

Although the methodological literature is replete with advice regarding the development and validation of multi-item scales based on reflective measures, the issue of index construction using formative measures has received little attention. The authors seek to address this gap by (1) examining the nature of formative indicators, (2) discussing ways in which the quality of formative measures can be assessed, and (3) illustrating the proposed procedures with empirical data. The aim is to enhance researchers' understanding of formative measures and assist them in their index construction efforts.

Index Construction with Formative Indicators: An Alternative to Scale Development

Following the publication of the *JMR* "Special Section on Measurement and Marketing Research" in February 1979—which included Churchill's (1979) and Peter's (1979) seminal articles—a great deal of attention has been devoted to developing marketing measures with sound psychometric properties (e.g., Churchill and Peter 1984; Peter 1981). The advent of structural equation modeling techniques further facilitated the assessment of multi-item measures in terms of dimensionality, reliability, and validity (e.g., Anderson and Gerbing 1982; Danes and Mann 1984; Steenkamp and van Trijp 1991). As a result of these developments, the usage of multi-item measures has increased substantially in the past 20 years (Bruner and Hensel 1993).

Although the methodological soundness of marketing studies has undoubtedly been improved as a result of these advances, existing measure development guidelines (e.g., DeVellis 1991; Spector 1992) focus almost exclusively on scale development, whereby items (i.e., observed variables) composing a scale are perceived as reflective (effect) indicators of an underlying construct (i.e., latent variable). This perspective reflects the conventional wisdom on measurement (see Bollen and Lennox 1991), which is largely based on classical test theory and, in particular, the domain-sampling model (Nunnally and Bernstein 1994).

An alternative measurement perspective is based on the use of formative (cause, causal) indicators and involves the creation of an index rather than a scale (Bollen and Lennox

1991).¹ In this context, formative indicators "are observed variables that are assumed to cause a latent variable. For effect indicators the latent variable causes the observed variables. Most researchers in the social sciences assume that indicators are effect indicators. Cause indicators are neglected despite their appropriateness in many instances" (Bollen 1989, p. 65). Such neglect is certainly the case in marketing, because practically all measures available are based—implicitly or explicitly—on reflective indicators (see, e.g., the compilations by Bearden, Netemeyer, and Mobley [1993] and Bruner and Hensel [1992, 1996]).

The purpose of this article is to provide some insights into the nature of formative indicators so that researchers can reach an informed choice as to the appropriate measurement model for their needs. In particular, we seek to complement existing guidelines on scale development with some suggestions regarding index construction. We first provide a brief background on formative indicators, drawing from the limited (and fragmented) literature on the topic. Next, we give an example of constructing an index based on actual survey data and highlight the procedures used to assess its quality. We conclude the article with some thoughts about the use of indexes in marketing studies.

FORMATIVE INDICATORS: THEORETICAL BACKGROUND

A measurement perspective based on formative indicators reflects the notion that "in many cases, indicators could be viewed as causing rather than being caused by the latent variable measured by the indicators" (MacCallum and Browne 1993, p. 533). A typical example is socioeconomic

*Adamantios Diamantopoulos is Chair of Marketing and Business Research, Loughborough University Business School (e-mail: a.diamantopoulos@lboro.ac.uk). Heidi M. Winklhofer is Lecturer in Marketing, Business School, The University of Nottingham (e-mail: heidi.winklhofer@nottingham.ac.uk). The authors thank the three anonymous *JMR* reviewers for helpful suggestions on a previous version of this article.

¹Note that the terminology can often be confusing in the literature; for example, Babbie (1973) uses the term "index" to describe what is really a scale, and several scales in Bruner and Hensel's (1996) compilation have "index" in their titles.

status (SES), which is formed as a combination of education, income, occupation, and residence (see Hauser 1971, 1973). If any one of these measures increases, SES would increase (even if the other indicators did not change); conversely, if a person's SES increases, this would not necessarily be accompanied by an increase in all four measures. The choice of a formative versus a reflective specification thus depends on the causal priority between the indicator and the latent variable (Bollen 1989).² More specifically,

constructs such as "personality" or "attitude" are typically viewed as underlying factors that *give rise to* something that is observed. Their indicators tend to be realized, then as reflective. On the other hand, when constructs are conceived as explanatory *combinations* of indicators (such as "population change" or "marketing mix") that are determined by a combination of variables, their indicators should be formative. (Fornell and Bookstein 1982, p. 292, emphasis in the original)³

A formative perspective (though not explicitly acknowledged as such) characterizes several composite measures used in the economics literature. Examples include the index of sustainable economic welfare (Daly and Cobb 1989), the human development index (United Nations Development Program 1990), the quality-of-life index (Johnston 1988), various country risk and market potential indexes (e.g., Dichtl and Köglmayr 1986; Economist Intelligence Unit 1996), and quality-adjusted (i.e., hedonic) price indexes (e.g., Griliches 1971).⁴ In contrast, use of formative measures is rather rare in marketing studies; only a handful of authors intentionally choose a formative specification for some of their measures (e.g., see the index of customer expectations by Ennew, Reed, and Binks [1993] or the market complexity measure by Homburg, Workman, and Krohmer [1999]).⁵

The origins of the formative perspective can be traced back to the "operational definition" model. Under strict operationalism, "a concept becomes its measure and has no meaning beyond that measure.... [T]he entire meaning of a theoretical concept is assigned to its measurement and any theoretical concept has one and only one measurement" (Bagozzi 1982, p. 15). Thus, if η represents the concept (i.e., latent variable) in question and x is an empirical measure (i.e., observed or manifest variable), then

$$(1) \quad \eta \equiv x.$$

A more contemporary view, which allows the possibility of multiple measures, x_i , ($i = 1, 2, \dots, n$), suggests that "a concept is assumed to be defined by, or to be a function of, its measurements" (Bagozzi and Fornell 1982, p. 34).

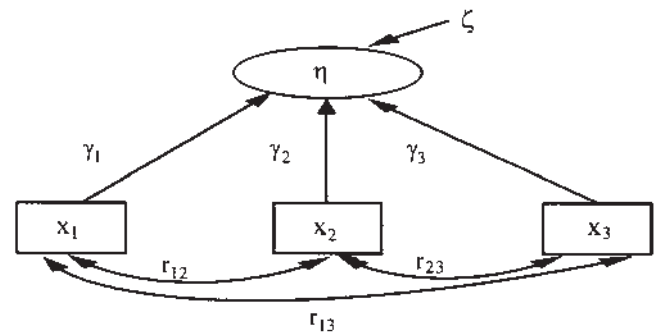
²As Nunnally and Bernstein (1994, p. 449) observe, "people have high socio-economic status because they are wealthy and/or well-educated; they do not become wealthy or educated because they are of high socio-economic status."

³Formative measurement, though potentially applicable to the measurement of individual characteristics (such as SES), is particularly relevant for dealing with organizational and social constructs—that is, when the unit of analysis is the firm or group (Bagozzi 1994).

⁴For further examples, we refer the reader to Horn (1993) and the references given therein.

⁵However, as is discussed subsequently (see Table 1), several marketing measures supposedly operationalized by means of reflective indicators are probably better viewed as combinations of formative indicators.

Figure 1
A FORMATIVE MEASUREMENT MODEL



According to this latter definition, a formative specification implies the following relationship:

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

where γ_i is a parameter reflecting the contribution of x_i to the latent variable η .

Another formative specification (shown in Figure 1 for $n = 3$) is provided by Bollen and Lennox (1991, p. 306):

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

Equation 1 characterizes much of the research in marketing during the 1960s and 1970s in which the use of single-item measures was common (Jacoby 1978). However, its utility as a measurement model is limited because it "rejects the possibility that multiple measurements may be taken for a theoretical concept" (Bagozzi 1982, p. 15); therefore, it makes the statistical assessment of measurement error problematic.

Regarding Equations 2 and 3, a key difference between them is that the former is consistent with a principal components model (Bagozzi and Fornell 1982; Nunnally and Bernstein 1994), whereas the latter is not (because of the presence of the disturbance term, ζ). Note also that unless all γ s are equal to 1 and, in the case of Equation 3, ζ is also equal to 0, a linear composite based on the sum of equally weighted indicators will not be equivalent to the latent variable, η (Bollen and Lennox 1991). This is an important point not least because discussions of formative indicators in the literature sometimes refer to both formulations and thus can lead to confusion. For example, according to Bagozzi (1994, p. 332, emphasis added), "when a latent variable is *defined* as a linear sum of a set of measurements *or* when a set of measures of a dependent variable is *determined* by a linear combination of measures of independent variables, the measures are termed formative indicators: the measures produce the constructs so to speak."

Formative indicators have several properties that sharply distinguish them from reflective indicators.⁶ First and fore-

⁶Formative indicators "are not indicators in the conventional sense as defined in factor analysis or covariance structure modelling. Rather they are exogenous measured variables that influence the composite defined as a causally indicated variable" (MacCallum and Browne 1993, p. 533). For a formal comparison between formative and reflective indicators, see Fornell, Rhee, and Yi (1991) or Fornell and Cha (1994).

most, whereas reflective indicators are essentially interchangeable (and therefore the removal of an item does not change the essential nature of the underlying construct),⁷ with formative indicators, “omitting an indicator is omitting a part of the construct” (Bollen and Lennox 1991, p. 308). Second, as can be seen in Figure 1, the correlations among formative indicators are not explained by the measurement model; the indicators being exogenously determined, it becomes more problematic to assess their validity (Bollen 1989). Third, and related to the previous point, there is no reason that a specific pattern of signs (i.e., positive versus negative) or magnitude (i.e., high versus moderate versus low) should characterize the correlations among formative indicators (Bollen 1984); indeed, “internal consistency is of minimal importance because two variables that might even be negatively related can both serve as meaningful indicators of a construct” (Nunnally and Bernstein 1994, p. 489). Fourth, unlike their reflective counterparts, formative indicators do not have error terms; error variance is represented only in the disturbance term, ζ , which is uncorrelated with the x s (i.e., $\text{cov}[x_i, \zeta] = 0$). It is worth noting in this context that with formative indicators, the variance in true scores is higher than the variance in observed scores; in the reflective specification, the opposite is true (see Fornell, Rhee, and Yi 1991).⁸ Fifth, taken in isolation, the formative indicator measurement model in Figure 1 is statistically underidentified (see Bollen and Lennox 1991); the model can be estimated only if it is placed within a larger model that incorporates consequences (i.e., effects) of the latent variable in question (Bollen 1989). In contrast, a model with three effect indicators is identified, and confirmatory factor analysis can be used to estimate its parameters (see Long 1983). Sixth, even if the model is embedded within a larger model, identification of all parameters may still prove problematic when a formative specification is involved. For example, MacCallum and Browne (1993) indicate that a necessary—but not sufficient—condition for identifying the residual variance (i.e., the disturbance term) is that the latent variable emits at least two paths to other latent variables measured with effect indicators.

As a result of these characteristics, conventional procedures used to assess the validity and reliability of scales composed of reflective indicators (e.g., factor analysis and assessment of internal consistency) are not appropriate for composite variables (i.e., indexes) with formative indicators. As Bollen (1989, p. 222) points out, “unfortunately, traditional validity assessments and classical test theory do not cover cause indicators.” Similarly, Bagozzi (1994, p. 333) warns that “reliability in the internal consistency sense and construct validity in terms of convergent and discriminant validity are not meaningful when indexes are formed as a linear sum of measurements.” These comments imply that alternative approaches must be followed to evaluate the

quality of measures that are based on formative indicators. This is the subject of the next section, in which several guidelines for index construction are presented and illustrated by means of an example.

INDEX CONSTRUCTION

Unlike scale development, for which detailed step-by-step guides exist for construct specification, item selection and purification, and scale validation (e.g., Churchill 1979; DeVellis 1991; Spector 1992), guidelines for constructing indexes based on formative indicators are much harder to find. However, inspection of the dispersed literature on formative indicators reveals that four issues are critical to successful index construction: content specification, indicator specification, indicator collinearity, and external validity. We illustrate each of these issues with reference to empirical data drawn from a survey on export sales forecasting practices. The focal construct for which we seek to generate an index is the resource commitment of a firm to the export sales forecasting task; the statistical analysis is based on 173 mail questionnaires completed by senior managers in exporting firms (for data collection details, see Appendix A).

Content Specification

The first issue in index construction is the specification of the scope of the latent variable, that is, the domain of content the index is intended to capture. This is particularly important because “an index is more abstract and ambiguous than a latent variable measured with reflective indicators” (Bagozzi 1994, p. 333). Because under formative measurement the latent variable is determined by its indicators rather than vice versa, content specification is inextricably linked with indicator specification (discussed subsequently). Consequently, “breadth of definition is extremely important to causal indicators” (Nunnally and Bernstein 1994, p. 484), not least because failure to consider all facets of the construct will lead to an exclusion of relevant indicators (and thus exclude part of the construct itself; see the previous section).

In our example, we specify the domain of content of the focal construct (resource commitment) as the monetary, informational, managerial, and time-related resources specifically devoted to the activity of forecasting sales in export markets. We thus seek to capture the range of resources that bear on the export sales forecasting task in fairly broad terms.

Indicator Specification

Whereas for reflective indicators, according to the domain sampling model, a set of items “is chosen randomly from the universe of items relating to the construct of interest” (DeVellis 1991, p. 55), a census of indicators is required for a formative specification (Bollen and Lennox 1991). More specifically, the items used as indicators must cover the entire scope of the latent variable as described under the content specification. In our example, this implies that the indicators selected must capture monetary, informational, managerial, and time aspects; failing to include at least one indicator for, say, informational resources would change the composition of the latent variable. This does not mean that no index purification is possible by the exclusion of indicators (see the next section). It simply stresses the need at the

⁷It may, however, adversely affect construct reliability and/or lead to identification problems if the number of remaining indicators becomes too low (i.e., less than three).

⁸Following Fornell, Rhee, and Yi (1991), in the formative specification, $T = O + \epsilon$, where T = true score, O = observed score, and ϵ = disturbance term. However, ϵ captures all the remaining causes of T other than O , and therefore $\text{Cov}(T, \epsilon) \neq 0$ and $\text{Cov}(O, \epsilon) = 0$. Thus $\text{Var}(T) > \text{Var}(O)$. For further details, see Fornell, Rhee, and Yi (1991) and Nambodiri, Carter, and Blalock (1975).

indicator specification stage to be sufficiently inclusive in order to capture fully the construct's domain of content.

Appendix B lists the items generated to be used as formative indicators for the resource commitment index; these were based on an extensive review of the forecasting literature as well as exploratory interviews with export managers. In writing up the items, conventional guidelines regarding clarity, length, directionality, lack of ambiguity, and avoidance of jargon were followed (e.g., DeVellis 1991; Spector 1992), and a five-point Likert format was used for scoring.

Indicator Collinearity

An issue particular to formative indicators is that of multicollinearity. This is because the formative measurement model is based on a multiple regression (see Equation 3), and therefore the stability of the indicator coefficients (γ s) is affected by the sample size and strength of the indicator intercorrelations. Excessive collinearity among indicators thus makes it difficult to separate the distinct influence of the individual x s on the latent variable, η . Two issues are relevant in this context. First, given that each indicator coefficient, γ_i , shows the direct structural relation between x_i and η , the magnitudes of the γ s can be interpreted as validity coefficients (Bollen 1989); thus, high multicollinearity would render the assessment of indicator validity problematic. Second, if a particular x_i turns out to be almost a perfect linear combination of the other x s, it is likely to contain redundant information and can therefore become a candidate for exclusion from the index (Bollen and Lennox 1991). Note that under reflective measurement, multicollinearity is not an issue because only simple regressions are involved (in which the indicator serves as the criterion and the latent variable as the predictor).

In our example, multicollinearity among the ten indicators listed in Appendix A did not seem to pose a problem; the maximum variance inflation factor came to 1.588, which is far below the common cut-off threshold of 10 (e.g., Kleinbaum, Kupper, and Muller 1988). Therefore, all ten items were retained for initial inclusion in the index.

External Validity

As already mentioned, the very nature of formative measurement renders an internal consistency perspective inappropriate for assessing the suitability of indicators; indeed, "the best we can do ... is to examine how well the index relates to measures of other variables" (Bagozzi 1994, p. 333). However, the literature is unclear as to exactly how this should be done and practically silent on the circumstances, if any, calling for the removal of invalid indicators from the index. From a theoretical perspective, elimination of indicators carries the risk of changing the construct itself (see the previous discussion) and should always be approached with caution. In contrast, from a practical perspective, an excessive number of indicators is undesirable because of both the data collection demands it imposes and the increase in the number of parameters when the construct is embedded within a broader structural model (e.g., in a LISREL context). How to balance these considerations is a question that has not yet been fully resolved; what follows, therefore, should be viewed as suggestive rather than conclusive and as subject to further conceptual and methodological development.

At a basic level, to obtain an initial idea of the quality of individual indicators, each indicator could be correlated to another variable (external to the index), and only those indicators that are significantly correlated with the variable of interest would be retained (see Spector 1992). There must be solid theoretical reasons why the indicators in the proposed index should relate to this variable. One possibility is to use as an external criterion a global item that summarizes the essence of the construct that the index purports to measure. For example, in our study, we developed an additional statement reading "our firm devotes few resources (people, money, time) to export sales forecasting" and used this as an initial screen for the indicators in Appendix B. All but one of the indicators (x_{10}) turned out to be significantly correlated with this statement (at $p < .05$ or better); subsequently, x_{10} was excluded, and further index validation was based on the remaining nine items (x_1 – x_9).

A more satisfactory approach to validation, allowing the assessment of the proposed indicators as a set (i.e., taking account of their interrelationships), is to include some reflective indicators and estimate a multiple indicators and multiple causes (MIMIC) model (Hauser and Goldberger 1971; Jöreskog and Goldberger 1975). In this model, the index indicators, x_i , act as direct causes of the latent variable, η , which is indicated by one or more reflective measures, y_j ($j = 1, 2, \dots, m$); the inclusion of reflective measures is necessary for identification purposes (Bollen 1989). Figure 2 illustrates a MIMIC model for $n = 3$ and $m = 2$. If the overall model fit proves acceptable, this can be taken as supporting evidence for the set of indicators forming the index. In addition, by focusing on the γ parameters, researchers can assess the contribution and significance of the individual indicators.

In our example, we estimated a MIMIC model with x_1 – x_9 as direct causes of resource commitment and two further items (y_1 and y_2) as reflective indicators of the construct (the

Figure 2
MIMIC MODEL

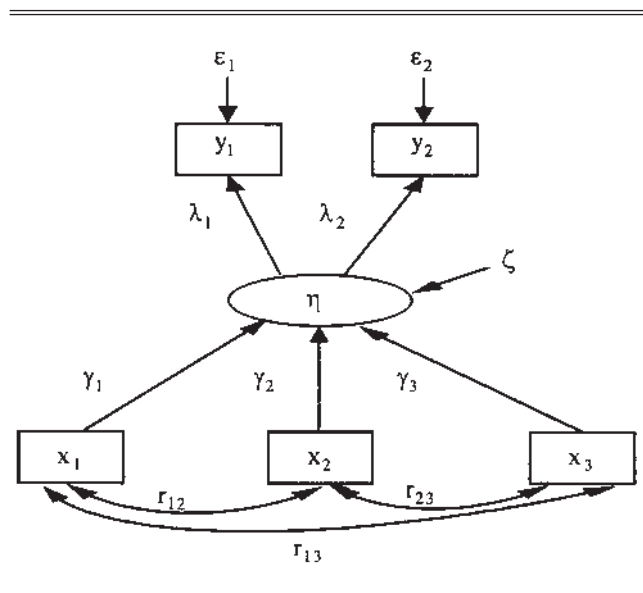
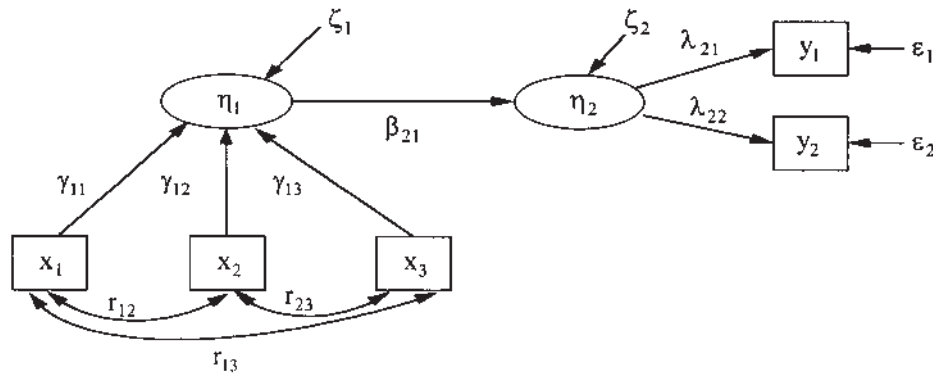


Figure 3
TWO-CONSTRUCT MODEL WITH FORMATIVE AND REFLECTIVE INDICATORS



latter are also shown in Appendix A).⁹ The two reflective items were chosen to represent effects of resource commitment; specifically, they captured the difficulty and costs associated with developing export sales forecasts, both of which are directly influenced by the various resource input to the export sales forecasting task.

Estimation of the model produced a good fit ($\chi^2 = 8.34$, d.f. = 8, $p = .401$; root mean square error of approximation [RMSEA] = .016, root mean square residual [RMSR] = .026, goodness-of-fit index [GFI] = .991, nonnormed fit index [NNFI] = .987, comparative fit index [CFI] = .998). However, several of the γ s turned out to be nonsignificant, which suggested that perhaps not all nine indicators should be included in the index. Recall that the γ s can be interpreted as validity coefficients (Bollen 1989); if the values of these are zero in the population, then arguably the indicators involved cannot be considered valid measures of the construct (and a nonsignificant t -statistic for γ fails to reject the zero value hypothesis). Bearing this in mind, we reestimated the model after eliminating nonsignificant indicators; the latter were excluded one at a time in an iterative process, starting with one displaying the lowest t -value. The revised model, comprising only four indicators (x_1 , x_3 , x_5 , and x_7), also produced an acceptable fit ($\chi^2 = 3.45$, d.f. = 3, $p = .328$; RMSEA = .029, RMSR = .024, GFI = .993, NNFI = .976, CFI = .995). A comparison of the nine-indicator model with the four-indicator model showed no significant deterioration in fit ($\Delta\chi^2 = 4.89$, d.f. = 5, $p > .10$), which is also reflected in the minimal reduction in variance explained (R^2) in the latent variable (the latter dropped from .453 to .414). Last, but by no means least, inspection of the four indicators still revealed sufficient breadth of coverage for capturing the content of the construct. This property is of major conceptual importance: Had indicator elimination resulted in an index comprising items that captured, say, only managerial resources, the nature of the construct would have been altered (Bollen and Lennox 1991; Nunnally and Bernstein

1994). Indicator elimination—by whatever means—should not be divorced from conceptual considerations when a formative measurement model is involved.

A final approach to validation, focusing on nomological aspects, involves linking the index to other constructs with which it would be expected to be linked (i.e., antecedents and/or consequences).¹⁰ Such validation is particularly relevant when indicators have been eliminated from the original index; under these circumstances, it becomes essential to establish that the new version functions in predictable ways. Validation along these lines requires (1) that information is gathered for at least one more construct than the one captured by the index, (2) that this other construct is measured by means of reflective indicators, and (3) that a theoretical relationship can be postulated to exist between the constructs. Figure 3 shows this approach with the latent variable captured by the index (η_1) acting as predictor of another latent variable (η_2); of particular interest is the sign and magnitude of the β_{21} parameter, as this represents the theoretical relationship between η_1 and η_2 .

In our example, we estimated a model similar to the one in Figure 3 in which η_1 represented resource commitment (comprising the four indicators selected by the MIMIC analysis) and η_2 represented managers' assessment of overall export forecast performance; the latter was measured by two reflective indicators (y_3 and y_4) (see Appendix A). Because greater resource commitment should improve forecast performance (e.g., see Dalrymple 1987), we expected that $\beta_{21} > 0$. Following MacCallum and Browne (1993), we scaled η_1 and η_2 by fixing one of their respective indicator paths to 1.0 and set the error variance of η_1 to 0.¹¹ Estimation of the model resulted in a good overall fit ($\chi^2 = 4.591$, d.f. = 3, $p = .204$; RMSEA = .056, RMSR = .023, GFI = .991, NNFI = .917, CFI = .983); the path between η_1 and η_2 was also found to be significant and consistent with expectations ($\beta_{21} = .108$, $p < .05$). Thus, evidence in support

⁹Following the suggestions of an anonymous reviewer, we repeated the MIMIC analysis by also including x_{10} (i.e., the item excluded at the previous stage); the results were practically identical to those reported in the text and confirmed the decision to eliminate x_{10} . We thank the reviewer for suggesting this analysis.

¹⁰After all, the substantive reason behind index construction is likely to be "how the index functions as a predictor or predicted variable" (Bagozzi 1994, p. 332).

¹¹The latter constraint is necessary when a latent variable measured by formative indicators emits only a single path (as is the case here).

of the validity of the resource commitment index was obtained.¹²

It goes without saying that when an index has been constructed and subjected to the validity tests discussed previously, it should be cross-validated on a fresh set of data (see Cudeck and Browne 1983). That formative rather than reflective indicators are involved does not in any way reduce the need to replicate the results on different samples; capitalization on chance constitutes a clear and present danger regardless of the type of indicator used.

CONCLUSION

Almost 30 years ago, in his influential article on unobserved variables, Blalock (1971, p. 346) stated that "the causal connections between unmeasured variables and their indicators ... should be made explicit so that implications for tests and estimating procedures can be noted." Much more recently, Bollen and Lennox (1991, p. 312) warned that "researchers should not automatically confine themselves to the unidimensional classical test model." In this article, we focus on the formative indicator model and attempt to provide some guidelines for constructing formative multi-item measures (indexes). The procedures presented here should prevent the use of indexes as "a handy excuse for low internal consistency" (Bollen and Lennox 1991, p. 312); although conventional scale development guidelines may be inappropriate, there are systematic steps that can be followed to help construct a valid measure from formative indicators.

The choice between a formative and a reflective specification should primarily be based on theoretical considerations regarding the causal priority between the indicators and the latent variable involved (Cohen et al. 1990; Fornell, Rhee, and Yi 1991; Namboodiri, Carter, and Blalock 1975). Additional considerations include the objectives of the study and empirical issues (see Fornell and Bookstein 1982). With regard to the former, "should the study intend to account for observed variances, reflective indicators ... are most suitable. If the objective is explanation of abstract or 'unobserved' variance, formative indicators ... would give greater explanatory power" (Fornell and Bookstein 1982, p. 292).¹³ With regard to empirical considerations, as already noted, indicator collinearity and sample size effects will affect the stability of the indicator coefficients in the formative model. If multicollinearity is a serious problem and the study's main concern is the explanation of structural model variance, a formative speculation may still be used, but interpretation should be based on loadings (as in canonical correlation analysis) rather than regression weights (γ s); procedures for doing this are described by Bagozzi, Fornell, and Larcker (1981).¹⁴

¹²Although the proportion of variance explained in η_2 by the resource index (η_1) is somewhat low ($R^2 = .101$), it should be appreciated that forecast performance is affected by several other factors, such as the turbulence of the environment (Mahmoud 1987) and the specific forecasting methods used (Small 1980).

¹³Note that using a combination of formative and reflective indicators is also an option. For example, "if one intends to explain variance in the observed criterion variables by way of the unobservables, the indicators of the endogenous construct should be reflective, and those of the exogenous formative" (Fornell and Bookstein 1982, p. 292).

¹⁴The authors thank an anonymous reviewer for alerting them to the potential use of loadings in the presence of multicollinearity among formative indicators.

In conclusion, we offer three thoughts to stimulate further debate and research: First, we believe that several marketing constructs currently operationalized by means of reflective indicators would be better captured if approached from a formative perspective. Table 1 shows a selection of multi-item measures that have been based on conventional scale development procedures (thus assuming reflective measurement) but that appear to be much more consistent with a formative perspective; inspection of the items constituting these scales reveals that the causal priority runs from the indicators to the construct, not the other way around. For example, Gaski and Nevin's (1985) coercive power measure is formed as a combination of diverse punitive actions that a supplier could take (e.g., delay delivery, refuse to sell); an increase (decrease) in the supplier's capability to take any one of these actions would result in an increase (decrease) in coercive power. Similarly, in Moorman and Matulich's (1993) health information utilization measure, there is no compelling rationale to expect that an increase (decrease) in information utilization would be reflected in the use of all five information sources involved. Thus, in both cases, the assumption of a reflective measurement model is questionable (the same applies to the rest of the measures in Table 1).

Second, we suspect that the poor fit experienced in covariance structure models may sometimes indicate that the wrong type of measurement model has been applied (as opposed to simply misspecifying the relations between the indicators and their underlying factors and/or relations between constructs). In this context, respecification of indicators from reflective to formative can often result in an improved model fit (for an illustration, see MacCallum and Browne 1993).

Third, we wonder whether the sheer availability of software for covariance structure analysis (e.g., LISREL, EQS, AMOS) has resulted in an almost automatic acceptance of reflective indicators in the minds of researchers. Although formative indicators can be accommodated within a conventional covariance structure framework (as we show in this article), it should not be forgotten that partial least squares (PLS) offers an alternative for incorporating formative (as well as reflective) measurement models. Given that PLS methodology has several attractive features (see Fornell and Bookstein 1982; Fornell and Cha 1994), the use of PLS for index construction purposes seems to be a particularly interesting area for further research.

APPENDIX A DATA COLLECTION

Survey data were collected by a mail questionnaire that was developed on the basis of a comprehensive review of the exporting and forecasting literature as well as exploratory interviews with exporting companies. The instrument was subsequently pretested by protocol interviews and two separate mail pilots with 100 exporters each. For the main survey, a size-stratified sample of 1330 exporters in the manufacturing field was targeted, which was derived from the *Dun & Bradstreet* database of exporters. A personalized mailing was undertaken to the employees responsible for export operations, a postage-paid return envelope was included, and the respondents were promised a copy of the results. Two weeks after the initial mailing, a follow-up letter was sent to a stratified sample of 300 identified nonre-

Table 1
EXAMPLES OF MULTI-ITEM MEASURES SUGGESTING A FORMATIVE PERSPECTIVE

<i>Author</i>	<i>Construct</i>	<i>Description (Scoring)</i>	<i>Items</i>
Gaski and Nevin (1985)	Perceived coercive power (in a marketing channel)	Supplier capability to take different kinds of action (0 = "no capability," 4 = "very much capability")	1. Delay delivery 2. Delay warranty claims 3. Take legal action 4. Refuse to sell 5. Charge high prices 6. Deliver unwanted products
McKee, Varadarajan, and Pride (1989)	Advertising expenditures (bank)	Spending on advertising compared with primary competitor (4 = "much more," 3 = "more," 2 = "same," 1 = "less")	1. Television 2. Radio 3. Newspaper 4. All media in total
Lumpkin and Hunt (1989)	Convenience (shopping)	Importance of different aspects (1 = "not important," 2 = "below average importance," 3 = "average importance," 4 = "above average importance," 5 = "very important")	1. Delivery to home 2. Telephone in order 3. Transportation to store 4. Convenient parking 5. Location close to home 6. Variety of stores close together
Burke (1984)	Company resource sharing	Extent of sharing of resources among business units (1 = "not all," 7 = "great")	1. Plant and equipment 2. Production personnel 3. Sales force 4. Distribution channels 5. Management services 6. Research and development facilities 7. Research and development personnel
Moorman and Matulich (1993)	Health information sources (nonpersonal)	Frequency of use of different information sources (1 = "none of the time," 7 = "all of the time")	1. Advertisements 2. Books, magazines, or pamphlets about health 3. Newspapers 4. Television and radio programming 5. Product labels
Richins and Dawson (1992)	Ecological awareness	Frequency of different behaviors (1 = "nearly never," 2 = "occasionally," 3 = "frequently," 4 = "usually," 5 = "almost always")	1. Recycle newspapers used at home 2. Recycle glass jars and bottles used at home 3. Intentionally eat meatless meals 4. Contribute to ecological or conservation organizations
Goldberg and Hartwick (1990)	Company reputation	Reputation among different publics (-3 = "a poor reputation," +3 = "a good reputation")	1. With its employees 2. With financial investors 3. With the U.S. public 4. With the Canadian public
Carlson and Grossbart (1988)	Coviewing television (parent/child)	Frequency of watching television with children (1 = "very seldom," 2 = "seldom," 3 = "sometimes," 4 = "often," 5 = "very often")	I watch television with my children on 1. Weekdays 2. Saturdays 3. Sundays

spondents (using the same strata proportions as in the original sample).

Altogether, 256 responses were obtained, of which 180 were usable (note that 173 cases are used in the analysis following listwise deletion of missing data); the distribution of the responses across size categories was practically identical to that of the original sample. To gather firsthand information with regard to reasons for nonresponse, a telephone follow-up of nonrespondents was undertaken. This showed that ineligibility (i.e., the firm was no longer exporting or the firm no longer existed) was the main reason for nonre-

sponse. Among eligible nonrespondents, the most common reasons were lack of time or company policy that prohibited the completion of any questionnaire. After adjustment for ineligibles (Wiseman and Billington 1984), the effective response rate came to 18.5%. This rate is comparable to other surveys conducted in an industrial setting (Jobber and Bleasdale 1987) and surveys on forecasting practices (e.g., White 1986). Although the telephone follow-ups gave no reason to suspect that nonresponse error was an issue of concern, further analysis involved a comparison of early and late responses, as recommended by Armstrong and Overton

(1977). A series of t-tests for independent samples failed to identify significant differences between early and late respondents, providing additional evidence that nonresponse bias was unlikely to be a major problem in this study.

Appendix B

MEASUREMENT ITEMS USED IN ILLUSTRATIVE MODELS

<i>Indicators for Resource Commitment Index</i>	
x ₁	Communication problems within our firm often complicate the preparation of export sales forecasts.
x ₂	Although we have a lot of data concerning our export markets, we do not really know how to use it in export sales forecast development.
x ₃	We have no problems in finding personnel with expertise in forecasting export sales.
x ₄	A major barrier to developing sound export sales forecasts is lack of relevant information.
x ₅	The costs of obtaining data useful for export sales forecasting purposes are often prohibitive.
x ₆	There is a lot of support from top management when it comes to export sales forecasting.
x ₇	Export sales forecasts are always prepared in a hurry, as our personnel are busy with other work.
x ₈	We have all the computer support we need for developing export sales forecasts.
x ₉	The people involved in export sales forecasting have a lot of knowledge about our export markets.
x ₁₀	The real problem in obtaining data useful for export sales forecasting is identifying the relevant sources rather than costs.
<i>Reflective Indicators Used in MIMIC Model</i>	
y ₁	Developing export forecasts is more difficult than preparing forecasts for the home market.
y ₂	The preparation of forecasts for the home market is less costly than for our export markets.
<i>Export Sales Forecast Performance Indicators</i>	
y ₃	Overall, we are as good in forecasting export sales as any firm in our industry.
y ₄	Compared to our competitors in export markets, our export sales forecasting capability is superior.

Notes: All items are scored as follows: 5 = "strongly agree," 4 = "agree," 3 = "neither agree nor disagree," 2 = "disagree," 1 = "strongly disagree."

REFERENCES

- 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), 453-60.
- Armstrong, J.B. and T.S. Overton (1977), "Estimating Non-response Bias in Mail Surveys," *Journal of Marketing Research*, 14 (August), 396-402.
- Babbie, Earl R. (1973), *Survey Research Methods*. Belmont, CA: Wadsworth.
- Bagozzi, Richard P. (1982), "The Role of Measurement in Theory Construction and Hypothesis Testing: Toward a Holistic Model," in *A Second Generation of Multivariate Analysis*, Vol. 1, C. Fornell, ed. New York: Praeger, 5-23.
- (1994), "Structural Equation Models in Marketing Research: Basic Principles," in *Principles of Marketing Research*, R. Bagozzi, ed. Oxford: Blackwell, 317-85.
- and Claes Fornell (1982), "Theoretical Concepts, Measurements, and Meaning," in *A Second Generation of Multivariate Analysis*, Vol. 1, C. Fornell, ed. New York: Praeger, 24-38.
- , ———, and D.F. Larcker (1981), "Canonical Correlation Analysis as a Special Case of a Structural Relations Model," *Multivariate Behavioral Research*, 16 (October), 437-54.
- Bearden, William O., Richard G. Netemeyer, and Mary F. Mobley (1993), *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Newbury Park, CA: Sage Publishing.
- Blalock, Hubert M., Jr. (1971), "Causal Models Involving Unmeasured Variables in Stimulus-Response Situations," in *Causal Models in the Social Sciences*, H.M. Blalock, ed. Chicago: Aldine, 335-47.
- Bollen, Kenneth (1984), "Multiple Indicators: Internal Consistency of No Necessary Relationship?" *Quality and Quantity*, 18, 377-85.
- (1989), *Structural Equations with Latent Variables*. New York: John Wiley & Sons.
- and Richard Lennox (1991), "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin*, 110 (2), 305-14.
- Bruner, Gordon C. and Paul J. Hensel (1992), *Marketing Scales Handbook: A Compilation of Multi-Item Measures*, Vol. 1. Chicago: American Marketing Association.
- and ——— (1993), "Multi-Item Scale Usage in Marketing Journals: 1980-1989," *Journal of the Academy of Marketing Science*, 21 (Fall), 339-44.
- and ——— (1996), *Marketing Scales Handbook: A Compilation of Multi-Item Measures*, Vol. 2. Chicago: American Marketing Association.
- Burke, Marian C. (1984), "Strategic Choice and Marketing Managers: An Examination of Business-Level Marketing Objectives," *Journal of Marketing Research*, 21 (November), 345-59.
- Carlson, Les and Sanford Grossbart (1988), "Parental Style and Consumer Socialization of Children," *Journal of Consumer Research*, 15 (June), 77-94.
- Churchill, Gilbert A., Jr. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (February), 64-73.
- and J. Paul Peter (1984), "Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis," *Journal of Marketing Research*, 21 (November), 360-75.
- Cohen, Patricia, Jacob Cohen, Jeanne Teresi, Margaret Marchi, and C. Noemi Velez (1990), "Problems in the Measurement Latent Variables in Structural Equations Causal Models," *Applied Psychological Measurement*, 14 (June), 183-96.
- Cudeck, Robert and Michael W. Browne (1983), "Cross-Validation of Covariance Structures," *Multivariate Behavioral Research*, 18, 147-67.
- Dalrymple, Douglas J. (1987), "Sales Forecasting Practices," *International Journal of Forecasting*, 3, 379-91.
- Daly, H.E. and J.B. Cobb (1989), *For the Common Good*. Boston: Beacon.
- Danes, Jeffrey E. and O. Karl Mann (1984), "Unidimensional Measurement and Structural Equation Models with Latent Variables," *Journal of Business Research*, 12 (3), 337-51.
- DeVellis, Robert F. (1991), *Scale Development: Theory and Applications*. Newbury Park, CA: Sage Publications.
- Dichtl, Erwin and Hans-Georg Köglmayr (1986), "Country Risk Ratings," *Management International Review*, 26 (49), 4-11.
- Economist Intelligence Unit (1996), *Global EIU Market Indexes*. London: Economist Intelligence Unit.
- Ennew, Christine T., Geoffrey V. Reed, and Martin R. Binks (1993), "Importance-Performance Analysis and the Measurement of Service Quality," *European Journal of Marketing*, 27 (2), 59-70.
- Fornell, Claes and Fred L. Bookstein (1982), "A Comparative Analysis of Two Structural Equation Models: LISREL and PLS Applied to Market Data," in *A Second Generation of*

- Multivariate Analysis*, Vol. 1, C. Fornell, ed. New York: Praeger, 289–324.
- and Jaesung Cha (1994), “Partial Least Squares,” in *Advanced Methods of Marketing Research*, R.P. Bagozzi, ed. Oxford: Blackwell, 52–78.
- , Byong-Duk Rhee, and Youjae Yi (1991), “Direct Regression, Reverse Regression, and Covariance Structure Analysis,” *Marketing Letters*, 2 (3), 309–20.
- Gaski, John F. and John R. Nevin (1985), “The Differential Effects of Exercised and Unexercised Power Sources in a Marketing Channel,” *Journal of Marketing Research*, 22 (May), 130–42.
- Goldberg, Marvin E. and Jon Hartwick (1990), “The Effects of Advertiser Reputation and Extremity of Advertising Claim on Advertising Effectiveness,” *Journal of Consumer Research*, 17 (September), 172–79.
- Griliches, Zvi (1971), *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Cambridge, MA: Harvard University Press.
- Hauser, Robert M. (1971), *Socioeconomic Background and Educational Performance*, Rose Monograph Series. Washington, DC: American Sociological Association.
- (1973), “Disaggregating a Social-Psychological Model of Educational Attainment,” in *Structural Equation Models in the Social Sciences*, Arthur S. Goldberger and Otis Dudley Duncan, eds. New York: Seminar Press, 255–84.
- and A.S. Goldberger (1971), “The Treatment of Unobservable Variables in Path Analysis,” in *Sociological Methodology 1971*, H.L. Costner, ed. San Francisco: Jossey-Bass, 81–117.
- Homburg, Christian, John P. Workman Jr., and Harley Krohmer (1999), “Marketing’s Influence Within the Firm,” *Journal of Marketing*, 63 (April), 1–17.
- Horn, Robert, V. (1983), *Statistical Indicators for the Economic and Social Sciences*. Cambridge, UK: Cambridge University Press.
- Jacoby, Jacob (1978), “Consumer Research: A State of the Art Review,” *Journal of Marketing*, 42 (April), 87–96.
- Jobber, D. and M.J.R. Bleasdale (1987), “Interviewing in an Industrial Market Research: The State-of-the-Art,” *Quarterly Review of Marketing*, 12 (2), 7–11.
- Johnston, D.F. (1988), “Towards a Comprehensive Quality of Life Index,” *Social Indicators Research*, 20, 473–96.
- Jöreskog, Karl G. 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*, 10, 631–39.
- Kleinbaum, D.G., L. Kupper, and K.E. Muller (1988), *Applied Regression Analysis and Other Multivariable Methods*, 2d ed. Boston: PWS-Kent.
- Long, J. Scott (1983), *Confirmatory Factor Analysis*. London; Beverly Hills, CA: Sage Publications.
- Lumpkin, James R. and James B. Hunt (1989), “Mobility as an Influence on Retail Patronage Behavior of the Elderly: Testing Conventional Wisdom,” *Journal of the Academy of Marketing Science*, 17 (Winter), 1–12.
- MacCallum, Robert C. and Michael W. Browne (1993), “The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues,” *Psychological Bulletin*, 114 (3), 533–41.
- Mahmoud, Essam (1987), “The Evaluation of Forecasts,” in *The Handbook of Forecasting: A Manager’s Guide*, Spiros Makridakis and Steven C. Wheelwright, eds. New York: John Wiley & Sons.
- McKee, Daryl O., P. Rajan Varadarajan, and William M. Pride (1989), “Strategic Adaptability and Firm Performance: A Market Contingent Perspective,” *Journal of Marketing*, 53 (July), 21–35.
- Moorman, Christine and Erika Matulich (1993), “A Model of Consumers’ Preventive Health Motivation and Health Ability,” *Journal of Consumer Research*, 20 (September), 208–28.
- Namboodiri, N. Krishnan, Lewis F. Carter, and Hubert M. Blalock Jr. (1975), *Applied Multivariate Analysis and Experimental Designs*. New York: McGraw-Hill.
- Nunnally, Jum C. and Ira H. Bernstein (1994), *Psychometric Theory*, 3d ed. New York: McGraw-Hill.
- Peter, J. Paul (1979), “Reliability: A Review of Psychometric Basics and Recent Marketing Practices,” *Journal of Marketing Research*, 16 (February), 6–17.
- (1981), “Construct Validity: A Review of Basic Issues and Marketing Practices,” *Journal of Marketing Research*, 18 (May), 133–45.
- Richins, Marsha L. and Scott Dawson (1992), “A Consumer Values Orientation for Materialism and Its Measurement: Scale Development and Validation,” *Journal of Consumer Research*, 19 (December), 303–16.
- Small, R.L. (1980), *Sales Forecasting in Canada: A Survey of Practices*, Study No. 66. Ottawa, Canada: The Conference Board of Canada.
- Spector, Paul E. (1992), *Summated Ratings Scales Construction*. Newbury Park, CA: Sage Publications.
- Steenkamp, Jan-Benedict E.M. and H.C.M. van Trijp (1991), “The Use of LISREL in Validating Marketing Constructs,” *International Journal of Research in Marketing*, 8, 283–99.
- United Nations Development Program (1990), *Human Development Index*. New York: United Nations Development Program and Oxford University Press.
- White, H.R. (1986), *Sales Forecasting: Timesaving and Profit-Making Strategies That Work*. London: Scott, Foresman and Company.
- Wiseman, F. and M. Billington (1984), “Comment on a Standard Definition of Response Rates,” *Journal of Marketing Research*, 21 (August), 336–38.