ_{11} ($F = 36,343.61$, $df = 1$, $p < .001$) and $\gamma 31$ ($F = 36,461.06$, $df = 1$, $p < .001$), but not β_{21} ($F = 1.09$, $df = 1$, $p > .05$) or β_{41} ($F = .01$, $df = 1$, $p > .05$). Moreover, it is worth noting that all of the individual Bonferroni pairwise comparisons across the three item intercorrelation conditions are significant for γ_{11} and γ_{31} (all p's $< .001$), but not for β_{21} or β_{41} (all p's $> .05$).

When the formatively indicated construct is in an endogenous position in the model (correctly specified model 2 in fig. 3), measurement model misspecification suppresses estimates of γ_{11}, negatively biasing the estimates by 88% in model 2A (formative item intercorrelation = .1), 89% in model 2B (intercorrelation = .4), and 93% in model 2C (intercorrelation = .7), but has no effect on estimates of γ_{31}. Estimates of β_{21} and β_{41} are both significantly inflated, again in a pattern negatively related to the magnitude of the item intercorrelations, with β_{21} biased by 555% in model 2A, 411% in model 2B, and 343% in model 2C; and β_{41} biased by 554% in model 2A, 410% in model 2B, and 344% in model 2C. Again, ANOVAs show that the differences between the average estimates of the misspecified models and the population values are statistically significant for γ_{11} ($F = 24,202.12$, $df = 1$, $p < .001$), β_{21} ($F = 25,147.50$, $df = 1$, $p < .001$), and β_{41} ($F = 26,214.16$, $df = 1$, $p < .001$), but not for γ_{31} ($F = .17$, $df = 1$, $p > .05$). All the Bonferroni pairwise comparisons of the three-item intercorrelation conditions are significant for β_{21} and β_{41} (all p's $< .001$), two of the three comparisons (.4 vs. .7, and .1 vs. .7) are significant for γ_{11}, and none are significant for γ_{31} (all p's $> .05$).

Table 4 shows the goodness-of-fit indices for the correctly and incorrectly specified models. As expected, all of the correctly specified models fit the data adequately, according to every one of the fit indices (e.g., nonsignificant chi-square statistics, goodness-of-fit indices [GFI] $> .90$, comparative fit indices [CFI] $> .95$, standardized root mean square residual [SRMR] $< .08$, root mean square error of approximation [RMSEA] $< .05$). So did all of the misspecified models, according to the CFI, SRMR, and the RMSEA indices. Only the chi-square and the GFI were able to detect the measurement model misspecification. More specifically, the chi-square and GFI identified all but one (model 1C) of the six misspecified models. This suggests that when the measurement model is misspecified, researchers may have difficulty detecting it based on the overall goodness-of-fit indices. Although the chi-square statistics are significant, they probably would have been discounted with a sample size of 500 based on the well-recognized dependence of the chi-square on sample size. Therefore, only the GFI would have indicated a lack of fit, and the overall pattern would have suggested that the models fit the data adequately. Indeed, if researchers relied on only the two indices (i.e., the CFI and SRMR) identified by Hu and Bentler (1999) in their simulation as being the best at balancing Type I and Type II errors, they would have been led to conclude that the misspecified models fit the data.

Our simulation results provide strong evidence that measurement model misspecification of even one formatively measured construct within a typical structural equation model can have very serious consequences for the theoretical conclusions drawn from that model. The entire model could appear to adequately fit the data, even though the structural parameter estimates within that model exhibit very substantial biases that would result in erroneous inferences. This is not simply a measurement model or construct validity problem, because its effects clearly extend into the estimates of the structural parameters that drive the development and testing of marketing theory. More specifically, the results indicate that paths emanating from a construct with a misspecified measurement model are likely to be substantially inflated, thus leading to Type I errors. However, paths leading into a construct with a misspecified measurement model are likely to be deflated, thus leading to Type II errors.

TABLE 4

SUMMARY OF GOODNESS-OF-FIT STATISTICS FOR THE MODELS SHOWN IN FIGURE 3

Model	Position of misspecified construct	Correlation level	$\chi 2$	*df*	*p*	GFI	CFI	SRMR	RMSEA
1A correctly specified	Exogenous	.10	163.50	160	.44	.94	.99	.025	.007
1A misspecified	Exogenous	.40	**282.20**	166	**.00**	**.20**	.98	.042	.037
1B correctly specified	Exogenous	.70	163.77	160	.45	.94	.99	.022	.007
1B misspecified	Exogenous	.10	**212.26**	166	**.05**	**.81**	.99	.031	.023
1C correctly specified	Exogenous	.40	167.74	160	.46	.94	.99	.09	.007
1C misspecified	Exogenous	.70	177.35	166	.33	.93	.99	.022	.010
2A correctly specified	Endogenous	.10	163.56	160	.44	.94	.99	.033	.008
2A misspecified	Endogenous	.40	**289.19**	166	**.00**	**.27**	.98	.046	.038
2B correctly specified	Endogenous	.70	162.38	160	.46	.94	.99	.031	.007
2B misspecified	Endogenous	.10	**242.70**	166	**.00**	**.47**	.99	.043	.030
2C correctly specified	Endogenous	.40	164.03	160	.43	.94	.99	.031	.008
2C misspecified	Endogenous	.70	**232.17**	166	**.01**	**.80**	.99	.045	.028

NOTE.—Goodness-of-fit indices shown in bold indicate a lack of model fit.

RECOMMENDATIONS FOR SPECIFYING FORMATIVE MODELS

Although it is difficult to know precisely why measurement model misspecification has become so pervasive, it is possible to speculate about the potential reasons. First, it may result from the fact that many marketing researchers simply do not think of measurement model relationships as hypotheses to be tested with differing theoretical implications. One indication of the failure to think about measurement relationships is that most articles devote little or no attention to this issue when describing the specification of their models. Another possible explanation is that researchers are simply unaware of the conceptual distinctions between formative and reflective measurement models. Although this is somewhat surprising since the distinction was made in the field over 20 years ago (e.g., Fornell and Bookstein 1982), it is still relatively uncommon to see references to this distinction in the literature. Still another explanation may be that authors are forced into overreliance on reflective measurement model specifications by journal reviewers who demand high internal consistency between measures and unidimensionality as a condition for acceptance and publication of latent variable research.

However, from our perspective, one of the most likely reasons is that researchers do not know how to correctly specify formative constructs in covariance structure models. Procedures for developing and evaluating measures for reflective principal factor models have been discussed in the literature (Churchill 1979; Nunnally 1978), but little research has provided guidance on how to specify formative measurement constructs in latent variable structural equation models. This is important, because there are several unique problems associated with the modeling of formative indicator constructs. Therefore, the remainder of this section will provide a series of recommendations on how to develop and specify formative indicator models.

The first issue to decide when designing a study is whether the construct of interest is formative or reflective in nature. This requires a clear conceptual definition of the construct, generation of a set of measures fully representing the domain of the construct, and careful consideration of the relationships between the construct and its measures. The latter judgment could be made on the basis of the decision rules specified in table 1. Our reading of the literature suggests that researchers spend more effort theoretically justifying structural relationships than they do theoretically justifying measurement relationships, even though both should be regarded as hypotheses to be conceptually justified and tested (Bagozzi 1984). In our experience, concern with establishing the proper direction of causality is common when structural relationships are the focus of attention but rare when measurement relationships are the focus.

Achieving Identification in Formative Indicator Models

Assuming that a formative indicator measurement model is appropriate, the next issue that needs to be addressed is that of model identification. Unfortunately, a model like the one shown in panel 1 of figure 5 is not identified due to indeterminacies associated with the scale of measurement for the latent construct and the construct level error term. MacCallum and Browne (1993) note two conditions necessary for the identification of formative indicator constructs. First, the scale of measurement for the latent construct must be established either by constraining a path from one of the construct's indicators to be equal to one or by constraining the residual error variance for the construct to be equal to one. Second, to resolve the indeterminacy associated with the construct level error term, formative constructs must emit paths to (*a*) at least two unrelated latent constructs with reflective indicators (see panel 2, fig. 5), (*b*) at least two theoretically appropriate reflective indicators (see panel 3, fig. 5), or (*c*) one reflective indicator and one latent construct with reflective indicators (see panel 4, fig. 5). Although this indeterminacy could also be resolved by fixing the error term to zero (cf. Diamantopoulos and Winklhofer 2001) or by equating it with the residual variance associated with the construct it is hypothesized to influence, MacCallum and Browne (1993) argue that these procedures may not be theoretically appropriate because the former assumes that the formative measures perfectly represent the latent construct, and the latter confounds construct level measurement error with structural error.

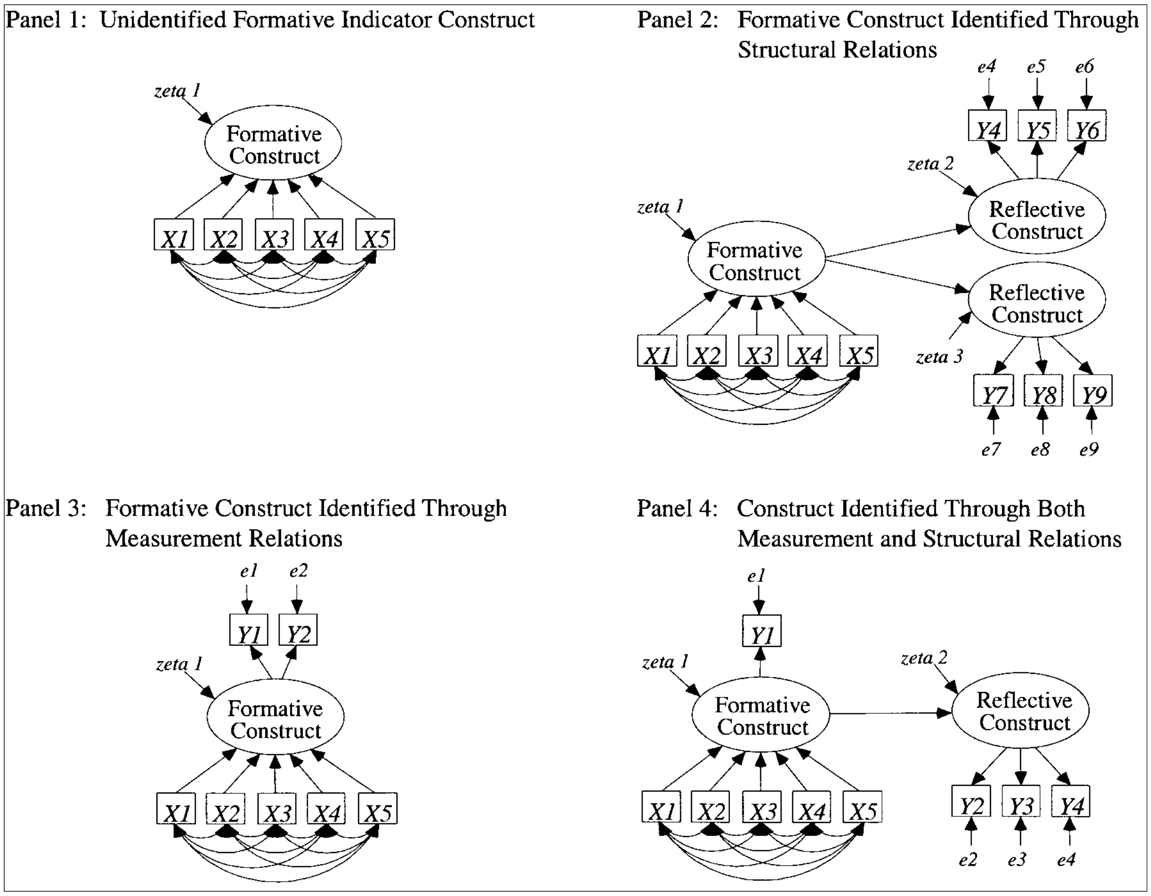
In our view, the best option for resolving the identification problem with the construct level error term is to add two reflective indicators to the formative construct, when conceptually appropriate. The advantages of doing this are that (*a*) the formative construct is identified on its own and can go anywhere in the model (e.g., as an exogenous or endogenous construct), (*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.

However, the disadvantage is that it is subject to alternative conceptual interpretations. In form, the model shown in panel 3 of figure 5 is identical to a multiple indicators of multiple causes (MIMIC) model (Jöreskog and Goldberger 1975). As such, one could interpret this model in three ways: (*a*) as a single construct with five formative and two reflective measures, (*b*) as five exogenous variables influencing a single endogenous construct with two reflective indicators, or (*c*) as a formatively measured construct that influences two manifest measures of two different constructs. Empirically, these interpretations are indistinguishable because they all produce identical estimates of the relationships between the measures and the constructs. The only difference is the conceptual interpretation attached to these relationships.

We would argue that when Y_1 and Y_2 in panel 3 of figure 5 are 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

FIGURE 5

ACHIEVING IDENTIFICATION IN FORMATIVE INDICATOR MODELS



of the key conceptual components of the construct (e.g., facets of satisfaction, risk, or trust), that it makes the most sense to interpret this structure as a single construct with five formative and two reflective indicators. Indeed, in this instance, it would not make sense to interpret this structure as five exogenous variables (e.g., facets of satisfaction, risk, or trust) influencing a separate and distinct endogenous latent construct (e.g., overall satisfaction, risk, or trust) 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. Similarly, it would not make sense to interpret it as a formatively measured construct (e.g., overall satisfaction, risk, or trust) that influences two manifest measures of different constructs because these two measures would obviously tap the exact same conceptual domain (e.g., overall satisfaction, risk, or trust). To be clear, our point is not that every model with a form like the one depicted in panel 3 of figure 5 should be interpreted as a single construct with both formative and reflective measures but, rather, that it can be if X_1–X_5 and Y_1–Y_2 are all conceptually appropriate measures of a single construct.

In order to make these issues more concrete, let us assume that a researcher is interested in studying a consumer's satisfaction with a product. Let us further assume that the consumer satisfaction construct has formative indicators, each of which captures the consumer's satisfaction with distinct attributes or aspects of the product. Thus, as shown in panel 1 of figure 5, the error term for consumer satisfaction is measured at the construct level, and the formative indicators are allowed to freely covary, per conventional practice. As specified in this model, the residual variance associated with consumer satisfaction (zeta 1) would not be identified. One way of obtaining identification would be to add two additional consequences of consumer satisfaction to the model (as shown in panel 2 of fig. 5). The likelihood of recommending the product to others and repurchase intentions might be such factors. Assuming that these factors have reflective indicators, and that they are not causally related to each other, the resulting model is identified, including zeta 1 and the new zeta 2 and zeta 3. Another way of achieving identification would be to add two reflective indicators of a consumer's satisfaction with the product (as shown in panel 3 of fig. 5). Examples of potentially appropriate items might be “Overall, how satisfied are you with this product?” and “All things considered, I am very pleased with this product.” Since these new indicators capture a consumer’s overall level of satisfaction, they are reflective in nature. Therefore, with the addition of these two reflective indicators, the consumer satisfaction construct would now have two paths emanating from it and be identified. It is important to note that this procedure results in measurement error terms (ϵ_1 and ϵ_2) for the reflective indicators (Y_1 and Y_2) plus a combined measurement error term (zeta 1) for the five formative indicators (X_1–X_5). A third option illustrated in panel 4 in figure 5 represents a compromise between these alternatives.

Of course, to implement any of these solutions, the researcher must be familiar enough with the identification problem to anticipate the need for either additional items or constructs. The construct conceptualization and nature of the indicators to be used should be determined in the questionnaire design stage, so that the need for extra indicators is anticipated early in the research process. This underscores the need for researchers to think as carefully about the measurement model relationships between constructs and their indicators as they do about the structural relationships between constructs. However, having done so, these recommendations would be easy to implement.

Modeling Exogenous Variable Intercorrelations in Formative Indicator Models

Another important issue that needs to be addressed when modeling formative indicator constructs is what to do with the covariances among the exogenous constructs and variables in the structural model. In the SEM literature, the convention is to free all covariances among the exogenous constructs on the grounds that they may be correlated due to spurious causes outside the system of relationships captured by the model. This would suggest that the same practice should be followed in structural equation models that include formative indicator constructs. However, the problem is that formative indicators are exogenous variables in a structural equation model, because they emit paths to a latent construct but do not have any paths coming into them. Consequently, following standard practice would involve adding not only covariances among the latent exogenous constructs in the model but also covariances among the formative indicators and between the formative indicators and the exogenous latent constructs. This often would add a considerable number of nonhypothesized covariances to the model.

Two ways of handling this problem have been proposed in the literature (MacCallum and Browne 1993). Neither of these solutions is ideal. The first is to constrain all of the covariances among the exogenous latent constructs and manifest variables to be equal to zero. The advantage of this method is that model parsimony is not undermined by the addition of the nonhypothesized paths and, consequently, the goodness-of-fit indices will be due predominately to the hypothesized (as opposed to nonhypothesized) relationships. However, as noted by MacCallum and Browne (1993), fixing these covariances at zero will typically result in large blocks of zeros in the predicted covariance (Σ-hat) matrix. This is a very strong theoretical statement. It assumes that the variables are perfectly uncorrelated. Thus, any common causes of these variables that are outside of the system of relationships represented in the model, any causal relationships among these variables that have been omitted from the model, or even methodological artifacts that are shared by these variables, will contribute to the lack of fit of the proposed model. For these reasons, this does not appear to be an acceptable solution.

Another method of handling the covariances is to follow standard practice and estimate the covariances among all exogenous latent constructs and manifest variables (MacCallum and Browne 1993). The advantage of this procedure is that the overall fit of the model is not unnecessarily penalized for covariances among the exogenous variables that are due to factors outside of the model. However, the problem with this approach is that it may greatly diminish model parsimony. By adding a substantial number of nonhypothesized covariances to the model, the proportion of nonhypothesized paths relative to hypothesized paths may become quite large and cause the overall goodness of fit of the model to be due more to the nonhypothesized relationships than to the hypothesized ones. This has the potential to lead to erroneous conclusions about the contribution of the hypothesized relationships to the overall fit of the model. Thus, although MacCallum and Browne (1993) recommend this approach as the best solution to the problem, it clearly becomes problematic when the number of formative indicators and exogenous latent constructs is large.

One possible way to mitigate this problem is to explicitly identify the impact of these covariances on the overall fit of the model and take them into account when judging model parsimony. This can be accomplished by estimating a series of hierarchically related (nested) models that sequentially free different sets of parameters. This would permit the effect of the exogenous covariances on the overall fit of the model to be evaluated. For example, one could estimate the following series of models: (*a*) a completely null model (i.e., no measurement or structural relationships), (*b*) a model that adds the measurement relationships only, (*c*) a model that adds covariances among the exogenous constructs and intraconstruct covariances among formative indicators, (*d*) a model that adds interconstruct covariances among formative indicators, (*e*) a model that adds covariances between formative indicators and exogenous latent variables, (*f*) a model that adds the hypothesized relationships between the constructs, and (*g*) a structurally saturated model. Once this series of models was estimated, the incremental contribution of each of these sets of parameters could be identified by comparing the goodness-of-fit indices associated with these models. In addition, indices similar to the relative normed fit index (RNFI) or relative parsimonious normed fit index

(RPNFI) could be used to evaluate the fit of the theoretical model relative to any of the baseline models that precede it in the model hierarchy. In our view, freeing up the covariances among all exogenous constructs and measures is the best approach for dealing with this issue when coupled with full disclosure of the impact of these nonhypothesized relationships on the fit of the model.

CONCLUSIONS

The goals of this research were to draw attention to the theoretical distinctions between formative and reflective measurement models, provide a set of guidelines for deciding on the appropriate measurement model, determine the extent to which measurement model misspecification has occurred in the top-tier marketing journals, examine the effects of measurement model misspecification with a Monte Carlo simulation, and recommend procedures for correctly modeling formative indicator constructs. Our discussion suggests that there are important theoretical and empirical distinctions between formative and reflective indicator measurement models, and that as many as 28% of the latent constructs with multiple indicators published in the top marketing journals were incorrectly specified as reflective when they should have been formative. This misspecification involved not only the relationships between first-order constructs and their indicators but also the relationships between second-order constructs and their first-order indicators. Indeed, our review indicates that this type of measurement model misspecification affects several of the most commonly used constructs in the field. In addition, the Monte Carlo simulation demonstrated that measurement model misspecification severely biases structural parameter estimates and can lead to inappropriate conclusions about hypothesized relationships between constructs. Therefore, measurement relationships must be appropriately modeled. Since most of the specification errors found in the literature involved the failure to properly specify constructs with formative indicators, we have provided a set of guidelines for specifying these types of models.

In closing, we believe that the failure to recognize the distinction between the measurement models discussed in this article potentially has had a number of detrimental effects on progress in the field. For example, it is likely that a large number of studies have been rejected in the review process because reviewers insisted on high internal consistency reliabilities and required a principal factor model to fit the data. As a consequence, constructs that are truly formative in nature may have received less attention in the literature and/or they may have been more likely to have been modeled as scale scores without taking measurement error into account. Perhaps an equally large number of studies have been published with severely restricted construct domains due to the same reviewer bias. This construct domain restriction undoubtedly contributes to the inconsistency in findings across studies (as slightly different subsets of the construct domain are tapped in different studies) and may partially account for the generally low proportion of variance explained in many criterion variables (Peterson, Albaum, and Beltramini 1985). In addition, as demonstrated by our simulation, the failure to correctly specify the measurement model can lead to different conclusions about the empirical relationships between latent constructs. Thus by implication, a substantial proportion of the empirical results in the literature may be potentially misleading. Therefore, it is imperative for the field as a whole to think more carefully about measurement model relationships and do a better job of making sure that the measurement models used match that conceptualization. Hopefully, this research will aid progress in this area.

[*David Glen Mick served as editor and William O. Bearden served as associate editor for this article.*]

REFERENCES

Achrol, Ravi S. and Louis W. Stern (1988), “Environmental Determinants of Decision-Making Uncertainty in Marketing Channels,” *Journal of Marketing Research*, 25 (February), 36–50.

Anderson, James C. and David W. Gerbing (1982), “Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement,” *Journal of Marketing Research*, 19 (November), 453–460.

Bagozzi, Richard P. (1980), *Causal Models in Marketing*, New York: Wiley.

——— (1981), “Evaluating Structural Equation Models with Unobservable Variables and Measurement Error: A Comment,” *Journal of Marketing Research*, 18 (August) 375–381.

——— (1984), “A Prospectus for Theory Construction in Marketing,” *Journal of Marketing*, 48 (Winter), 11–29.

Bagozzi, Richard P. and Todd F. Heatherton (1994), “A General Approach to Representing Multifaceted Personality Constructs: Application to State Self-Esteem,” *Structural Equation Modeling*, 1 (1), 35–67.

Baumgartner, Hans and Christian Homburg (1996), “Applications of Structural Equation Modeling in Marketing and Consumer Research: A Review,” *International Journal of Research in Marketing*, 13 (2), 139–161.

Bentler, Peter M. and David G. Weeks (1980), “Linear Structural Equations with Latent Variables,” *Psychometrika*, 45 (3), 289–308.

Blalock, Hubert M., Jr. (1964), *Causal Inferences in Nonexperimental Research*, New York: Norton.

Bollen, Kenneth (1989), *Structural Equations with Latent Variables*, New York: Wiley.

Bollen, Kenneth and Richard Lennox (1991), “Conventional Wisdom on Measurement: A Structural Equation Perspective,” *Psychological Bulletin*, 110 (2), 305–314.

Bone, Paula Fitzgerald and Pam Scholder Ellen (1992), “The Generation and Consequences of Communication-Evoked Imagery,” *Journal of Consumer Research*, 19 (June), 93–104.

Calantone, Roger J., Jeffrey B. Schmidt, and X. Michael Song (1996), “Controllable Factors of New Product Success: A Cross-National Comparison,” *Marketing Science*, 15 (4), 341–358.

Cannon, Joseph P. and William D. Perreault, Jr. (1999), “Buyer-Seller Relationships in Business Markets,” *Journal of Marketing Research*, 36 (November), 439–460.

Chandon, Pierre, Brian Wansink, and Gilles Laurent (2000), “A

Benefit Congruency Framework of Sales Promotion Effectiveness," *Journal of Marketing*, 64 (October), 65–84.

Churchill, G. A., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64–73.

Cote, Joseph and M. Ronald Buckley (1987), "Estimating Trait, Method, and Error Variance: Generalizing across 70 Construct Validation Studies," *Journal of Marketing Research*, 24 (August), 315–318.

Cravens, David W., Thomas N. Ingram, Raymond W. LaForge, and Clifford E. Young (1993), "Behavior-Based and Outcome-Based Salesforce Control Systems," *Journal of Marketing*, 57 (October), 47–59.

Crosby, Lawrence A., Kenneth R. Evans, and Deborah Cowles (1990), "Relationship Quality in Services Selling: An Interpersonal Influence Perspective," *Journal of Marketing*, 54 (July), 68–81.

Crosby, Lawrence A. and Nancy Stephens (1987), "Effects of Relationship Marketing on Satisfaction, Retention and Prices in the Life Insurance Industry," *Journal of Marketing Research*, 24 (November), 404–411.

Diamantopoulos, Adamantios and Heidi M. Winklhofer (2001), "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research*, 38 (May), 269–277.

Edwards, Jeffery R. and Richard P. Bagozzi (2000), "On the Nature and Direction of Relationships between Constructs and Measures," *Psychological Methods*, 5 (2), 155–174.

Fishbein, Martin and Icek Ajzen (1975), *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*, Reading, MA: Addison-Wesley.

Fornell, Claes and Fred L. Bookstein (1982), "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research*, 19 (November), 440–452.

Fornell, Claes and David F. Larcker (1981), "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, 18 (February), 39–50.

Gaski, John F. (1986), "Interrelations among a Channel Entity's Power Sources: Impact of the Exercise of Reward and Coercion on Expert, Referent, and Legitimate Power Sources," *Journal of Marketing Research*, 23 (February), 62–77.

Gerbing, David W. and James C. Anderson (1984), "On the Meaning of Within-Factor Correlated Measurement Errors," *Journal of Consumer Research*, 11 (June), 572–580.

——— (1988), "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research*, 25 (2), 186–192.

Hallen, Lars, Jan Johanson, and Nazeem Seyed-Mohammed (1991), "Interfirm Adaptation in Business Relationships," *Journal of Marketing*, 55 (April), 29–37.

Hartline, Michael D. and O. C. Ferrell (1996), "The Management of Customer-Contact Service Employees: An Empirical Investigation," *Journal of Marketing*, 60 (October), 52–70.

Hu, Li-tze and Peter M. Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives," *Structural Equation Modeling*, 6 (1), 1–55.

John, George (1984), "An Empirical Investigation of Some Antecedents of Opportunism in a Marketing Channel, *Journal of Marketing Research*, 21 (August), 278–289.

Jöreskog, Karl and A. S. Goldberger (1975), "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 70 (September), 631–639.

Kohli, Ajay K., Bernard J. Jaworski, and Ajith Kumar (1993), "MARKOR: A Measure of Market Orientation," *Journal of Marketing Research*, 30 (November), 467–477.

Law, Kenneth, and Chi-Sum Wong (1999), "Multidimensional Constructs in Structural Equation Analysis: An Illustration Using the Job Perception and Job Satisfaction Constructs," *Journal of Management*, 25 (2), 143–160.

Lichtenstein, Donald R., Nancy M. Ridgway, and Richard G. Netemeyer (1993), "Price Perceptions and Consumer Shopping Behavior: A Field Study," *Journal of Marketing Research*, 30 (May), 234–245.

MacCallum, Robert C. and Michael W. Browne (1993), "The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues," *Psychological Bulletin*, 114 (3), 533–541.

MacCorquodale, K. and Paul E. Meehl (1948), "On a Distinction between Hypothetical Constructs and Intervening Variables," *Psychological Review*, 55(2), 307–321.

MacKenzie, Scott B., Richard J. Lutz, and George E. Belch (1986), "The Role of Attitude toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations," *Journal of Marketing Research*, 23 (May), 130–143.

Michaels, Ronald E., Ralph L. Day, and Erich A. Joachimsthaler (1987), "Role Stress among Industrial Buyers: An Integrative Model," *Journal of Marketing*, 51 (April), 28–45.

Mohr, Jakki J., Robert J. Fisher, and John R. Nevin (1996), "Collaborative Communication in Interfirm Relationships: Moderating Effects of Integration and Control," *Journal of Marketing*, 60 (July), 103–115.

Moorman, Christine and Erika Matulich (1993), "A Model of Consumers' Preventative Health Behaviors: The Role of Health Motivation and Health Ability," *Journal of Consumer Research*, 20 (September), 208–228.

Murry, John P., Jr. and Peter A. Dacin (1996), "Cognitive Moderators of Negative-Emotion Effects: Implications for Understanding Media Context," *Journal of Consumer Research*, 22 (March), 439–447.

Nunnally, Jum C. (1978), *Psychometric Theory*, New York: McGraw Hill.

Paxton, Pamela, Patrick J. Curran, Kenneth A. Bollen, Jim Kirby, and Feinian Chen (2001), "Monte Carlo Experiments: Design and Implementation," *Structural Equation Modeling*, 8 (2), 287–312.

Peter, J. Paul (1981), "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research*, 18 (May), 133–145.

Peterson, Robert A., Gerlad Albaum, and Richard F. Beltramini (1985), " A Meta Analysis of Effect Sizes in Consumer Behavior Experiments," *Journal of Consumer Research*, 12 (June), 97–103.

Podsakoff, Philip M. and Scott B. MacKenzie (1994), "Organizational Citizenship Behaviors and Sales Unit Effectiveness," *Journal of Marketing Research*, 31 (August), 351–363.

Reilly, Michael D. (1982), "Working Wives and Convenience Consumption," *Journal of Consumer Research*, 8 (March), 407–418.

Rose, Gregory M. (1999), "Consumer Socialization, Parental Style, and Developmental Timetables in the United States and Japan," *Journal of Marketing*, 63 (July), 105–119.

Ryan, Michael J. (1982), "Behavioral Intention Formation: The Interdependency of Attitudinal and Social Influence Variables," *Journal of Consumer Research*, 9 (December), 263–278.

Shimp, Terence A. and Alician Kavas (1984), "The Theory of Reasoned Action Applied to Coupon Usage," *Journal of Consumer Research*, 11 (December), 795–809.

Siguaw, Judy A., Penny M. Simpson, and Thomas L. Baker (1998), "Effects of Supplier Market Orientation on Distributor Market Orientation and the Channel Relationship: The Distributor Perspective," *Journal of Marketing*, 62 (July), 99–111.

Singh, Jagdip (1988), "Consumer Complaint Intentions and Behavior: Definitional and Taxonomical Issues," *Journal of Marketing*, 52 (January), 93–107.

Srinivasan, Narasimhan and Brian T. Ratchford (1991), "An Empirical Test of a Model of External Search for Automobiles," *Journal of Consumer Research*, 18 (September), 233–242.