EVALUATING EFFECT, COMPOSITE, AND CAUSAL INDICATORS IN STRUCTURAL EQUATION MODELS¹

Kenneth A. Bollen

Carolina Population Center, Department of Sociology, University of North Carolina at Chapel Hill,
Chapel Hill, NC 27599-3210 U.S.A. {bollen@unc.edu}

Although the literature on alternatives to effect indicators is growing, there has been little attention given to evaluating causal and composite (formative) indicators. This paper provides an overview of this topic by contrasting ways of assessing the validity of effect and causal indicators in structural equation models (SEMs). It also draws a distinction between composite (formative) indicators and causal indicators and argues that validity is most relevant to the latter. Sound validity assessment of indicators is dependent on having an adequate overall model fit and on the relative stability of the parameter estimates for the latent variable and indicators as they appear in different models. If the overall fit and stability of estimates are adequate, then a researcher can assess validity using the unstandardized and standardized validity coefficients and the unique validity variance estimate. With multiple causal indicators or with effect indicators influenced by multiple latent variables, collinearity diagnostics are useful. These results are illustrated with a number of correctly and incorrectly specified hypothetical models.

Keywords: Causal indicators, effect indicators, formative indicators, reflective indicators, measurement, validity, structural equation models, scale construction, composites

Introduction

After years of neglect, alternative measures to effect indicators are finally receiving more attention in a variety of social and behavioral science disciplines. As consideration of causal, formative, and composite indicators becomes more common so do questions about how to analyze them. Some research on identification and estimation of causal indicator models is available (e.g., Bollen 1989, pp. 311-313, 331; Bollen and Davis 2009a; Diamantopoulos et al. 2008; MacCallum and Browne 1993). But the validity, selection, and elimination of causal, composite, or formative indicators have received scant attention. The focus of this paper is on these issues.

Some authors have broached these topics in the course of writing about causal or formative indicators (e.g., Bollen 1989, pp. 222-223; Bollen and Lennox 1991; Diamantopoulos et al. 2008; Jarvis et al. 2003; Petter et al. 2007). One clear message from these works is that researchers should not use effect indicator selection tools such as item-total correlations or Cronbach's alpha when evaluating other types of indicators.

Another common theme is the greater importance of individual causal indicators in measuring a concept versus the interchangeability of effect indicators. For instance, Bollen and Lennox (1991, p. 307-308) argue that we can sample effect indicators, but we need a "census" of causal indicators. Jarvis et al. (2003, p. 202) state that "dropping a causal indicator may omit a unique part of the composite latent construct and change the meaning of the variable." Although these discussions are helpful, our fields would benefit from further discussion.

¹Detmar Straub was the accepting senior editor for this paper. Thomas Stafford served as the associate editor.

The purpose of this paper is to present ideas on evaluating the validity of causal, composite, or formative indicators in measurement models. This task is facilitated by considering the selection and elimination of effect indicators in that the similarities and differences between these traditional and nontraditional indicators are illuminating. A division is also made between models whose only possible flaw is that one or more of the indicators are not a measure of the intended concept and a second set of models where more serious structural misspecifications are present. But the task is the same: evaluating the validity of indicators.

The next section presents background information to clarify my terminology, the role played by theory, a summary table of validity checks, and my position on when to eliminate indicators. Next are sections that examine these different indicators in models that are correctly specified, followed by a section on these indicators in incorrectly specified models. A concluding section summarizes the discussion.

Background Information

Terminology

Structural equation models (SEMs) and concerns about measurement cover a broad range of disciplines. The diversity of disciplines and groups writing on SEMs has resulted in a diversity of terms with sometimes the same term applied in different ways. For this reason I clarify my basic terminology and assumptions.

Concepts are the starting point in measurement. Concepts refer to ideas that have some unity or something in common. The meaning of a concept is spelled out in a theoretical definition. The dimensionality of a concept is the number of distinct components that it encompasses. Each component is analytically separate from other components, but it is not easily subdivided. The theoretical definition should make clear the number of dimensions in a concept. Each dimension is represented by a single latent variable in a SEM, so if a concept is unidimensional we need a single latent variable; if it is bidimensional, we will need two latent variables.

SEMs are sets of equations that encapsulate the relationships among the latent variables, observed variables, and error variables. Parameters such as coefficients, variances, and covariances from the same model are estimable in different ways (e.g., ML, PLS) with the different estimators having different properties. So the model and estimation method are distinct and should not be confused.

Indicators are observed variables that measure a latent variable. I use indicator and measure to mean the same thing. One primary distinction is whether the indicator is influenced by the latent variable or vice versa. Effect (or reflective) indicators depend on the latent variable. Causal indicators affect the latent variable. Blalock (1964) appears to be the first social or behavioral scientist to make the distinction between causal and effect indicators. The term *formative indicator* came much later than causal indicators and had a meaning more restrictive than causal indicators. Fornell and Bookstein (1982, p. 292) defined formative indicators as

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.

As this quote illustrates, an important difference between causal indicators and formative indicators is that formative indicators completely determine the “latent” variable. Crudely put, the R-square for the composite variable is 1.0 since it is a linear combination of the formative indicators. Causal indicators have no such restriction.

Another distinction is that causal indicators should have conceptual unity in that all the variables should correspond to the definition of the concept whereas formative indicators are largely variables that define a convenient composite variable where conceptual unity is not a requirement. That is, the composite variable with its formative indicators might just be a useful summary device for the effect of several variables on other variables (e.g., see Heise’s [1972] sheaf coefficient). In other situations, researchers’ use of the term *formative indicators* departs from its original meaning and they appear to mean the same as causal indicators.

To avoid the ambiguity of the term formative indicator, I use *composite indicator* to refer to an indicator that is part of a linear composite variable (Grace and Bollen 2006). Although composite indicators might share similarities, they need not have conceptual unity. In other words, I use composite indicators to refer to the original use and sometimes current use of formative indicators. Causal indicators differ from the composite indicators in that causal indicators tap a unidimensional concept (or dimension of a concept) and the latent variable that represents the concept is not completely determined by the causal indicators. That is, an error term affects the latent variable as well. As I will explain later, whether an indicator is a composite or causal indicator has implications for evaluating validity.

The Necessity of Theory

Theory enters measurement throughout the process. We need theory to define a concept and to pick out its dimensions. We need theory to develop or to select indicators that match the theoretical definition and its dimensions. Theory also enters in determining whether the indicator influences the latent variable or vice versa. Each relation specified between a latent variable and an indicator is a theoretical hypothesis. Empirical estimates and tests aid in assessing whether the set of relations are consistent with the data. However, empirical consistency between the data and model alone is insufficient to support the validity of measures. There must be theory behind the relations to make the validity argument credible.

On the other hand, theory might suggest that an indicator is valid, but the empirical estimates linking the latent variable and indicator might be weak to zero. If this occurs in a well-fitting model, these results cast doubt on the validity of the measure even if theory suggests its validity. The burden would be on the researcher to explain why a valid measure is not related to its latent variable. It could be that the measure is valid, but that the model is misspecified in a way that leads to low values of the validity statistics. This explanation is most convincing if the researcher is specific on the nature of the misspecifications and why these problems would lead to low validity estimates.

Although I will not repeatedly discuss this in the remaining sections, I assume throughout that there is measurement theory to support the models and the relations between the latent variables and indicators within them. The empirical estimation and tests are intended to evaluate these ideas.

Validity Checks

I define the validity of a measure as follows: “*The validity of a measure x_i of ξ_i is the magnitude of the direct structural relation between ξ_i and x_i* ” (Bollen 1989, p. 197). Causal indicators occur when the observed indicators have a direct effect on the latent variable whereas effect indicators are indicators where the latent variable has a direct effect on the indicator.

Table 1 gives a summary of the validity assessment that I would recommend (Bollen 1989, pp. 197-206). The first step is to compare the definition of the concept (or dimension of the concept) that is the object of measurement to the indicator and to check whether the indicator corresponds to this definition. This is a type of face validity check in that the researcher assesses whether it is sensible to view the indicator as matching the idea embodied by the concept.

The second check in Table 1 is to examine the overall fit of the model. Assuming that the researcher is using an estimator such as ML that provides a test of whether the model adequately reproduces the variances, covariances, and means of the observed variable, a chi-square test statistic and various fit indexes are available to help in assessing model fit. The fit indexes to choose and the weight to give to them is subject to controversy and this is an issue that cannot be easily resolved. However, the researcher will need to make the judgment as to whether model fit is adequate enough to use the other validity checks that are specific to the individual indicators.

Another aspect of assessing measurement validity is having the factor and measures embedded in a fuller model with additional causes and consequences of the main latent variable. Part of assessing validity is determining whether a measure and latent variable relate to other distinct latent variables as expected. Sometimes this is called *external validity*. For instance, there often are variables thought to be determinants or consequences of a specific latent variable. If I estimate a model with this full set of relationships and I find the expected effects, then this is further support for the validity of the latent variable and its measures. Failure to find such relationships undermines validity even if the preceding validity measures are satisfactory. This idea is closely related to construct validity where part of the validity assessment is done by evaluating if a construct relates to other constructs as predicted by theory or substantive considerations. When such variables are available, they enhance the validation process.

If the researcher is satisfied with model fit, the remaining validity checks bear examination. Using the definition of validity from the previous section, the factor loadings (λ) for effect indicators and the regression coefficients (γ) for causal indicators provide the *unstandardized validity coefficient*. Here the researcher makes sure that the coefficients have the correct sign and looks for their statistical and substantive significance. The *standardized validity coefficients* are treated similarly except that these coefficients give the expected standard deviation shift in dependent variable for a one standard deviation shift in the other variable net of other influences on the dependent variable.

The unique validity variance in the last row gives the variance in the dependent variable uniquely attributable to a variable. If we have an effect indicator, it gives the explained variance uniquely attributable to the latent variable that it measures. If we have a causal indicator, it gives the uniquely attributable variance in the latent variable due to the causal indicator. When a single latent variable influences the effect indicator, the unique validity variance is the same as the usual R^2 for an indicator. Typically there are two or more causal indicators,

Table 1. Validity Assessment for Causal and Effect Indicators

Validity Check	What to Check	What to Look For
Definition of Concept	Indicator	Indicator corresponds to definition
Overall Model Fit	Chi-square and fit indexes	High p-value and acceptable fit indexes
External Validity	Relation to other latent variables	Relations consistent with theory
Unstandardized Validity Coefficients	Coefficients (λ , γ)	Correct sign, statistical and substantive significance
Standardized Validity Coefficients	Coefficients (standardized λ , γ)	Correct sign, statistical and substantive significance
Unique Validity Variance	Incremental R^2	Explained variance added

so in this situation the unique validity variance is the additional explained variance due to a specific causal indicator once all other variables enter the latent variable equation.

I will say more about these validity checks in the other sections.

Indicator Elimination

Viewpoints on when an indicator should be eliminated vary with one extreme answering “never” and the other suggesting whenever an indicator is found inadequate. I would place myself between these two endpoints. If a definition of a concept (or a dimension of it) leads a researcher to include an indicator as a measure of it, this is strong *a priori* evidence to keep it. On the other hand, if repeated empirical tests find a statistically and substantively weak relation of the indicator to its purported latent variable, then it is difficult to maintain the validity of the indicator. The distinction between causal and effect indicators has relevance here. Effect indicators of the same latent variable that have the same explained variance are in some sense interchangeable. Although more indicators are generally better than fewer, which indicators are used is less important.

With causal indicators, removal of an indicator can be more consequential. The causal indicators help to determine the latent variable, so that removal of a causal indicator might change the nature of the latent variable. But here also if the coefficient of the causal indicator is not significant or is the wrong sign, it is possible that you have an invalid causal indicator. Decisions on whether to eliminate indicators must be made taking account of the theoretical appropriateness of the indicator and its empirical performance in the researcher’s and the studies of others.

Indicators in Structurally Correct Models

We can analytically separate the task of indicator evaluation under two broad conditions. The first condition is when a researcher has correctly identified the dimensionality and the latent variables that represent a concept. Furthermore, most if not all of the indicators selected to measure the concept are correct in that there is a direct relationship between the indicator and the latent variable that represents the concept. The goal is to find which indicators do not adequately measure the concept. I refer to these as *structurally correct models* in that these are essentially valid models where the only possible flaw is that some indicators might have low to near zero validity.

It will help our understanding of the validity of causal indicators if we first present the better-known effect indicator model and then contrast it with causal indicators. To make the discussion more concrete, I consider five indicators and one latent variable. In one situation I have effect indicators, in another I have causal indicators, and finally I have composite indicators.

Five Effect Indicators, One Latent Variable

The equations of the measurement model² on which I focus are

²These equations describe the hypothesized relation between the indicators and their latent variable without assuming that any particular estimator (e.g., ML or PLS) is being used.

$$\begin{aligned}
y_1 &= \alpha_1 + \lambda_1\eta + \varepsilon_1 \\
y_2 &= \alpha_2 + \lambda_2\eta + \varepsilon_2 \\
y_3 &= \alpha_3 + \lambda_3\eta + \varepsilon_3 \\
y_4 &= \alpha_4 + \lambda_4\eta + \varepsilon_4 \\
y_5 &= \alpha_5 + \lambda_5\eta + \varepsilon_5
\end{aligned}
\tag{1}$$

where y_j is the j^{th} effect indicator of η , α_j is the intercept in the measurement equation, λ_j is the factor loading or regression coefficient of the effect of the latent factor on y_j , η is the factor or latent variable that we hope to measure, and ε_j is the unique factor that contains all other influences on y_j besides η . I assume that $E(\varepsilon_j) = 0$, $\text{COV}(\varepsilon_j, \varepsilon_j) = 0$, and that $\text{COV}(\varepsilon_j, \eta) = 0$ for $i, j = 1, 2, 3, 4, 5$ and $i \neq j$. This is equivalent to a single factor model with five indicators and uncorrelated unique factors (errors).³ The path diagram is in Figure 1.

Suppose that I am not sure that all five measures are valid. How can I examine their validity and discard any that are not valid? Assume that I am certain of the validity of at least one of these measures, say y_1 , so I scale the latent η to this indicator by setting $\lambda_1 = 1$ and $\alpha_1 = 0$ leading to $y_1 = \eta + \varepsilon_1$. Scaling the latent variable is sufficient to identify this model; that is, it is possible to find unique values for the remaining intercepts, factor loadings, and other parameters in this model.⁴ (Without scaling the latent variable, the model is underidentified.)

Referring to Table 1, the first check for the remaining indicators is whether they correspond to the definition of the concept that is represented by η . If they do not, then this is sufficient grounds to eliminate an indicator. Assuming that they do match the definition, then the next step is to assess whether the model represented in equation (1) fits the data. This requires that we estimate the model with a procedure that provides overall fit statistics. For instance, if the ML estimator is appropriate, then a likelihood ratio chi-square test is available to compare the hypothesized model to a saturated model. Adjustments to this chi-square test are readily available when the data exhibit significant nonnormality. There also are additional fit indexes that can supplement the chi-square test. If the fit is not adequate, then the researcher will need to find alternative specifications that might be better suited to the data. If the fit is judged to be good, then the other means of assessing validity are turned to next.

³Here, as with the causal indicator and composite indicator model, we can estimate a measurement model and need not form indexes prior to analysis.

⁴If the scaling indicator has zero validity, I would expect estimation problems and even if estimates are obtained, the R-square for this indicator would be near zero.

Employing a consistent estimator (e.g., maximum likelihood, two-stage least squares) of the parameters in this model, I can begin to assess the validity of the indicators.⁵ The factor loadings ($\hat{\lambda}$) for effect indicators provide the *unstandardized validity coefficient*. Here the researcher makes sure that the coefficients have the correct sign and looks for their statistical and substantive significance. It provides the expected difference in the indicator (y_j) given a one unit difference in the latent variable (η). If a measure is valid, then I expect that the loading will be large in a substantive sense and that its estimate is statistically significant as well. Invalidity would be suggested by the opposite: a substantively small or statistically insignificant factor loading. For instance, suppose that the scaling indicator variable, y_1 , has a range from 0 to 100 as does y_2 . If y_2 has a coefficient (λ_2) of 0.1, then its validity appears much less than y_1 whose coefficient is 1.0.

Closely related is the *standardized validity coefficient* which gives the expected number of standard deviations that the indicator (y_j) differs for a one standard deviation difference in the latent variable (η). The unique validity variance provides the variance in the measure that is uniquely attributable to a specific latent variable. Assuming that only η directly affects y_j , the *unique validity variance* is equivalent to the R-square for the indicator (y_j). This in turn is the amount of variation in the indicator explained by the latent variable. Finally, the collinearity analysis is not relevant for this example because only a single latent variable affects the measure so there is no correlation of the latent variable with anything else. This would be an issue if, say, there were two correlated latent variables influencing the same measure.

Another aspect of assessing measurement validity is having the factor and measures embedded in a fuller model with additional causes and consequences of η . In this hypothetical example and sometimes in practice, we do not have additional latent variables for external validity checks but this would be possible if this measurement structure were embedded in a larger model.

Returning to the primary task of indicator selection in the effect indicator model in Figure 1, I can now provide some guidance. First, each indicator should have a sufficiently

⁵A consistent estimator is an estimator that converges on the true parameter values of the model as the sample size goes to infinity. A related concept is *asymptotically unbiased*. Asymptotically unbiased means that in large samples the expected value of the estimator is the population parameter. The usual maximum likelihood estimator that I use in this paper has these two properties. Partial least squares (PLS)—sometimes called a component-based SEM estimator—does not, so this makes it difficult to use PLS to make validity assessments. That is, if you have biased estimates of parameters in the model, using these to guide validity assessments could be misleading.

Figure 1. Five Effect Indicators, One Latent Variable

large and statistically significant unstandardized and standardized factor loading and R-square. For instance, an indicator with an R-square of less than 0.10 would be undesirable in that 90 percent or more of its variance would be due to error. Similarly a statistically nonsignificant factor loading would suggest a possibly invalid indicator. Any indicator with a statistically insignificant loading is a candidate for elimination since it suggests that the measure is not valid. Finally, when η is embedded in a fuller model it and its effect indicators should behave similarly to when it is not. For instance, if a researcher reanalyzes this model keeping y_1 as the scaling indicator, but using only three indicators, the factor loadings for the same variable in the three indicator and the original five indicator model should be within sampling fluctuations of each other if the indicators are valid. Large differences suggest structurally misspecified models and problems with indicator validity.⁶ These judgments need to be made in the context of what other researchers have found with these indicators and taking account of the overall fit of the model. But measures that perform poorly on these criteria are candidates for elimination. The next subsection contrasts these criteria in the case of causal indicators.

Five Causal Indicators, One Latent Variable

In this section I move from effect indicators to causal indicators. The equation corresponding to causal indicators is

$$\eta = \alpha_{\eta} + \gamma_{\eta 1}x_1 + \gamma_{\eta 2}x_2 + \gamma_{\eta 3}x_3 + \gamma_{\eta 4}x_4 + \gamma_{\eta 5}x_5 + \zeta \quad (2)$$

where α_{η} is the intercept, x_j is a causal indicator, $\gamma_{\eta j}$ is the coefficient of x_j and its effect on η , and ζ is the error or

⁶This is a point where the estimator employed is relevant in that some estimators are less sensitive to structural misspecifications than others. For example, the ML estimator is generally more sensitive to structural misspecifications than the latent variable two stage least squares estimator.

disturbance with $E(\zeta) = 0$ and $COV(x_j, \zeta) = 0$ for $j = 1$ to 5. Figure 2 is a path diagram of the equation.

The presence of the error (ζ) in equation (2) is in recognition that other variables not in the model are likely to determine the values of the latent variable. It seems unlikely that we could identify and measure all of the variables that go into the latent variable and the error collects together all of those omitted causes.

An important difference from the effect indicator model in equation (1) is that the causal indicator model in (2) is not identified. This means that without additional information I cannot find unique values for the parameters in the model and hence cannot assess indicator validity.

To overcome this obstacle I assume that I have two effect indicators added to this model. Now instead of equation (2) alone, I also have

$$\begin{aligned} y_1 &= \alpha_1 + \lambda_1\eta + \varepsilon_1 \\ y_2 &= \alpha_2 + \lambda_2\eta + \varepsilon_2 \end{aligned} \quad (3)$$

with the same assumptions made as I did for equation (1). These two effect indicators are sufficient to identify the causal indicator model once I scale the latent variable (e.g., $\lambda_1 = 1$, $\alpha_1 = 0$). I now turn to assessing the validity of the causal indicators.⁷

Referring to Table 1 again, our starting point is the same as with effect indicators: Do the causal indicators correspond to the definition of the concept? The causal indicators should be initially chosen for their face validity in that on the face of it they seem to tap the concept (or dimension of a concept) that is being measured. If not, then they should not be used.

⁷The model becomes a MIMIC model (Jöreskog and Goldberger 1975) when these two effect indicators are added.

Figure 2. Five Causal Indicators, One Latent Variable

Our second validity criterion is to check the overall fit of the model. The model as specified is overidentified, so using an estimator like ML we can test its fit compared to the saturated model and supplement this with other fit indexes. A poor fit means we need to stop to determine the source of the problem before going further. A good fit suggests that we move on to the other validity checks.

The definition of validity given earlier refers to the direct structural relation between the measure and the latent variable. In the causal indicator model, the unstandardized validity coefficient is $\gamma_{\eta j}$ and the standardized validity coefficient is its standardized counterpart⁸ and these provide evidence relevant to the validity of the causal indicators. A researcher should examine the sign, magnitude, and statistical significance of these coefficients. In addition, the unique validity variance reveals the incremental variance explained by each causal indicator. A valid causal indicator should have a strong effect on η as gauged by these standards.

It is impossible to give exact numerical values for cutoff for each of these validity measures since the minimum validity acceptable would depend on several factors such as the findings from prior studies, the degree to which collinearity prevents crisp estimates of effects, the difficulty of measuring the latent variable, etc. However, a coefficient of a causal indicator with the wrong sign or that is not statistically significant would appear to be invalid and a candidate for exclusion. Alternatively, a causal indicator with a correct sign and statistically significant coefficient with an increase of the squared multiple correlation by 10 percent or more has support for its validity.

⁸More specifically, the standardized validity coefficient is

$$\gamma_{\eta j} \left(\frac{s.d.(x_j)}{s.d.(\eta)} \right).$$

Before eliminating a causal indicator, a researcher must consider whether the allegedly invalid causal indicator suffers from high multicollinearity with the other causal indicators. Since η has more than one determinant, our validity assessment is more complicated than the previous effect indicator example in that we need to assess the impact of each x_j *separately* from the other causal indicators. If x_j were uncorrelated (or nearly so) with all other causal indicators, then the task would be easier in that I would not attribute to one causal indicator the effects of another. But when the causal indicators are highly correlated, there is the potential for the well-known multiple regression problem of multicollinearity: An explanatory variable (a causal indicator) is highly correlated with the other explanatory variables (causal indicators) and it is more difficult to estimate its unique effects. This difficulty is reflected in larger standard errors than if there were lower collinearity and these larger standard errors imply less precise (more variable) coefficient estimates and validity coefficients. In the situation of high multicollinearity, a competing explanation for low unstandardized and standardized validity coefficients and low unique validity variance is that the causal indicator was so highly associated with the other causal indicators that estimates of distinct impact were not possible.

One factor that makes this collinearity issue less worrisome for causal indicators is that such indicators need not correlate (Bollen 1984). This means in general we would not expect high degrees of collinearity among causal indicators, although it is possible. Fortunately, there are regression diagnostics available to assess when collinearity is severe or mild (e.g., Belsley et al. 1980) and many of these are adaptable to the causal indicator model.

Are effect indicators immune from the collinearity problem? The brief answer is “no.” We did not face this issue in equation (1) because each effect indicator was only influenced by a single latent variable. But sometimes effect indicators are depended on by two or more latent variables and if these

latent variables are highly correlated, the same collinearity issue might emerge. If an effect indicator has factor complexity of one, then the collinearity issue is absent.

Five Composite Indicators, One Composite

The distinction between composite (formative) and causal indicators is subtle. I represent the composite indicator model for five indicators below:

$$C = \alpha_C + \gamma_{C1}x_1 + \gamma_{C2}x_2 + \gamma_{C3}x_3 + \gamma_{C4}x_4 + \gamma_{C5}x_5 \quad (4)$$

This equation shares with equation (2) that the indicators (x 's) help set the value of the left-hand side variable, but the differences from causal indicators are more important than this similarity. For one thing, the equation with composite indicators has no error (or disturbance) term since the composite indicators completely determine the composite. The composite variable C rather than the latent variable η is on the left hand side of the equation to reflect that we are dealing with a linear composite variable (C) rather than a latent variable (η).

At least as important, but less obvious, is that composite indicators need not share unidimensional conceptual unity. That is, composite indicators might be combined into a composite as a way to conveniently summarize the effect of several variables that do not tap the same concept although they may share a similar "theme." For instance, age, sex, gender, and race effects could be summarized in a demographic composite even though in this context demographic is not seen as a unidimensional latent variable. Indeed, another researcher might have a different set of demographic variables to form a demographic composite.

A troubling aspect of composite indicators not sharing unidimensional conceptual unity is whether it even makes sense to discuss the validity of these variables. Using either the preceding definition of validity (*The validity of a measure x_i of ξ_j is the magnitude of the direct structural relation between ξ_j and x_i*) or the less formal definition of whether a variable measures what it is suppose to measure, if the composite indicators are just convenient ingredients to form a linear composite variable, it is not clear that assessments of validity apply. In other words, if we have no unidimensional theoretical concept in mind, how can we ask whether the composite indicators measure it? With composite indicators and a composite variable, the goal might not be to measure a scientific concept, but more to have a summary of the effects of several variables.

Even if the unidimensional concept criterion is satisfied by the composite indicators, it seems unlikely that there are many

situations where an error term would be absent as in (4). This would mean that the latent variable that represents the unidimensional concept is an exact linear function of its indicators, which would seem to be a rarity.⁹ Given that the concept of validity makes little sense for arbitrary composites formed from composite indicators, I will not discuss the validity measures for composite indicators.

Many researchers gloss over the distinction between composite indicators and causal indicators that I make here.¹⁰ This distinction likely underlies some of the controversies in this area. For instance, if composite indicators form a composite without having unidimensional conceptual unity, then their coefficients could well be unstable as the composite is allowed to influence different outcome variables. Returning to the demographic composite, we would not expect the effects of age, race, and gender to be the same regardless of the outcome variable. On the other hand, if these are causal indicators intended to tap a single latent variable, then the coefficients of the causal indicators should be stable across different outcomes. This could explain some of the differences expressed in Howell et al. (2007) versus Bollen (2007) where the former is referring to composite (formative) indicators and the latter refers to causal indicators.

Summary of Structurally Correct Models

When we have "clean" models in which the researcher has essentially the correct specification and the only question is whether one or more of either the effect or causal indicators are valid, the task of indicator assessment is simplified. We can use the regression coefficients from the causal indicator to the latent variable (as in (2)) or from the latent variable to the effect indicator (as in (1)) to form the unstandardized and standardized validity coefficients and test their statistical significance to assess measurement validity. And we can use the unique validity variance as another measure of an indicator's validity. Low or statistically insignificant values on these measures are symptoms of low validity and suggest potential removal of such measures. Moderate to high values reinforce their validity. Exact cutoff values for indicators do not exist since their performance must be assessed in the context of what prior research has found under similar conditions.

⁹Linear identities might be present where this formative formulation makes sense. For instance, Profit = (Sales Price – Cost) or Population Size = Births – Deaths +/- Net Migration are identities that require no error term assuming accurate measures.

¹⁰My earlier writings have not made this distinction clear.

One complication with causal indicators is that we cannot identify or estimate any of the validity measures in a model when only causal indicators are present. The latent variable with causal indicators must emit at least one path to identify and estimate the unstandardized validity coefficients and it must emit at least two paths to identify and estimate the standardized validity measures (Bollen and Davis 2009b).

The main remaining difficulty in indicator selection occurs with causal indicators or with effect indicators that depend on more than one collinear latent variable. In either of these cases, high collinearity might make it difficult to obtain sharp estimates of the distinct effects of the explanatory variables. In practice, causal indicators are often not highly collinear and effect indicators often depend on one latent variable, so the collinearity issue is less prevalent. Finally, composite indicators as an arbitrary linear combination of indicators are not subject to validity analysis in that the composite need not represent a unidimensional concept so the question of validity is not appropriate.

Indicators in Structurally Misspecified Models

The discussion to this point has focused on models that are structurally correctly specified. Real world modeling is more complex and makes the task of indicator selection more complicated. I turn now to some of these issues. As before, I treat effect indicator models first and then turn to causal indicator models to facilitate understanding the similarities and differences between these models. The effect indicators are included to contrast the similarities and differences between causal and effect indicators. For the reasons given in the last subsection, I do not discuss composite indicators further.

Five Effect Indicators, Two Latent Variables

Suppose we once again have five effect indicators, but now have two latent variables in the true model. The equations are

$$\begin{aligned} y_1 &= \eta_1 + \varepsilon_1 \\ y_2 &= \alpha_2 + \lambda_{21} \eta_1 + \varepsilon_2 \\ y_3 &= \alpha_3 + \lambda_{31} \eta_1 + \lambda_{32} \eta_2 + \varepsilon_3 \\ y_4 &= \alpha_4 + \lambda_{42} \eta_2 + \varepsilon_4 \\ y_5 &= \eta_2 + \varepsilon_5 \end{aligned} \quad (5)$$

where I make the usual assumptions that the unique factors (ε_j) have means of zero and are uncorrelated with the latent variables and with each other. Figure 3a is a path diagram of

this model. In practice, researchers do not know the true generating model and are likely to make specification mistakes. Figures 3b to 3d are three examples of misspecified models. In Figure 3b, I mistakenly assume a single factor for all five indicators; Figure 3c also assumes a single factor but only uses y_1 , y_3 , and y_5 as indicators; and Figure 3d has a single factor but takes y_1 , y_2 , and y_3 as the only indicators.

Simulating data that conforms to Figure 3a ($\lambda_{11} = 1$, $\lambda_{21} = .8$, $\lambda_{31} = .6$, $\lambda_{32} = .5$, $\lambda_{42} = .8$, $\lambda_{52} = 1$, $\alpha_1 = 0$, $V(\varepsilon_j) = 1$, $V(\eta_k) = 1$, $COV(\eta_1, \eta_2) = .6$, $N = 1000$), I estimated the factor loadings and R-squares for the models that matched Figures 3a to 3d using the ML estimator. These results are reported in Table 2. The first columns with the results for the true model in Figure 3a show estimates that are within two standard errors of the population values. The low chi-square and high p-value and the other fit indices consistently point toward a good fitting model. I could interpret the validity measures for this correct version of the model as I illustrated with the other correct models in the first part of the paper, but will not do so. Rather I examine what happens in the incorrect models.

Figure 3b mistakenly has one factor rather than two. All the factor loadings remain statistically and substantively significant, which might lead a researcher to view these as valid measures. But we know that y_4 and y_5 are invalid measures of the first factor (η_1) and that y_3 measures both η_2 and η_1 , although the model only permits η_1 to have an impact.

In applied settings we are not privileged to the knowledge of the true model, so how can we know that there is a validity issue with some of the measures in Figure 3b? This is where the previous discussion about overall model fit is relevant. The chi-square of 70.8 with 5 *df* is highly statistically significant for the model in Figure 3b. The Relative Noncentrality Index (RNI) and Incremental Fit Index (IFI) pass conventional standards, but the Tucker-Lewis Index (TLI) and Bayesian Information Criterion (BIC) point toward an unacceptable fit.¹¹ The overall poor fit for some of the measures serves as a cautionary note about structural misspecifications and the misspecification creates uncertainty about the accuracy of the validity statistics.

The models in Figures 3c and 3d provide additional signs of trouble and potential invalidity. If these five indicators truly measured a single factor, then I would expect that fitting a single factor to a subset of the five indicators would lead to the same factor loadings and other parameter estimates for any parameters shared between the models. Any differences in factor loadings for the same variable should be due to just

¹¹ Other fit indexes could be used, but in my experience these represent a useful variety of measures that perform well.

Figure 3. Five Indicators with Correct (3a) and Three Incorrect Specifications (3b–3d)

Table 2. Maximum Likelihood Estimates of Models in Figure 3 (N = 1000)

	Figure 3a		Figure 3b		Figure 3c		Figure 3d	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Lambda								
y1	1.000	—	1.000	—	1.000	—	1.000	—
y2	0.762	0.065	0.795	0.065			0.745	0.066
y3 (from η_1)	0.628	0.093	1.197	0.083	1.418	0.145	0.897	0.080
y3 (from η_2)	0.408	0.076						
y4	0.752	0.062	0.860	0.076				
y5	1.000	—	1.110	0.090	1.040	0.091		
R-square								
y1	0.464		0.338		0.300		0.474	
y2	0.331		0.263				0.324	
y3	0.479		0.520		0.648		0.410	
y4	0.369		0.274					
y5	0.506		0.353		0.275			
Chi-Square	1.142		70.801		0.000		0.000	
DF	3		5		0		0	
p-value	0.767		0.000					
RNI	1.002		0.928					
IFI	1.002		0.928					
TLI	1.007		0.856					
BIC	-19.581		36.262					

sampling fluctuations. In Figure 3c, I keep the assumption of a single factor, but only use y_1 , y_3 , and y_5 as indicators. Figure 3d also is a single factor, but I only use y_1 , y_2 , and y_3 . Contrasting the estimates from the models that correspond to Figure 3c and 3d reveals a large difference in the factor loading estimates for y_3 (1.418 versus 0.897). Furthermore the R-squares for y_1 go from 0.30 to 0.47 and from 0.65 to 0.41 for y_3 . Finding differences this large raises questions about the model specification and cast doubt on any estimates of validity that use these model estimates as a basis. This illustrates that using different indicators in the model can provide evidence of an invalid structure and should deter the analyst from using the estimates to make validity assessments.

To minimize errors in indicator selection when working with effect indicators, a researcher should have a model with good overall fit and one in which the coefficients do not exhibit large changes when effect indicators of the same latent variable are added or subtracted from the model. Our misspecified models in Figures 3b to 3d did not exhibit these characteristics and the result was that at least some of the validity estimates were seriously misleading.

Five Causal Indicators, Two Latent Variables

In this subsection, I have a model with five causal indicators that influence two latent variables according to the following equations:

$$\begin{aligned}\eta_1 &= \alpha_{\eta_1} + \gamma_{\eta_1 x_1} + \gamma_{\eta_1 x_2} + \gamma_{\eta_1 x_3} + \zeta_1 \\ \eta_2 &= \alpha_{\eta_2} + \gamma_{\eta_2 x_3} + \gamma_{\eta_2 x_4} + \gamma_{\eta_2 x_5} + \zeta_2\end{aligned}\quad (6)$$

where the errors (ζ 's) have a mean of zero and are uncorrelated with the causal indicators (x 's) and with each other. Note that x_1 is the only causal indicator to influence both η 's. The model is set up to parallel the situation of effect indicators, which are linked to two different latent variables. With no effect indicators both equations are underidentified. Suppose that I have one effect indicator for each latent variable

$$\begin{aligned}y_1 &= \eta_1 + \varepsilon_1 \\ y_2 &= \eta_2 + \varepsilon_2\end{aligned}\quad (7)$$

where the unique factors (ε 's) have means of zero and are uncorrelated with the latent variables (η 's), but correlated with each other. Figure 4a is a path diagram of this model. Not knowing the true model, in practice researchers are likely to formulate models with at least some errors. Figure 4b is an example where a researcher incorrectly assumes a single latent variable (η) for the five causal indicators and that the latent variable has two effect indicators, y_1 and y_2 . Figure 4c

is a model where once again a single latent variable is incorrectly used, but only one effect indicator (y_1) is available. Figure 4d is the same setup except that only y_2 is available as an effect indicator.

Simulating data for the model in Figure 4a ($\lambda_{11} = 1$, $\lambda_{22} = 1$, $\gamma_{11} = .5$, $\gamma_{12} = .7$, $\gamma_{13} = .2$, $\gamma_{23} = .3$, $\gamma_{25} = .7$, $\alpha_j = 0$, $V(\varepsilon_j) = 1$, $COV(\varepsilon_1, \varepsilon_2) = .6$, $V(\zeta_k) = 1$, $N = 1000$), I estimated each of the models in Figures 4a to 4d using the ML estimator. The results are reported in Table 3. The model in Figure 4a is not identified. More specifically, we cannot separately identify the error variances for the η 's and the error variances of the ε 's. Had I two effect indicators each for η_1 and η_2 , then this would not be necessary since the original model would be identified.¹² But to illustrate how to proceed in the situation of a single effect indicator for each latent variable, I do not assume additional effect indicators. In Figure 4a the coefficients of the causal indicators are identified and I can estimate them by using the partially reduced form model as described in Bollen and Davis (2009a).¹³ A similar situation holds for the models in Figures 4c and 4d and I also estimate these using the partially reduced form versions of these models.

The column of Table 3 that corresponds to Figure 4a shows that the estimates of the identified parameters are all within ± 2 times the standard errors of the true parameters. In addition, the likelihood ratio chi-square and the fit indices all indicate an excellent fitting model. Since the γ 's are identified and estimated, I have the unstandardized validity coefficients to gauge the validity of each causal indicator as a measure of either η_1 or η_2 . The low to moderate association among the causal indicators does not create collinearity problems that prevent reasonable estimates of validity. The statistical significance and magnitude of coefficients reveal that x_1 to x_2 are valid measures of η_1 . Similarly, x_3 to x_4 are valid measures of η_2 . However, the lack of identification of the variances of ζ 's (error variances of the η 's) means that I cannot estimate the standardized validity coefficients for the causal indicators. (I could estimate the incremental validity variance by comparing the explained variance with and without each causal indicator.) Fortunately, the results for the coefficients are sufficient to highlight the validity of the causal indicators. Furthermore, if I modified the model to allow all five causal indicators to influence both η_1 and η_2 , then I also could establish that x_4 and x_5 are invalid measures of η_1 and that x_1 and x_2 are invalid measures of η_2 .

Figure 4a is the true model and it is reassuring to have accurate estimates of the validity of the causal indicators. But

¹²This assumes that the errors of the new indicators are uncorrelated.

¹³This strategy leads the variances of the errors in the y 's to be the sum of the variances of the equation and indicator disturbances (e.g., $VAR(\varepsilon_1 + \zeta_1)$).

Figure 4. Causal Indicator Models with Correct (4a) and Three Incorrect Specifications (4b–4d)

Table 3. Maximum Likelihood Estimates of Models in Figure 4 (N = 1000)

	Figure 4a		Figure 4b		Figure 4c		Figure 4d	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
gamma								
x1	0.536	0.053	0.148	0.049	0.493	0.060	-0.069	0.060
x2	0.661	0.059	0.288	0.055	0.682	0.061	0.066	0.061
x3 (from η ₁)	0.198	0.056	0.191	0.042	0.180	0.059		
x3 (from η ₂)	0.261	0.050					0.243	0.059
x4	0.727	0.051	0.372	0.038	0.009	0.053	0.721	0.053
x5	0.651	0.049	0.377	0.040	0.066	0.056	0.688	0.056
lamda								
y1	1.000	–	1.000	–	1.000	–		
y2	1.000	–	1.227	0.099			1.000	–
R-square								
η ₁	1.000		0.898		1.000			
η ₂	1.000						1.000	
y1	0.401		0.312		0.0402			
y2	0.444		0.433				0.449	
Chi-Square	3.269		331.022		0.000		0.000	
DF	4		4		0		0	
p-value	0.514		0.000					
RNI	1.000		0.870					
IFI	1.000		0.871					
TLI	1.002		0.316					
BIC	-24.362		303.391					

what happens when I use the wrong model? Figure 4b mistakenly assumes that there is a single latent variable with two effect indicators. The column in Table 3 that corresponds to it shows all causal indicators with statistically significant coefficients. Their nonnegligible magnitude suggests that these measures are valid. Indeed, the R-square for η^* suggests that these causal indicators collectively explain nearly 90 percent of the variance in the latent variable. It is only when I note the large and statistically significant chi-square and other fit indices that doubt is cast on the validity assessment. The model's fit is not good. This raises questions about the validity analysis since structural misspecifications in the model could distort these estimates. A comparison of the results for Figure 4b to those for the true Figure 4a model reveals the problem of incorrectly assuming a single latent variable when two are present.

Figures 4c and 4d represent a case where a researcher might have different single effect indicators with which to estimate a model. Table 3 has columns that represent the results of estimating these models. Interestingly, the estimates that correspond to the model in Figure 4c correspond to estimates of the impact of these causal indicators on η_1 while the estimates that correspond to the model in Figure 4d are the estimates of the impact of these same causal indicators but on η_{12} . This implies that as long as a researcher recognizes that the effect indicators (y_1 and y_2) correspond to *different* latent variables, then it is possible to estimate the unstandardized validity coefficients and determine which causal indicator is valid for which latent variable. I tested the single latent variable model in the column that corresponds to Figure 4b. I found a poor fit as is expected given that the effect indicators actually measure two different latent variables. A good fit would be more likely if they did tap the same latent variable.

Some researchers have been puzzled by estimates of the impact of causal indicators that differ depending on the effect indicator (or other outcome variable) in the analysis. They have interpreted this as suggesting that this is a special problem with causal indicators. However, this is more of a problem with indicators not measuring the same latent variable. An analogous problem was found with effect indicators in the models represented in Figure 3b to 3d. In Figure 3b, I mistakenly assumed that all five effect indicators measure a single latent variable and the factor loadings differ from the true model with two factors. In Figures 3c and 3d, the factor loadings differ depending on which three effect indicators are used, a symptom of the incorrect number of latent variables, not of an inherent problem with effect indicators. So assuming a single latent variable when there are more and having indicators that tap different latent variables can lead to different assessments of validity whether a researcher is evaluating causal *or* effect indicators. For a discussion of this

issue, see Howell et al. (2007), Bagozzi (2007), and Bollen (2007) for contrasting views.

Conclusion

Indicator selection and elimination begins with clear measurement theory that defines a concept, its dimensions, and the necessary latent variables to represent them. The definitions also guide the selection of indicators and specifying whether the measures are causal or effect indicators.¹⁴ Once these relations are specified in SEM, a researcher can begin to assess the validity of indicators. An important step in the empirical analysis is determining whether the measurement model is an adequate description of the data using statistics such as the likelihood ratio chi-square test that compares the model fit to a saturated model. Additional fit indices often supplement the chi-square test. In addition, if the estimates are relatively stable using different effect indicators for the same latent variable, this is supportive of the model specification. We would not expect such stability when using a subset of causal indicators. Embedding this measurement structure in a larger model and checking whether the coefficients change beyond sampling fluctuations is another check on validity. If the results differ, this raises doubt about the specification.

If the fit of the model is adequate, then the researcher can examine the unstandardized and standardized validity coefficients and the unique validity variance measures to gauge the degree of validity. The analyst needs to consider the substantive and statistical significance of these estimates. For causal indicators or effect indicators that depend on two or more latent variables, it would be wise to perform collinearity diagnostics. All of these statistics provide information on the validity of the indicators. Indicators with low validity coefficients and low unique validity variance (in the absence of high collinearity) are ones that are candidates for elimination even if theory initially supported their use. This is true whether they are causal or effect indicators. On the other hand, statistically significant and substantively large unstandardized and standardized validity coefficients with moderate to high unique validity variance support the validity of an indicator.

What makes these assessments most difficult is deciding whether the model as a whole is adequate for the data. In many SEMs with large samples there is considerable statistical power to detect even minor mistakes in the model specification. In practice, there will be ambiguity in assessing

¹⁴The study by Bollen and Ting (2000) has an empirical test to help determine whether a subset or all indicators are causal or effect indicators.

overall model fit. This in turn leaves open the possibility of a structural misspecification that is sufficient to alter the parameter estimates vital to assessing indicator validity. Thus we must recognize and be willing to live with the fact that mistakes are possible where an invalid indicator is retained or a valid one discarded. It is this fact that reinforces the need for a theoretical basis to back up our specifications and empirical tests.

Finally, if composite (formative) indicators are combined to be a convenient summary of a set of variables, then the issue of validity is not present since there is not a claim that these indicators share conceptual unity. With the composite (formative) indicators, there is no expectation of stable coefficients across different dependent variables that the composite might predict. This distinction between causal indicators and composite (formative) indicators is not typically made and when ignored can contribute to controversy.

Acknowledgments

I gratefully acknowledge the support of NSF SES 0617276. My thanks to Shawn Bauldry for research assistance and to the reviewers and editors for helpful comments.

References

- Bagozzi, R. P. 2007. "On the Meaning of Formative Measurement and How it Differs from Reflective Measurement," *Psychological Methods* (12), pp. 229-237.
- Belsley, D. A., Kuh, E., and Welch, R. E. 1980. *Regression Diagnostics*, New York: Wiley.
- Blalock, H. M. 1964. *Causal Inferences in Nonexperimental Research*, Chapel Hill, NC: University of North Carolina Press.
- Bollen, K. A. 1984. "Multiple Indicators: Internal Consistency or No Necessary Relationship," *Quantity and Quality* (18:4), pp. 377-385.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*, New York: Wiley.
- Bollen, K. A. 2007. "Interpretational Confounding Is Due to Misspecification, Not to Type of Indicator: Comment on Howell, Breivik, and Wilcox," *Psychological Methods* (12:2), pp. 219-228.
- Bollen, K. A., and Davis, W. R. 2009a. Causal Indicator Models: Identification, Estimation, and Testing," *Structural Equation Modeling* (16:3), pp. 498-522.
- Bollen, K. A., and Davis, W. R. 2009b. "Two Rules of Identification for Structural Equation Models," *Structural Equation Modeling* (16:3), pp. 523-536.
- Bollen, K. A., and Lennox, R. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2), pp. 305-314.
- Bollen, K. A., and Ting, K. 2000. "A Tetrad Test for Causal Indicators," *Psychological Methods* (5:1), pp. 3-22.
- Diamantopoulos, A., Riefler, P., and Roth, K. P. 2008. "Advancing Formative Measurement Models," *Journal of Business Research* (61:12), pp. 1203-1218.
- Fornell, C., and Bookstein, F. L. 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, pp. 289-324.
- Grace, J. B., and Bollen, K. A. 2008. "Representing General Theoretical Concepts in Structural Equation Models: The Role of Composite Variables," *Environmental and Ecological Statistics* (15), pp. 191-213.
- Heise, D. R. 1972. "Employing Nominal Variables, Induced Variables, and Block Variables in Path Analyses," *Sociological Methods and Research* (1), pp. 147-173.
- Howell R. D., Breivik, E., and Wilcox, J. B. 2007. "Reconsidering Formative Measurement," *Psychological Methods* (12:2), pp. 205-218.
- Jarvis, C., MacKenzie, S., and Podsakoff, P. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research* (30:2), pp. 199-218.
- Jöreskog, K. G., and Goldberger, A. S. 1975. "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association* (70:351), pp. 631-639.
- MacCallum, R. C., and Browne, M. W. 1993. "The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues," *Psychological Bulletin* (114:3), pp. 533-541.
- Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp. 623-656.

About the Author

Kenneth A. Bollen is the H. R. Immerwahr Distinguished Professor of Sociology at the University of North Carolina at Chapel Hill. He is a member of the Statistical Core and a Fellow of the Carolina Population Center and an adjunct professor of Statistics at UNC-Chapel Hill. He was Director of the Odum Institute for Research in Social Science from 2000–2010.

Dr. Bollen is a Fellow of the American Association for the Advancement of Science (AAAS) and Chair of the Social, Economic, and Political Sciences Section K of AAAS. He is a former Fellow at the Center for Advanced Study in the Behavioral Sciences and is the Year 2000 recipient of the Lazarsfeld Award for Methodological Contributions in Sociology. The ISI named him among the World's Most Cited Authors in the Social Sciences. He is coauthor of *Latent Curve Models: A Structural Equations Approach* (with P. Curran, 2006, Wiley) and author of *Structural Equation Models with Latent Variables* (1989, Wiley) and of over 100 published papers. Bollen has been an instructor at the Interuniversity Consortium for Political and Social Research (ICPSR) Summer Training Program in Quantitative Methods at the University of Michigan in Ann Arbor since 1980. His primary areas of statistical research are structural equation models, measurement models, and latent growth curve models.

Copyright of MIS Quarterly is the property of MIS Quarterly & The Society for Information Management and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.