MIS Quarterly

RESEARCH COMMENTARY

CONSTRUCT MEASUREMENT AND VALIDATION PROCEDURES IN MIS AND BEHAVIORAL RESEARCH: INTEGRATING NEW AND EXISTING TECHNIQUES

Scott B. MacKenzie
Department of Marketing, Kelley School of Business, Indiana University,
Bloomington, IN 47405 U.S.A. {mackenz@indiana.edu}

Philip M. Podsakoff
Department of Management, Kelley School of Business, Indiana University,
Bloomington, IN 47405 U.S.A. {podsakof@indiana.edu}

Nathan P. Podsakoff
Department of Management and Organizations, Eller College of Management,
University of Arizona, Tucson, AZ 85721 U.S.A. {podsakoff@email.arizona.edu}

Appendix A

Summary of Steps for Scale Purification and Refinement

	Constructs with Reflective Indicators	Constructs with Formative Indicators
Evaluate goodness of fit of the measurement model	Examine significance of the χ^2 (p > .10), CFI ($\geq$ .95), RMSEA ($\leq$ .06), and SRMR ($\leq$ .08). Conduct a simultaneous test of the vanishing tetrads implied by the model (Bollen and Ting 2000).	Examine significance of the χ^2 (p > .10), CFI ($\geq$ .95), RMSEA ($\leq$ .06), and SRMR ($\leq$ .08). For models including 2+ reflective indicators, conduct a simultaneous test of the vanishing tetrads implied by the model (Bollen and Ting 2000).
Assess validity of the set of indicators at the construct level	**For first-order constructs (Figure 3, Panel A):** Examine whether the average variance extracted (AVE) for the set of indicators is greater than .50 (Fornell and Larcker 1981).	**For first-order constructs (Figure 3, Panel B)**: Assess the validity of the set of indicators using Edwards' (2001) adequacy coefficient (R^2_a). Values of R^2_a greater than .50 would mean that, on average, a majority of the variance in the indicators is shared with the construct. Alternatively, if the only antecedents of the composite latent construct are its own formative indicators, the magnitude of the construct level error term could be used to assess validity (Diamantopoulos et al. 2008; Williams et al. 2003). The construct level error variance should be small and constitute no more than half of the total variance of the construct (the smaller the better).
	For second-order constructs (Figure 3, Panel C): Assess the validity of the set of sub-dimensions using Edwards' (2001) multivariate coefficient of determination (R^2_m). Alternatively, the average variance	**For second-order constructs (Figure 3, Panel D)**: Assess the validity of the set of sub-dimensions using Edwards' (2001) adequacy coefficient (R^2_a). Values of R^2_a greater than .50 would mean that, on average, a majority of the variance in the

	Constructs with Reflective Indicators	Constructs with Formative Indicators
	extracted (AVE) could be calculated for the second-order construct by averaging the squared multiple correlations for the first-order indicators. In either case, values greater than .50 would mean that, on average, a majority of the variance in the first-order sub-dimensions is shared with the second-order latent construct.	first-order sub-dimensions is shared with the second-order latent construct. Alternatively, if the only antecedents of the second-order construct are its own first-order sub-dimensions, the magnitude of the second-order construct level error term could be used to assess validity (Diamantopoulos et al. 2008; Williams et al. 2003). An error variance that is less than half of a construct's total variance suggests that a majority of the construct's variance is due to the indicators.
Assess reliability of the set of indicators at the construct level	**For first-order constructs (Figure 3, Panel A)**: Evaluate the reliability of the set of indicators by examining whether (1) the internal consistency reliability (Cronbach 1951) of the set of indicators is greater than .70; and/or (2) Fornell and Larcker's (1981) index of construct reliability is greater than .70 (Netemeyer et al. 2003).	**For first-order constructs (Figure 3, Panel B)**: Traditional notions of internal consistency reliability do not apply to formative indicator measurement models, because the model does not predict that the indicators will be correlated (Bollen 1989; Bollen and Lennox 1991). Consequently, neither Cronbach's alpha nor Fornell and Larcker's index of construct reliability are relevant.
	For second-order constructs (Figure 3, Panel C): The reliability of the first-order sub-dimensions as indicators of the second-order construct can be examined by calculating Fornell and Larcker's (1981) index of construct reliability for the second-order construct. This can be done by using the completely standardized estimates of the second-order factor loadings and error variances associated with the first-order sub-dimensions.	**For second-order constructs (Figure 3, Panel D)**: Traditional notions of internal consistency reliability do not apply to the set of sub-dimensions serving as formative indicators of a second-order construct because the measurement model does not predict that the sub-dimensions will be correlated (Bollen and Lennox 1991; Edwards 2003). Indeed, Edwards (2001, p. 160) argues that "reliability is not an issue of debate when a multidimensional construct and its dimensions are treated as latent variables that contain no measurement error."
Evaluate individual indicator validity and reliability (this assumes that each indicator is associated with only one factor)	**For first-order constructs (Figure 3, Panel A)**: Test indicator validity by examining whether each indicator is significantly related to the latent construct. Assess the degree of validity of each indicator by examining the *unique* proportion of variance in the indicator accounted for by the latent construct. In the typical case where each indicator is hypothesized to load on only one construct, this will be equal to the square of the indicator's completely standardized loading; and its value should be greater than .50 (see Fornell and Larcker 1981). The reliability of each indicator can be assessed by examining the squared multiple correlation for the indicator; typically a value greater than .50 is desired because it suggests that the majority of the variance in the indicator is due to the latent construct. (Of course, in models where each indicator loads on only one construct, the squared multiple correlation and the square of the completely standardized loading will be equal.)	**For first-order constructs (Figure 3, Panel B)**: Test indicator validity by examining whether each indicator is significantly related to the latent construct (Bollen 1989; Bollen and Lennox 1991). Assess the degree of validity of each indicator using the *unique* proportion of variance in the construct accounted for by the indicator. This is calculated by subtracting the proportion of variance in the construct accounted for by all of the indicators *except for the one of interest* from the proportion of variance in the construct accounted for by all of the indicators (see Bollen 1989, pp. 200 and 222). Examine indicator reliability using (1) test-retest reliability (if the indicator is expected to be stable over time), and/or (2) inter-rater reliability (if different raters are expected to agree).
	For second-order constructs (Figure 3, Panel C): Test the validity of each first-order sub-dimension by examining whether it is significantly related to the second-order latent construct. Assess the degree of validity of each sub-dimension by examining the *unique* proportion of variance in the sub-dimension accounted for by the second-order construct. This will be equal to the square of the sub-dimension's completely standardized loading on the second-order construct in the typical case where each sub-dimension is hypothesized to load on only one second-order construct; and its value should be greater than .50 (see Fornell and Larcker 1981). Evaluate the reliability of each sub-dimension by determining	**For second-order constructs (Figure 3, Panel D)**: Test sub-dimension validity by examining whether each sub-dimension is significantly related to the second-order latent construct (Bollen 1989; Bollen and Lennox 1991). Assess the degree of validity of each sub-dimension using the *unique* proportion of variance in the construct accounted for by the sub-dimension. This is calculated by subtracting the proportion of variance in the construct accounted for by all of the sub-dimensions *except for the one of interest* from the proportion of variance in the construct accounted for by all of the sub-dimensions (see Bollen 1989, pp. 200, 222). The reliability of each sub-dimension can be assessed by using Fornell and Larcker's construct reliability index.

	Constructs with Reflective Indicators	Constructs with Formative Indicators
	whether the second-order latent construct accounts for the majority of its variance; this will be shown by a squared multiple correlation for the sub-dimension that is greater than .50 (Fornell and Larcker 1981).	
Eliminate Problematic Indicators	**For first-order constructs (Figure 3, Panel A)**: Provided that all of the essential aspects of the construct domain are captured by the remaining indicators, consider eliminating indicators that have (1) nonsignificant loadings on the hypothesized construct, (2) squared completely standardized loadings that are less than .50, and (3) large and significant measurement error covariances with other measures. Nonsignificant or weak loadings are an indication of a lack of validity, and measurement error covariances may be a sign of multidimensionality (Gerbing and Anderson 1984). Significant measurement error covariances can be identified by looking at the modification indices and their magnitude can be assessed by examining the completely standardized expected change estimates.	**For first-order constructs (Figure 3, Panel B)**: Indicators that have weak or nonsignificant relationships with the latent construct may be candidates for elimination because this may suggest that the indicator lacks validity. However, because this could also be due to multicollinearity, it is important to calculate the VIF to examine multicollinearity among the indicators before deciding whether to eliminate any of them. Indicators with a nonsignificant relationship with the latent construct and a VIF greater than 10 are redundant and should be considered for sequential elimination (Diamantopoulos et al. 2008; Diamantopoulos and Winklhofer 2001; Neter et al. 1996). However, this should only be done if all of the essential aspects of the construct domain are captured by the remaining indicators (Bollen and Lennox 1991; Diamantopoulos and Winklhofer 2001; MacKenzie 2003).
	For second-order constructs (Figure 3, Panel C): The first step in the process is to eliminate problematic indicators of each individual first-order sub-dimension. Provided that all of the essential aspects of the sub-dimension's domain are captured by the remaining items, consider eliminating indicators that have (1) nonsignificant loadings on the hypothesized sub-dimensions, (2) squared completely standardized loadings that are less than .50, (3) large and significant measurement error covariances with other measures (especially measures of other sub-dimensions), and (4) large and significant cross-loadings on nonhypothesized sub-dimensions. Nonsignificant or weak loadings are an indication of a lack of validity, measurement error covariances may be a sign of multidimensionality, and strong and significant cross-loadings are an indication of conceptual confounding. Significant measurement error covariances and cross-loadings can be identified by looking at the modification indices and their magnitude can be assessed by examining the completely standardized expected change estimates. In addition, first-order sub-dimensions that have weak or nonsignificant loadings on the second-order construct may be candidates for elimination because this would suggest that the sub-dimension lacks validity. However, instances where an entire sub-dimension can be dropped without eliminating an essential aspect of the construct domain may be rare.	**For second-order constructs (Figure 3, Panel D)**: The first step in the process is to eliminate problematic indicators of each individual first-order sub-dimension. Provided that all of the essential aspects of the sub-dimension's domain are captured by the remaining items, consider eliminating indicators that have (1) nonsignificant loadings on the hypothesized sub-dimensions, (2) squared completely standardized loadings that are less than .50, (3) large and significant measurement error covariances with other measures (especially measures of other sub-dimensions), and (4) large and significant cross-loadings on non-hypothesized sub-dimensions. Nonsignificant or weak loadings are an indication of a lack of validity, measurement error covariances may be a sign of multidimensionality, and strong and significant cross-loadings are an indication of conceptual confounding. Significant measurement error covariances and cross-loadings can be identified by looking at the modification indices and their magnitude can be assessed by examining the completely standardized expected change estimates. In addition, first-order sub-dimensions that have weak or nonsignificant relationships with the second-order construct may be candidates for elimination because this may suggest that the sub-dimension lacks validity. However, because this could also be due to multicollinearity, it is important to calculate the VIF to examine multicollinearity among the sub-dimensions before deciding whether to eliminate any of them. Sub-dimensions with a nonsignificant relationship with the second-order construct and a VIF greater than 10 are redundant and should be considered for sequential elimination (Diamantopoulos et al. 2008, Diamantopoulos and Winklhofer 2001; Neter et al. 1996). However, this should only be done if all of the essential aspects of the construct domain are captured by the remaining sub-dimensions (Bollen and Lennox 1991; Diamantopoulos and Winklhofer 2001; MacKenzie 2003). Instances where an entire sub-dimension can be dropped without eliminating an essential aspect of the construct domain may be rare.

Appendix B

Summary of Techniques to Assess Scale Validity

	Constructs with Reflective Indicators	Constructs with Formative Indicators
Assess the effect of an experimental manipulation on the construct	Develop an experimental manipulation that theoretically should cause changes in the focal construct. Test whether a dummy variable representing the experimental manipulation is significantly related to scores on the scale.	Develop an experimental manipulation of the attribute/ characteristic that is measured by the formative indicator, and test whether this manipulation influences the scores on the indicator. Note that this implies that when the measurement model is like the one in Panel D of Figure 3, it is the sub-dimensions that serve as the formative indicators of the second-order construct that should be individually manipulated.
Assess known-groups validity of the construct	Identify groups of individuals, organizations, etc. that theoretically should differ on the focal construct. Test whether a dummy variable representing group membership is significantly related to scores on the scale.	Identify groups of individuals, organizations, etc. that theoretically should differ on attribute/characteristic measured by the formative indicator, and test whether a dummy variable representing group membership is significantly related to scores on the measure. Note that this implies that when the measurement model is like the one in Panel D of Figure 3, it is the individual sub-dimensions that serve as the formative indicators of the second-order construct that the groups should differ on. This means that different groups may be used for establishing the known-groups validity of different sub-dimensions.
Assess nomological and/or criterion-related validity of the construct	Test whether the focal construct is significantly related to other constructs hypothesized to be in its nomological network.	Test whether the focal construct is significantly related to other constructs hypothesized to be in its nomological network.
Use the nomological network to assess the validity of the multidimensional structure	**For endogenous constructs [See Figure 5, Panel A]**: Test whether an antecedent construct has direct effects on the sub-dimensions of the focal construct over and above the indirect effects that this antecedent has through the focal construct. This can be tested with a chi-square difference test of the model with and without the direct paths, or by examining the modification indices.	**Endogenous constructs [See Figure 5, Panel B]:** Test whether the R^2 for the effect of the antecedent construct on the focal construct is greater than the R^2_m for the focal construct's sub-dimensions using a confidence interval based on standard errors obtained using bootstrap procedures (Edwards 2001). If the two are equivalent, it can be interpreted as support for the hypothesized multidimensional structure of the focal construct. This test is based on the assumption that a single coefficient (i.e., the coefficient of the effect of the antecedent on the focal construct), must *completely* represent the effect of the antecedent on all sub-dimensions, or the focal construct is concealing potentially useful information. An alternative test of the multidimensional structure (also based on this assumption) is to examine whether the direct effects of the antecedent on each sub-dimension (without the focal construct in the model) are equal (Edwards 2001). If the direct effects are equal, it can be interpreted as support for the hypothesized multidimensional structure of the focal construct. However it is important to note that, although this test is reasonable, it is not necessarily implied by the structure of the model so its appropriateness depends upon whether it makes sense conceptually. One could argue that this requirement is too stringent; particularly if the single coefficient captures the vast majority of the effect of the antecedent on all of the sub-dimensions of the construct.
	For exogenous constructs [See Figure 5, Panel C]: Test whether the sub-dimensions of the multidimensional construct have significant direct effects on the consequence construct over and above the direct effects that the focal construct has on this	**For exogenous constructs [See Figure 5, Panel D]**: Test whether the sub-dimensions of the multidimensional construct have significant direct effects on the consequence construct over and above the indirect effects that these sub-dimensions have through the focal construct. This can be tested with a chi-

	Constructs with Reflective Indicators	Constructs with Formative Indicators
	consequence. This can be tested with a chi-square difference test of the model with and without the direct paths, or by examining the modification indices.	square difference test of the model with and without the direct paths, or by examining the modification indices.
Assess discriminant validity of the construct	Test whether the focal construct is less than perfectly correlated with conceptually similar constructs (Anderson and Gerbing 1988). For a more stringent test, examine whether the average variance explained (AVE) in the indicators by their underlying latent construct is greater than the squared correlation between the focal construct and conceptually similar constructs (Fornell and Larcker 1981).	Test whether the focal construct is less than perfectly correlated with conceptually similar constructs (Anderson and Gerbing 1988). For a more stringent test, examine whether the correlation is less than .71 (which would mean they have less than half of their variance in common).

References

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

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-F. 2000. "A Tetrad Test for Causal Indicators," *Psychological Methods* (5:1), pp. 3-22.

Cronbach, L. J., and Meehl, P. E. 1955. "Construct Validity in Psychological Tests," *Psychological Bulletin* (52:4), pp. 281-302.

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

Edwards, J. R. 2001. "Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework," *Organizational Research Methods* (4:2), pp. 141-192.

Edwards, J. R. 2003. "Construct Validation in Organizational Behavior Research," in *Organizational Behavior: The State of the Science*, J. Greenberg (ed.), Mahwah, NJ: Lawrence Erlbaum Associates, pp. 327-371.

Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), pp. 39-50.

Gerbing, D. W., and Anderson, J. C. 1984. "On the Meaning of Within-Factor Correlated Measurement Errors," *Journal of Consumer Research* (11:1), pp. 572-580.

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

Netemeyer, R. G., Burton, S., and Lichtenstein, D. R. 1995. "Trait Aspects of Vanity: Measurement and Relevance to Consumer Behavior," *Journal of Consumer Research* (21:4), pp. 612-626.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. 1996. *Applied Linear Statistical Models* (4th ed.), Boston: McGraw-Hill.

Williams, L. J., Edwards, J. R., and Vandenberg, R. J. 2003. "Recent Advances in Causal Modeling Methods for Organizational and Management Research," *Journal of Management* (29:6), pp. 903-936.