MIS Quarterly

RESEARCH NOTE

THE NEGATIVE CONSEQUENCES OF MEASUREMENT MODEL MISSPECIFICATION: A RESPONSE TO AGUIRRE-URRETA AND MARAKAS

Cheryl Burke Jarvis
College of Business, Southern Illinois University, 1025 Lincoln Drive,
Carbondale, IL 62901 U.S.A. {cbjarvis@siu.edu}

Scott B. MacKenzie and Philip M. Podsakoff
Kelley School of Business, Indiana University, 1309 East Tenth Street,
Bloomington, IN 47405 U.S.A. {mackenz@indiana.edu} {podsakof@indiana.edu}

It has been more than 40 years since Blalock (1964) noted the distinction between what he called "cause" (formative) and "effect" (reflective) indicators of latent variables, and three decades since the academic literature recognized that some SEM measurement models don't fit classical test theory's assumptions about the direction of causality of the relationships between constructs and their indicators (Bagozzi 1981; Fornell and Bookstein 1982). Since that time, interest in modeling constructs with formative indicators has significantly increased in the business, social science, and information systems literatures (e.g., Bollen 2007; Cenfetelli and Bassellier 2009; Diamantopoulos and Papadopoulos 2010; Diamantopoulos et al. 2008; Diamantopoulos and Winklhofer 2001; Edwards and Bagozzi, 2000; Jarvis et. al. 2003; Law and Wong 1999; MacKenzie et al. 2005; Marakas et al. 2007; Petter et al. 2007), and an increasing number of academic studies are incorporating these types of measurement models in their substantive investigations. However, recently researchers in a variety of disciplines (Franke et al. 2008; Howell et al. 2007; Kim et al. 2010) have raised questions surrounding the correct conceptualization and operationalization of formative indicator measurement models.

In this context, the articles by Jarvis et al. (2003) and MacKenzie et al. (2005) were written with several goals in mind. In these articles we attempted not only to illustrate the extent of measurement model misspecification in our literatures and the potential consequences of such misspecification but, more importantly, to provide needed guidance to researchers about how to (1) determine which type of measurement model is conceptually appropriate, (2) develop and purify scales using formative indicators, and (3) specify structural equation models incorporating these composite latent variables.

Others have expanded on this work, including Petter et al. (2007), Cenfetelli and Bassellier (2009), and Kim et al. (2010) in the IS literature specifically. Aguirre-Urreta and Marakas now have continued this discussion with an investigation of just one of the many potential negative consequences of measurement model misspecification—specifically, the potential for bias in structural parameter estimates. They make the argument that, since standardized parameter estimates do not show the same extent of bias as do unstandardized estimates in a misspecified measurement model, the current belief in the negative consequences of misspecification is unwarranted. Although we find there are many areas of agreement between our work and that of Aguirre-Urreta and Marakas, we disagree wholeheartedly with their conclusion. Therefore, we would like to take this opportunity to highlight our points of agreement, as well as offer some alternative insights on the points upon which we disagree. We also will outline some additional implications of misspecification that we believe Aguirre-Urreta and Marakas have overlooked.

First and most importantly, we fully agree with Aguirre-Urreta and Marakas that, regardless of whether or not mea-

[1]Joseph Valacich was the accepting senior editor for this paper.

surement model misspecification biases structural parameter estimates, it is essential that researchers correctly specify their measurement models to match their theoretical conceptualizations. As Aguirre-Urreta and Marakas (p. 134) state,

> Great care should be taken when interpreting our results. Our findings should not be taken to indicate that proper construct specification is not important—much to the contrary. To the extent that the ultimate goal of researchers is to better model the theoretical relationships underlying patterns of manifest variables, then certainly properly specifying causality relationships between components of a theory is central to such a goal. Moreover, and although not generally discussed, properly understanding the direction of causality has major implications for researchers interested in the development of interventions or manipulations of those constructs of interest. The definition of the construct itself, as well as its nomological network, are impacted when defined by one specification or the other. While we do not challenge the conclusion that misspecification is a major issue in the information systems discipline, we believe there is much to be gained from a clear and accurate perspective on the magnitude of its consequences, and why those occur.

We also agree with the authors' last statement in this quote—namely, that an accurate perspective on the magnitude of the consequences of misspecification is critical. However, we disagree that the magnitude of bias in the structural parameter estimates fully captures all of the important consequences of misspecification. In fact, we believe it fails to even address the most important consequences of misspecification. Therefore, we next outline and discuss four critical consequences of measurement model misspecification, including but not limited to the effect of measurement model misspecification on standardized and unstandardized structural parameter estimates.

Consequences for Construct Validity

Before we can consider the impact of measurement model misspecification on a structural model, we must consider its effect on construct validity. A measurement model is a set of theoretical hypotheses about both the conceptual meaning of a latent construct and its relationship to its measures—hypotheses that are just as critical to the validity of a research study as those pertaining to the relationships between the latent constructs in a structural model (Bagozzi 1984; MacKenzie 2003). For any construct, the goal of the item generation process is to develop a set of items that faithfully captures all of the essential aspects of the construct domain. As noted by Diamantopoulos and Siguaw (2006), this is true regardless of whether the measures are viewed as formative or reflective. In addition, in instances where the construct has multiple distinct facets or subdimensions, one also must consider how the subdimensions relate to the higher-order construct. This primarily is a conceptual issue, not an empirical one. It is essential to specify the relationships between the indicators of these subdimensions—and especially the relationships between the subdimensions and the higher-order latent construct—in a manner that is consistent with the conceptual definition. Decisions about the nature of these relationships (i.e., formative or reflective) should be based on a number of important conceptual considerations, which, if ignored, result in a conceptually misspecified measurement model (Bollen and Lennox 1991; Edwards and Bagozzi 2000; Jarvis et al. 2003; Law and Wong 1999; Law et al. 1998, MacKenzie et al. 2005). The use of a conceptually misspecified measurement model will undermine validity not only because it influences estimates of the effects of the focal construct on other constructs, but also because it fundamentally misrepresents the relations between the indicators and the focal construct. This is crucial because, as noted by Bagozzi (1984, p. 14),

> Without consideration of the relationships between theoretical and empirical concepts, it is not possible to assess the meaning of one's terms in a theory. Closely related to this is the problem of construct validity….Construct validity is the extent to which an operationalization measures the theoretical concept which it is intended to measure. It thus depends on the definitional and theoretical meaning of the concept, the behavior of its measurements, and the semantic and syntactic relationships between the two. Our ability to detect true empirical generalizations and interpret them as evidence of underlying nonobservational propositions is dependent fundamentally on construct validity and the relationships between theoretical and empirical concepts. Failure to specify correspondence rules and employ valid measurements of concepts in our theories not only creates a lacuna between theory construction and hypothesis testing but prevents one from addressing the degree of confirmability or falsifiability of theories.

Although they do not emphasize this point in their paper, Aguirre-Urreta and Marakas (p. 124) make a statement that is consistent with this perspective when they write: "Research on the proper specification of constructs is still of critical importance to uncover the true relationships between variables and their structure." Thus, we are in agreement that a proper specification of the conceptual relationships inherent in the measurement model is critical, *whether or not the choice of specification affects structural parameter estimates*.

Measurement model misspecification also can result in the misapplication of scale purification procedures that would have additional serious consequences for the content validity of our scales. As noted by a number of authors (Bollen 1984; Bollen and Lennox 1991; Diamantopoulos et al. 2008; Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003), attempting to force a reflective indicator measurement model on a construct that is more appropriately conceptualized as having formative indicators is likely to lead to the discarding of items with low item-to-total correlations. This is a problem because internal consistency reliability is not an appropriate standard for evaluating formative indicators (Bollen and Lennox 1991). In fact, it is entirely possible (although not necessary) for formative indicators to be completely uncorrelated. In this situation, removing items that don't correlate positively and strongly with other items is likely to result in omitting unique (but critical) parts of the construct domain, thus culminating in a set of measures that is deficient. Clearly, if a latent construct's indicators do not adequately capture the conceptual domain of the construct, there is little value in examining how it relates to other constructs in a structural model.

In addition, the above-identified problems resulting from measurement model misspecification also could affect the publishability of research studies in academic journals, and the amount of attention paid in the literature to certain types of constructs. For example, if reviewers fail to recognize the distinction between formative and reflective measurement models, it is likely that studies using constructs appropriately modeled as having formative indicators may be rejected in the review process because the reviewers may insist on high internal consistency reliabilities and require a reflective indicator measurement model to fit the data. Similarly, if researchers fail to recognize the distinction between formative and reflective measurement models, they may abandon or discard research projects before submitting them for review, simply because they are unable to force the data to fit a reflective indicator measurement model that, in reality, may be inappropriate. The end result may be that some constructs are neglected in the literature because their indicators do not conform to a conceptually inappropriate (i.e., reflective) measurement model, or—perhaps even worse—they may be published with severely restricted construct domains.

Aguirre-Urreta and Marakas (p. 134) seem to understand this potential when they state: "Note we are not claiming here that a researcher can merely change the direction of the relationship between latent and manifest variables and obtain the same results." However, they do not explicitly acknowledge the potential for misspecification to damage construct validity and the extension of that damage forward onto our understanding of critical theoretical concepts in our literatures. In fact, they continue that sentence by saying, "rather, that the bias…due to this change is limited." We would argue that although the bias to some standardized structural parameter estimates might be limited depending on several factors (including the position of the focal construct within the structural model and the level of intercorrelation between indicators), the negative consequences for the valid conceptualization and operationalization of critical theoretical concepts in our literature are unrelated to any potential bias in a structural model's parameter estimates.

Consequences for Model Fit Assessments

Moreover, focusing solely on the bias in structural parameter estimates also ignores the impact of measurement model misspecification on our overall assessment of the goodness of fit of the structural models themselves. Before hypotheses can be tested or path estimates evaluated, the fit of the model should be evaluated. As demonstrated in MacKenzie et al. (2005), the fit statistics of a structural model can be significantly biased by misspecification in the measurement model. We found that if a formative indicator measurement model is misspecified as reflective, the most commonly relied-upon fit statistics—for example, comparative fit index (CFI), goodness-of-fit index (GFI), root-mean-square error of approximation (RMSEA), and standardized root-mean-square residual (SRMR)—can be extremely biased, depending on the location of the misspecified construct within the structural model, the sample size, and the inter-item correlations. Specifically, we found errors of inference as great as 16 percent on common fit statistics in which a correctly specified model was falsely rejected as being inconsistent with the data, and errors of inference in situations in which a misspecified model was falsely judged to fit the data ran as high as 100 percent for the CFI, GFI, and SRMR, and as high as 94 percent for the RMSEA.

These findings have important consequences for our overall evaluation of a structural model. That is, a researcher could be misled by the fit statistics, falsely judging that a misspecified model fits the data well; or vice versa, falsely rejecting a correctly specified model. In the first case, the results of the model and its hypotheses could be used to make predictions about relationships among variables that, in reality, aren't supported by the evidence. In the second case, good data might be thrown away and valuable findings lost.

In either case, these consequences of misspecification of a single construct in a measurement model also can extend into adversely affecting the publishability of a particular piece of research. That is, "bad" models may get published, while "good" ones are rejected or abandoned early in the model testing process. In either case, these errors of inference have implications for the quality of evidence of the models in our literature.

Consequences for Structural Parameter Estimates

We also agree with Aguirre-Urreta and Marakas that misspecification of a measurement model can sometimes result in substantial bias in the model's structural parameter estimates, and this bias appears not only in the unstandardized estimates (as demonstrated in our work), but also in the standardized estimates as reported by Aguirre-Urreta and Marakas. As Aguirre-Urreta and Marakas demonstrate in their results, the misspecification of a formative construct can result in a relative bias in the standardized estimates ranging between -29 percent and -68 percent (for the path linking ξ_1 to η_1 in their Model 2, in which the misspecified construct is placed in the endogenous position). Such bias would be considered substantial by any standards.

In addition, Aguirre-Urreta and Marakas also report relative bias amounts of as much as 5 percent (for the paths from η_1 to η_2 and η_1 to η_4 in Model 2) and as much as 7 percent (for the paths from ξ_1 to η_1 and ξ_1 to η_3 in their Model 1 in which the misspecified formative construct is placed in the exogenous position in the structural model) when the correlations among the formative indicators are low (as will often be the case). These values also are noteworthy, based on widely accepted standards used in other simulations testing for bias in parameter estimates in structural equation models. Indeed, Hoogland and Boomsma (1998) argue that relative bias greater than 5 percent is unacceptable, and many others have adopted this standard in their simulation studies (e.g., Chen et al. 2010; Cheung 2007; Leite 2007).

However, we strongly disagree with Aguirre-Urreta and Marakas' fundamental assertion that the standardized estimates are the "right" values to consider when investigating bias due to misspecification. They support their assertion with the statement (footnote 3, p. 126) that: "Standardized parameters figure prominently in applied research," and cite such evidence as the fact that rules-of-thumb for proxy measures of construct validity recommend standardized estimates greater than 0.70, and that the software package PLS only reports standardized estimates. As a result, they conclude (p. 127) that the "key" to interpreting their results

> …is to remember that relationships in unstandardized form are expressed in a metric that is dependent on the particular estimated variances for the factors involved, whereas those expressed in standardized form have been rescaled to make the variances of all latent and manifest variables equal to one.

However, although standardized estimates are appropriate for comparing the relative influence among a set of predictors *within* a single sample, they are not appropriate for comparing differences in influence *across* different samples. In many respects, our disagreement with them on this issue is a revival of a long-standing debate in the research literature about whether it is better to use unstandardized or standardized estimates for comparisons of effects across samples. Aguirre-Urreta and Marakas' simulation procedure—and that of Jarvis et al. (2003 and MacKenzie et al. (2005)—contrasts the average effect of one construct on another from literally thousands of samples for which the measurement model is correctly specified, to the average effect from thousands of samples for which the measurement model is incorrectly specified. Thus, both analyses are fundamentally based upon comparisons of the estimates of the effect of one construct on another across samples. Many researchers (Alwin 1988; Blalock 1964; Borhnstedt 1969; Duncan 1975; Greenland et al. 1991; Greenland et al. 1986; Kim and Mueller 1976; King 1986; Schumacker and Lomax 2004) have argued that it is misleading to compare standardized estimates across samples because the standardization procedure implicitly equates the variances of the latent constructs across these samples and there is no reason to believe that these variances would be equal. This happens because the standardization process transforms the variances of the constructs to the same value of 1, which implicitly equates theses variances across the samples. In the specific case at hand, this is a particularly serious problem because we *know* that the variances of the constructs are *not* equivalent across samples because the manipulation of the measurement model specification altered the variance of the correctly and incorrectly specified constructs (Law et al. 1998). It is for precisely this reason that so many researchers (such as those cited above) have argued that standardized estimates should not be used for cross-sample comparisons of parameter estimates, and that unstandardized estimates should be used instead. Indeed, this recommendation is consistent with current modeling practice in the SEM literature whenever measurement or structural model estimates are compared across samples (see Byrne et al. 1989; Byrne and Stewart 2006; Cheung and Rensvold 1999; Diamantopoulos and Papadopoulos 2010; Steenkamp and Baumgartner 1998; Vandenberg and Lance 2000), and with recent recommendations by several authors in the meta-analysis literature (see Baguley 2009; Cummings 2004). It is also consistent with the recommendations of Alwin (1988, pp. 19-20, emphasis in original), who noted that,

> The conventional logic governing the choice of metric for variables or the scaling of coefficients in structural equation models is usually given as follows. *In order to compare the effects of variables in a given equation or set of equations in terms of their relative importance, one refers to coefficients for variables scaled in a common metric,* typically variables scaled to have a standard deviation of unity, that is, standardized coefficients….By contrast, when the objective is *to compare the magnitudes of coefficients for a given variable in equations specified in different populations, the general practice is to compare the regression coefficients in their original metric,* rather than to rely on standardized units. Differing magnitudes of coefficients reflect both the units of measurement and the sizes of effects, and because the magnitudes of such standardized coefficients are clearly dependent upon within-population variances, and to the extent these vary significantly across populations/subpopulations, it is generally thought that the unstandardized values are the more stable and therefore the more appropriate.…As the discussion above suggests, it is not simply because these metric coefficients are unaffected by inter-population differences in variation that they are of value for this purpose, as is often supposed. The rationale in fact is different: structural coefficients are often assumed to be invariant reflections of the causal processes generating the distributions and joint distributions of the variables.

In view of the clear consensus in the research literature that it is better to compare unstandardized than standardized structural estimates across samples, we believe that Aguirre-Urreta and Marakas are wrong to assert that standardized estimates should have been used in Jarvis et al. (2003) and MacKenzie et al. (2005). In addition, we find the arguments they provide in support of their position to be unpersuasive for several reasons.

First, it is true that the unstandardized parameter estimates produced by researchers using "real data" will be influenced to some unknown extent by the probability that the specific non-zero values used to set the metric for the latent constructs differ from the true population values for that item loading. However, the estimates produced in our Monte Carlo simulations specifically controlled for this additional source of bias by choosing a value to set the metric in the estimated models that *exactly* matched the true population value for that item loading. Thus, we know that the estimates of bias due to measurement model misspecification that we reported in our 2003 and 2005 articles did not include any additional bias due to the choice of value for setting the metric, and the implications for the effects of measurement model misspecification on the unstandardized parameter estimates are exactly as we described.

Second, there is no evidence that Aguirre-Urreta and Marakas' estimates of bias (which are based on the standardized estimates) are any more accurate than those using unstandardized estimates. Indeed, they are likely to be less accurate because the standardization procedure equates the variances of the latent constructs across the samples even though they are known to differ (Law et al. 1998).

And finally, their argument is unpersuasive because the scaling problem is common to every single latent construct model, regardless of whether the constructs have formative or reflective indicators. Indeed, by their logic, if this problem makes it inappropriate to use unstandardized estimates as the basis for hypothesis tests in our study, then it would also be inappropriate to use unstandardized estimates for hypothesis testing *in every other study*. This, of course, would be inconsistent with the recommendations of most SEM textbooks, software manuals, and experts, including Bentler (1995), Bollen (1989), Jöreskog and Sörbom (1993), MacCallum and Browne (1993), and Schumacker and Lomax (2004), among many others. Consequently, we don't believe Aguirre-Urreta and Marakas' conclusion is warranted.

Consequences for Hypothesis Testing

Aguirre-Urreta and Marakas state in their opening paragraph (p. 123),

> When attempting to empirically test their theoretical propositions, researchers must find ways to make these constructs manifest in order to ascertain whether the expected relationships among them hold as predicted by the proposed research model.

We couldn't agree more. This is the crux of the issue. The most important consequence of measurement model misspecification is not the bias in the structural estimates we report, but rather the erroneous conclusions we draw from our models. Regardless of whether standardized or unstandardized estimates are the "appropriate" choice to report in a statistical analysis, the fact remains that hypotheses about the relationships between constructs generally are tested using the values of the unstandardized estimates and their standard errors. Thus, any bias in either or both of those two values, as demonstrated in our and others' work, would potentially bias the theoretical conclusions we draw from our data.

Aguirre-Urreta and Marakas seem to assume that our original argument was that researchers should only care about measurement model misspecification if the misspecification causes bias in the structural parameters. For example, they state (p. 124) "in several of these articles, Monte Carlo simulations were used to illustrate the deleterious effects of construct misspecification on the estimation of structural parameters of interest." Although it certainly is true that we reported the effects on structural parameters, our focus was not on the biases in the estimates themselves, but rather on the implications of such biases on drawing accurate conclusions from structural models. As we stated in our original work, "as demonstrated by our simulation, the failure to correctly specify the measurement model can lead to different conclusions about the empirical relationships between latent constructs" (Jarvis et al. 2003, p. 216). Our primary goal never was to determine the precise impact of misspecification on the magnitude of a specific path value, or on the relative influence of a set of predictor variables. Rather, our goal always has been to investigate the impact that measurement model misspecification has on the testing of hypotheses about the causal relationships among constructs, and on the evaluation of structural models of those relationships.

Conclusion

Aguirre-Urreta and Marakas appear to take umbrage with the suggestion of our work, and that of Petter et al. (2007), that—because of the substantial proportion of misspecified models in our respective literatures—some of the empirical results in our literatures may be misleading. They argue that this assertion is overstated because the standardized estimates are not as biased as the unstandardized estimates. However, their argument ignores the substantial negative impact of measurement model misspecification on construct development and purification, structural model fit statistics, hypothesis testing, and the resulting choices of which constructs and models appear in the academic literature.

What we most strongly disagree with is the authors' conclusion that their results somehow prove that "the consequences of misspecification seem to be much less dire than previously thought" (p. 124). We believe that the consequences of measurement model misspecification are *exactly* as dire as previously thought. Those consequences simply never had anything to do with the standardized parameter estimates. However, we are happy to note that we all agree that the proper conceptualization and specification of measurement models is, as Aguirre-Urreta and Marakas (p. 137) note, "a worthy goal." The continuing attention paid to the proper application and use of appropriately specified measurement models can serve only to improve the theoretical development and empirical testing of constructs and structural models in many social science and business disciplines.

References

Aguire-Urreta, M. I., and Marakas, G. M. 2011. "Revisiting Bias Due to Construct Misspecification: Different Results from Considering Coefficients in Standardized Form," *MIS Quarterly,* (forthcoming).

Alwin, D. F. 1988. "Measurement and the Interpretation of Effects in Structural Equation Models," in *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research*, J. S. Long (ed.), Newbury Park, CA: Sage, pp. 15-45.

Bagozzi, R. P. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error: A Comment," *Journal of Marketing Research* (18:3), pp. 375-381.

Bagozzi, R. P. 1984. "A Prospectus for Theory Construction in Marketing," *Journal of Marketing* (48:1), 1984, pp. 11-29.

Baguley, T. 2009. "Standardized or Simple Effect Size: What Should be Reported?," *British Journal of Psychology* (100:3), pp. 603-617.

Bentler, P. M. 1995. *EQS Structural Equations Program Manual*, Encino, CA: Multivariate Software Inc.

Blalock Jr., H. M. 1964. *Causal Inferences in Nonexperimental Research*, New York: W. W. Norton.

Bohrnstedt, G. W. 1969. "Observations on the Measurement of Change," in *Sociological Methodology,* E. Borgatta (ed.), San Francisco, CA: Jossey-Bass, pp. 113-133.

Bollen, K. A. 1984. "Multiple Indicators: Internal Consistency or No Necessary Relationship?," *Quality and Quantity* (18:4), pp. 377-385.

Bollen, K. A. 1989. *Structural Equations with Latent Variables*, New York: John Wiley & Sons, Inc.

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., and Lennox, R. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2), pp. 305-314.

Byrne, B. M., Shavelson, R. I., and Muthen, B. 1989. "Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance," *Psychological Bulletin* (105:3), pp. 456- 466.

Byrne, B. M., and Stewart, S. A. 2006. "The MACS Approach to Testing for Multigroup Invariance of a Second-Order Structure: A Walk Through the Process," *Structural Equation Modeling* (13:2), pp. 287-321.

Cenfetelli, R. T., and Bassellier, G. 2009. "Interpretation of Formative Measurement in Information Systems Research," *MIS Quarterly* (33:4), pp. 689-706.

Chen, Q., Kwok, O-M., Luo, W., and Willson, V. L. 2010. "The Impact of Ignoring a Level of Nesting Structure in Multilevel Growth Mixture Models: A Monte Carlo Study," *Structural Equation Modeling* (17:4), pp. 570-589.

Cheung, G. W., and Rensvold, R. B. 1999. "Testing Factorial Invariance Across Groups: A Reconceptualization and Proposed New Method," *Journal of Management* (25:1), pp. 1-27.

Cheung, M. W.-L. 2007. "Comparison of Methods of Handling Missing Time-Invariant Covariates in Latent Growth Models Under the Assumption of Missing Completely at Random," *Organizational Research Methods* (10:4), pp. 609-634.

Cummings, P. 2004. "Meta-Analysis Based on Standardized Effects is Unreliable," *Archives of Pediatrics & Adolescent Medicine* (158:6), pp. 595-597.

Diamantopoulos, A., and Papadopoulos, N. 2010. "Assessing the Cross-National Invariance of Formative Measures: Guidelines for International Business Researchers," *Journal of International Business Studies* (41:2), pp. 360-370.

Diamantopoulos, A., Reifler, P., and Roth, K. P. 2008. "Advancing Formative Measurement Models," *Journal of Business Research* (61:12), pp. 1203-1218.

Diamantopoulos, A., and Siguaw, J. A. 2006. "Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration," *British Journal of Management* (17:4), pp. 263-282.

Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.

Duncan, O. D. 1975. *Introduction to Structural Equation Modeling*, New York: Academic Press.

Edwards, J. R., and Bagozzi, R. P. 2000. "On the Nature and Direction of Relationships Between Constructs and Measures," *Psychological Methods* (5:2), pp. 155-174.

Fornell, C., and Bookstein, F. L. 1982. "A Comparative Analysis of Two Structural Equation Models: LISREL and PLS Applied Consumer Exit-Voice Theory," *Journal of Marketing Research* (19:4), pp. 440-452.

Franke, G. R., Preacher, K. J., and Rigdon, E. E. 2008. "Proportional Structural Effects of Formative Indicators," *Journal of Business Research* (61:12), pp. 1229-1237.

Greenland, S., Maclure, M., Schlesselman, J. J., Poole, C., and Morgenstern, H. 1991. "Standardized Regression Coefficients: A Further Critique and Review of Some Alternatives," *Epidemiology* (2:5), pp. 387-392.

Greenland, S., Schlesselman, J. J., and Criqui, M. H. 1986. "The Fallacy of Employing Standardized Regression Coefficients and Correlations as Measures of Effect," *American Journal of Epidemiology* (123:2), pp. 203-208.

Hoogland, J. J., and Boomsma, A. 1998. "Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis," *Sociological Methods & Research* (26:3), pp. 329-367.

Howell, R. D., Breivik, E., and Wilcox, J. B. 2007. "Reconsidering Formative Measurement," *Psychological Methods* (12:2), pp. 205-218.

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 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 Sörbom, D. 1993. *LISREL 8 User's Reference Guide*, Chicago: Scientific Software International, Inc.

Kim, J., and Mueller, C. W. 1976. "Standardized and Unstandardized Coefficients in Causal Analysis," *Sociological Methods & Research* (4:4), pp. 423-438.

Kim, G., Shin, B., and Grover, V. 2010. "Investigating Two Contradictory Views of Formative Measurement in Information Systems Research," *MIS Quarterly* (34:2), pp. 345-365.

King, G. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science," *American Journal of Political Science* (30:3), pp. 666-687.

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

Law, K. S., Wong, C. S., and Mobley, W. H. 1998. "Toward a Taxonomy of Multidimensional Constructs," *Academy of Management Review* (23:4), pp. 741-755.

Leite, W. L. 2007. "A Comparison of Latent Growth Models for Constructs Measured by Multiple Items," *Structural Equation Modeling* (14:4), pp. 581-610.

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.

MacKenzie, S. B. 2003. "The Dangers of Poor Construct Conceptualization," *Journal of the Academy of Marketing Science* (31:3), pp. 323-326.

MacKenzie, S. B., Podsakoff, P. M., and Jarvis, C. B. 2005. "The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions," *Journal of Applied Psychology* (90:4), pp.710-730.

Marakas, G. M, Johnson, R. D., and Clay, P. F. 2007. "The Evolving Nature of the Computer Self-Efficacy Construct: An Empirical Investigation of Measurement Construction, Validity, Reliability and Stability Over Time," *Journal of the Association for Information Systems* (8:1), pp. 16-46,

Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in IS Research," *MIS Quarterly* (31:4), pp. 623-656.

Schumacker, R. E., and Lomax, R. G. 2004. *Beginner's Guide to Structural Equation Modeling* (2nd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.

Steenkamp, J-B. E. M., and Baumgartner, H. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research* (25:1), pp. 78-90.

Vandenberg, R. J., and Lance, C. E. 2000. "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research," *Organizational Research Methods* (3:1), pp. 4-70.

About the Authors

Cheryl Burke Jarvis (Ph.D., Indiana University) is an associate professor of marketing at Southern Illinois University, and a Research Faculty Fellow of the Center for Services Leadership at Arizona State University. She also has served on the faculties of the Kelley School of Business at Indiana University and the W. P. Carey School of Business at Arizona State University. Her research interests focus on improving our understanding of customer relationships with the firm, including service failure and recovery, and on improving analytical methodologies used in social and behavioral sciences. Her work has been published in top academic journals including *Journal of Marketing*, *Journal of Consumer Research, Journal of Applied Psychology, Journal of the Academy of Marketing Science,* and the *International Journal of Research in Marketing*, among others.

Scott B. MacKenzie (Ph.D., UCLA) is the Neal Gilliatt Chair and Professor of Marketing at the Kelley School of Business at Indiana University. He is the author of nearly 60 articles in leading marketing and management journals on research methodology, advertising effectiveness, leadership, and organizational citizenship behavior; and currently serves on the editorial boards of the *Journal of Marketing, Journal of Marketing Research, Journal of Consumer Research, International Journal of Research in Marketing,* and *Journal of the Academy of Marketing Science.* He is a previous winner of the American Marketing Association's Harold H. Maynard Award, the Society for Industrial and Organizational Psychology's William A. Owens Scholarly Achievement Award, and received Honorable Mention in the Association for Consumer Research's Robert Ferber Award competition.

Philip M. Podsakoff (DBA, Indiana University) is professor of organizational behavior and human resource management, and holder of the John F. Mee Chair of Management in the Kelley School of Business at Indiana University. His current research focuses on the determinants of leadership effectiveness, the antecedents and consequences of organizational citizenship behaviors, relationships between employee attitudes and behaviors, and methodological issues in the social and behavioral sciences. He currently serves on the editorial boards of *Journal of Applied Psychology, Organizational Behavior and Human Decision Processes, Personnel Psychology,* and *Journal of Management*.