MIS Quarterly

RESEARCH NOTE

REVISITING BIAS DUE TO CONSTRUCT MISSPECIFICATION: DIFFERENT RESULTS FROM CONSIDERING COEFFICIENTS IN STANDARDIZED FORM

Miguel I. Aguirre-Urreta
School of Accountancy and MIS, College of Commerce, DePaul University,
1 East Jackson Boulevard, Chicago, IL 60604 U.S.A. {maguirr6@depaul.edu}

George M. Marakas
School of Business, University of Kansas, Summerfield Hall,
1300 Sunnyside Avenue, Lawrence, KS 66045 U.S.A. {gmarakas@ku.edu}

Appendix A

Effect of Scale Metric on Standardized Results (Example)

We illustrate the importance of this issue with a simple example. Following the conventions set by Wright (1934), all models depicted in this research show causal influence by unidirectional arrows from cause to effect, and relationships not analyzed in causal terms (correlations, covariances) by curved two-headed arrows. From the population model shown in Figure A1, we generated one dataset with 100,000 observations, and then fit this data to models with identification constraints differing on both approach (reference loading or variance of the exogenous latent variable) and nonzero values (1, 0.50, and 2) yielding a total of 18 possible combinations. The size of the generated dataset was chosen not for its representativeness in applied research, but rather to ensure the estimated parameters would be virtually free from sampling error and thus would remove this source of influence from the point we are trying to make. The population value for the standardized path was 0.630 and the proportion of variance explained in the endogenous variable 39.5 percent. Results from this exercise are shown in Table A1.

Table A1. Results from Alternative Identification Constraints

Identification of Exogenous Variable	Identification of Endogenous Variable								
	Loading = 1			Loading = 0.5			Loading = 2		
	Unstd.	Std.	χ^2	Unstd.	Std.	χ^2	Unstd.	Std.	χ^2
Loading = 1	0.879	0.630	3.605	1.759	0.630	3.605	0.440	0.630	3.605
Loading = 0.5	0.440	0.630	3.605	0.879	0.630	3.605	0.220	0.630	3.605
Loading = 2	1.759	0.630	3.605	3.518	0.630	3.605	0.879	0.630	3.605
Variance = 1	1.491	0.630	3.605	2.983	0.630	3.605	0.746	0.630	3.605
Variance = 0.5	2.109	0.630	3.605	4.218	0.630	3.605	1.055	0.630	3.605
Variance = 2	1.055	0.630	3.605	2.109	0.630	3.605	0.527	0.630	3.605

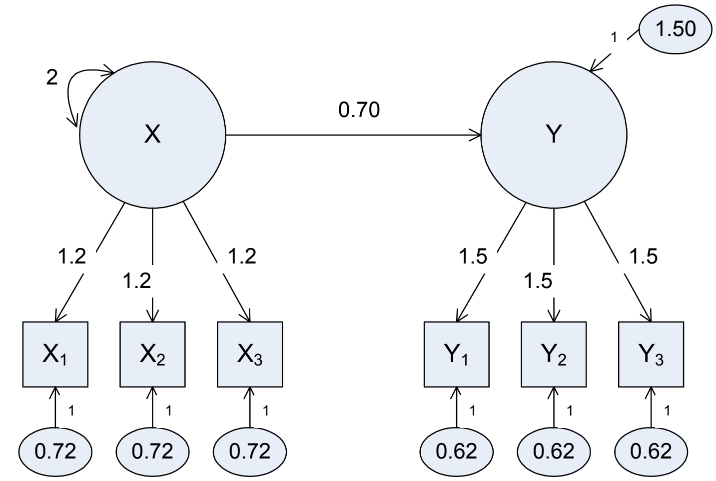

Figure A1. Population Model for Identification Example

Before attempting to interpret these results, it should be noted that all models are correctly specified, as follows: the direction of causality between the latent variables matches the one at the population level, all observed variables load on their respective factors, and the direction of these relationships is also correct. Therefore, all parameter estimates obtained should be unbiased with respect to the corresponding population values, aside from the effects of sampling error. Table A1, however, shows unstandardized estimates that are markedly different from each other, and from known population values. As shown by the χ^2 value, however, model fit is identical across models with different identification constraints, as are the respective standardized path coefficients, from which it follows that the proportion of variance explained in the endogenous latent variable is also the same across all conditions.

The interpretation of these results is quite straightforward, and entirely consistent with our discussion of identification and latent variable metric issues. Whereas unstandardized regression coefficients are a function of the particular approach and particular nonzero values used to set the scale for the latent variables, the nature of the underlying relationship between the two latent variables remains unchanged as a result of these choices. The appearance of bias in the estimate is the result of having used values to set the scales of the latent variables that differ from those at the population level which, incidentally, would be unknown to a researcher attempting to estimate these models. Moreover, even if we had restricted our choice of values to the commonly used unity, we would still have seen apparent bias in the estimate, as evidenced by the result shown in the top-left cell in Table A1 (where the unstandardized regression estimate of 0.879 clearly differs from the population value of 0.700). Other researchers (e.g., Marsh et al. 2004) have recognized the importance of this issue and have rescaled items accordingly to avoid the confounding that would arise from interpreting results expressed in different metrics.

Unstandardized estimates, both regression coefficients as well as covariances between latent variables, are expressed in a metric that is dependent on the particular estimates of the variances for the two latent variables involved in the relationship, and cannot be easily interpreted as to their magnitude absent knowledge about those. In this particular example, the smallest unstandardized estimate was 0.220 (from the combination of a loading set at a value of 2 in the endogenous variable, and at a value of 0.50 in the exogenous one), and the largest 4.218 (from the combination of a loading in the endogenous variable set at 0.50 and the variance of the exogenous latent variable also fixed at 0.50), yet both represent the same proportion of variance explained in the endogenous factor by the exogenous one when the variances of each variable are taken into account. In the first case, the estimated variance of the exogenous variable was 11.503 and that of the endogenous one 1.340, for an R^2 of 39.77 percent [$(11.503 \times 0.220^2) / 1.340$; see Appendix B for a discussion on how to decompose the variances in latent variable models], not different from the population R^2 of 39.5 percent. In the second case, the variance of the exogenous variable was estimated at 0.50 (by virtue of being fixed at that value) and the variance for the endogenous one at 22.380, for an R^2 of 39.75 percent [$(0.50 \times 4.218^2) / 22.380$].

On the other hand, while unstandardized coefficients are a function of the estimated variances of the involved variables, standardized coefficients are obtained by rescaling the research model so that all involved variables have a variance of one. All commonly used statistical packages (e.g., LISREL, MPlus, EQS, etc.) provide these estimates. As shown in Table A1, the standardized regression coefficients were always the same across all conditions, and matched their population level counterparts, as did the proportions of variance explained in the endogenous variable across all conditions.

Appendix B

Decomposition of Variances in the Research Models

The formulas employed in this appendix are

1. The variance of the sum of uncorrelated variables equals the sum of their variances, such that VAR (X + Y) = VAR (X) + VAR (Y).

2. The variance of the sum of correlated variables equals the sum of their variances plus two times their covariance, such that VAR (X + Y) = VAR (X) + 2 COV (X , Y) + VAR (Y).

3. The variance of a random variable multiplied by a constant equals the variance of the random variable times the square of the constant, such that VAR (cX) = c^2 VAR (X).

The main contention of this research is that the scale in which results are expressed has not been carefully considered, and that has led to equivocal conclusions about the different effects discussed here. One example of these issues is the statement by Jarvis et al. (2003) that the models shown in Figure 1 were built so that the item error variances would average 32 percent per factor, consistent with the average amount of random and systematic error found in marketing studies, and thus making these models more representative of that literature. This figure, 32 percent, arises from averaging the residual variances of each of the four items loading on the reflectively specified factors in Figure 1, namely 0.40, 0.24, 0.34, and 0.30 (see, for example, those for η_1 in Models 1A, 1B, and 1C). An average error of 0.32 would represent an average 32 percent of the variance in these indicators only if the average indicator variance amounted to one. Our decomposition of the full and explained variances for all correctly specified models depicted in Figure 1 shows this is not the case. Rather, the average item error variance for η_1 and η_3 in Model 1A above comes to 24 percent and 27 percent respectively and decreases as the formative item intercorrelation increases, reaching 16 percent and 22 percent, respectively, when those are 0.70 (i.e., Model 1C).

This decrease occurs as follows. As the item intercorrelation for the formatively specified construct increases, so does the variance of that construct. This in turn leads to an increase in the variance of those constructs on the receiving end of a path emitting from the formatively specified construct, which in turn results in an increase in the variance for their indicators. Since the residual variance is fixed and does not change from model to model, the proportion this residual item variance represents of the overall total decreases accordingly. Similar results arise for all other correctly specified models, being more markedly different for those in Models 2A, 2B, and 2C. The decomposition shown below also reveals that the proportion of variance explained in the constructs themselves is rather high, ranging from 42 percent to 70 percent in the reflectively specified constructs in Models 1A, 1B, and 1C, and from 17 percent to 71 percent in the reflectively specified constructs in Models 2A, 2B, and 2C above.

We next show how to calculate both the variance and the proportion of variance explained for each variable in all four models. When the formatively specified construct only emits paths to other latent variables (e.g., Models 1A, 1B, and 1C in Figure 1), the equations relating observed with latent variables, and latent variables among themselves are (intercepts omitted for clarity, as they do not affect the calculations below)

$$\xi_1 = \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \zeta_0$$
$$\eta_1 = \beta_1 \xi_1 + \zeta_1$$
$$\eta_3 = \beta_3 \xi_1 + \zeta_3$$
$$\eta_2 = \beta_2 \eta_1 + \zeta_2$$
$$\eta_4 = \beta_4 \eta_1 + \zeta_4$$
$$y_i = \lambda_i \eta_j + \delta_i$$

Therefore, and using the formulas presented above, the following decomposition of the variances for each observed and latent variable can be made, using Model 1A as an example. First, the variance of the formative variable equals (all gamma coefficients were set to one in the population model used in the simulations) $VAR(\xi_1) = VAR(1X_1 + 1X_2 + 1X_3 + 1X_4 + \zeta_0)$. Since the disturbance term ζ_0 is uncorrelated with the exogenous variables $X_{1–4}$, then (by formula 1 above) $VAR(1X_1 + 1X_2 + 1X_3 + 1X_4 + \zeta_0) = VAR(1X_1 + 1X_2 + 1X_3 + 1X_4) + VAR(\zeta_0)$. The X variables, however, are correlated among themselves, thus (by formula 2 above and omitting the ones)

$$VAR(1X_1 + 1X_2 + 1X_3 + 1X_4) + VAR(\zeta_0) = VAR(X_1) + VAR(X_2) + VAR(X_3) + VAR(X_4) + 2\ COV(X_1, X_2) + 2\ COV(X_1, X_3) + 2\ COV(X_1, X_4) + 2\ COV(X_2, X_3) + 2\ COV(X_2, X_4) + 2\ COV(X_3, X_4) + VAR(\zeta_0)$$

Replacing with the population values from Model 1A (e.g., the variances of the X variables equal 1, their correlation equals 0.10, and the variance of the disturbance term, that is, the residual variance, equals 0.50)

$$VAR(\xi_1) = 1 + 1 + 1 + 1 + 6 \times 2 \times 0.10 + 0.50 = 5.70$$

Therefore, the proportion of the variance of ξ_1 explained by X_1, X_2, X_3, and X_4 is 91.23 percent [(5.70 – 0.50) / 5.70]. Next, since η_1 and η_3 are related to ξ_1 by the same regression parameter, their variances will be the same; for η_1 $VAR(\eta_1) = VAR(\beta_1\xi_1 + \zeta_1)$; by formulas 1 and 3 above, the decomposition becomes $VAR(\beta_1\xi_1 + \zeta_1) = \beta_1^2 VAR(\xi_1) + VAR(\zeta_1)$, and replacing with the population parameters (0.30 for the beta coefficient, 5.70 for the variance of the formative variable as calculated above, and 0.50 for the variance of the disturbance term) obtains $VAR(\eta_1) = 0.30^2 \times 5.70 + 0.50 = 1.013$. Therefore, the proportion of the variance of η_1 explained by ξ_1 equals 50.64 percent [(1.013 – 0.50) / 1.013]. The same values apply to η_3.

The variances of all other variables in the model (latent and manifest) can be similarly obtained. Table B1 shows these values, together with the proportions of variance explained in each, for Models 1A, 1B, and 1C in Figure 1 (which differ only in terms of the correlation among the cause indicators of the formatively specified construct).

Table B1. Variances of Variables

Variable	Model 1A (Correlations = 0.10)		Model 1B (Correlations = 0.40)		Model 1C (Correlations = 0.70)	
	Variance	R^2	Variance	R^2	Variance	R^2
ξ_1	5.700	91.23%	9.300	94.62%	12.900	96.12%
η_1, η_3	1.013	50.64%	1.337	62.60%	1.661	69.90%
η_2, η_4	0.865	42.20%	0.981	49.03%	1.098	54.46%
Y_1, Y_9	1.413	71.69%	1.737	76.97%	2.061	80.60%
Y_2, Y_{10}	1.253	80.85%	1.577	84.78%	1.901	87.38%
Y_3, Y_{11}	1.353	74.87%	1.677	79.73%	2.001	83.00%
Y_4, Y_{12}	1.313	77.15%	1.637	81.67%	1.961	84.70%
Y_5, Y_{13}	1.265	68.38%	1.381	71.04%	1.498	73.30%
Y_6, Y_{14}	1.105	78.28%	1.221	80.34%	1.338	82.06%
Y_7, Y_{15}	1.205	71.78%	1.321	74.26%	1.438	76.36%
Y_8, Y_{16}	1.165	74.25%	1.281	76.58%	1.398	78.54%

The proportion of variance explained and, by subtraction, the residual item variance for each indicator loading on a reflectively specified factor can be obtained from Table B1 as follows: For those indicators loading on η_1 and η_3 ($Y_{1–4}$ and $Y_{9–12}$), the average variance explained is 76.14 percent [(71.69 + 80.85 + 74.87 + 77.15) / 4] in Model 1A, 80.79 percent in Model 1B, and 83.92 percent in Model 1C; therefore, the average item error variance for these indicators is 23.86 percent, 19.21 percent, and 16.08 percent for Models 1A, 1B, and 1C respectively. For indicators loading on η_2 and η_4 ($Y_{5–8}$ and $Y_{13–16}$), the corresponding values are 73.17 percent, 75.56 percent, and 77.57 percent for the proportion of variance explained, and 26.83 percent, 24.45 percent, and 22.44 percent for the average item error variance, for Models 1A, 1B, and 1C, respectively.

Using the same formulas and the appropriate equations relating observed to latent variables, as well as latent variables among themselves, Table B2 shows variances and the proportions explained in each, for Models 2A, 2B, and 2C in Figure 1 (differing in terms of the correlation among the cause indicators of the formatively specified construct).

Table B2. Variances and Proportions Explained

Variable	Model 2A (Correlations = 0.10)		Model 2B (Correlations = 0.40)		Model 2C (Correlations = 0.70)	
	Variance	R^2	Variance	R^2	Variance	R^2
ξ_1	0.104	—	0.104	—	0.104	—
η_1	6.636	92.46%	10.236	95.11%	13.836	96.39%
η_3	0.604	17.22%	0.604	17.22%	0.604	17.22%
η_2, η_4	1.097	54.43%	1.421	64.82%	1.745	71.35%
X_1	0.504	20.63%	0.504	20.63%	0.504	20.63%
X_2	0.344	30.23%	0.344	30.23%	0.344	30.23%
X_3	0.444	23.42%	0.444	23.42%	0.444	23.42%
X_4	0.404	25.74%	0.404	25.74%	0.404	25.74%
Y_9	1.004	60.16%	1.004	60.16%	1.004	60.16%
Y_{10}	0.844	71.56%	0.844	71.56%	0.844	71.56%
Y_{11}	0.944	63.98%	0.944	63.98%	0.944	63.98%
Y_{12}	0.904	66.81%	0.904	66.81%	0.904	66.81%
Y_5, Y_{13}	1.497	73.28%	1.821	78.04%	2.145	81.35%
Y_6, Y_{14}	1.337	82.05%	1.661	85.55%	1.985	87.91%
Y_5, Y_{15}	1.437	76.34%	1.761	80.70%	2.085	83.69%
Y_8, Y_{16}	1.397	78.53%	1.721	82.57%	2.045	85.33%

For the manifest variables loading on ξ_1 (X_{1-4}) the average proportion of explained variance in them by this factor was 25.01 percent in all three cases, as the variance of this factor did not change from model to model. This result in the average item error for these indicators is74.99 percent, quite different from the 32 percent stated by Jarvis et al. (2003). For manifest variables loading on η_2 and η_4 (Y_{5-8} and Y_{13-16}), the average variance explained is 77.55 percent, 81.72 percent, and 84.57 percent for Models 2A, 2B, and 2C, yielding an average item error variance of 22.45 percent, 18.29 percent, and 15.43 percent, respectively.

Finally, the items loading on η_3 also display the same explained and residual variance from model to model, since changes in the intercorrelation of the items for the formatively specified constructs had no effect on them. The average variance explained in these indicators amounts to 65.63 percent and the average item residual to 34.37 percent.

Appendix C

Calculation of Standardized Path Coefficients

The calculation of standardized values for all population parameters in the simulated models shown in Figure 1 is quite straightforward and can be performed with any common SEM software package by fixing all parameter estimates at their population values and applying the resulting model to any covariance matrix with the same number of observed variables (since most packages do require some input, even if it is not used in any meaningful way since models contain only fixed values; in our case we used an identity matrix). The following code uses MPlus 3.0 to estimate Model 1A:

```
TITLE: Model 1A Population Values
DATA:
FILE IS C:\I-MATRIX20V.TXT;
TYPE IS COVARIANCE;
NOBSERVATIONS ARE 100000;
VARIABLE:
NAMES ARE x1-x4 y1-y16;
ANALYSIS:
TYPE = GENERAL;
ESTIMATOR = ML;
MODEL:
E1@0.50
E1 BY y1-y4@1;
y1@0.40 y2@0.24 y3@0.34 y4@0.30;
E3@0.50
E3 BY y9-y12@1;
y9@0.40 y10@0.24 y11@0.34 y12@0.30;
K1 BY E1@0.30 E3@0.30;
K1 ON x1-x4@1;
K1@0.50;
x1 WITH x2@0.10 x3@0.10 x4@0.10;
x2 WITH x3@0.10 x4@0.10;
x3 WITH x4@0.10;
x1-x4@1;
E4@0.50
E4 BY y13-y16@1;
y13@0.40 y14@0.24 y15@0.34 y16@0.30;
E2@0.50
E2 BY y5-y8@1;
y5@0.40 y6@0.24 y7@0.34 y8@0.30;
E2 ON E1@0.60;
E4 on E1@0.60;
E4 WITH K1@0;
E2 WITH K1@0;
E2 WITH E4@0;
OUTPUT: RESIDUAL STAND;
```

Running this analysis provides the researcher, among other things, with the population values in standardized metric for each path or parameter of interest. For Model 1A, the relevant standardized values are 0.712 for the K → E1 and K → E3 paths, and 0.649 for the E1 → E2 and E1 → E4 paths. Generally, Table C1 shows the standardized population values for all models, and their variations, investigated in this research. It should be noted that these values for Models 3 and 4 are identical to their counterparts in Models 1 and 2 since the former test for the presence of a nonexistent relationship, which does not alter the population values for these paths.

Table C1. Standardized Population Values

Model	K → E1	K → E3	E1 → E2	E1 → E4
Model 1A (formative indicators correlated at 0.10)	0.712	0.712	0.649	0.649
Model 1B (formative indicators correlated at 0.40)	0.791	0.791	0.700	0.700
Model 1C (formative indicators correlated at 0.70)	0.836	0.839	0.738	0.738
Model 2A (formative indicators correlated at 0.10)	0.376	0.415	0.738	0.738
Model 2B (formative indicators correlated at 0.40)	0.302	0.415	0.805	0.805
Model 2C (formative indicators correlated at 0.70)	0.260	0.415	0.845	0.845

Note: These models and paths correspond to those shown in Figure 1.

It may seem puzzling that, while the structural coefficients in the three variations of both Models 1 and 2 (see Figure 1) do not change, their standardized counterparts do, as shown in Table C1. This is, however, a direct result of the way formatively specified latent variables perform in these models. We work out one example in more detail to show why this is the case.

Consider the relationship between ξ_1 and η_1 in Model 1. The unstandardized coefficient in this relationship is always 0.300 in all variations shown in Figure 1, yet its standardized version, as shown in Table C1, varies from Model 1A to 1B to 1C. We emphasize, however, that the nature of the relationship does not change, as follows: From Appendix B, the variances of ξ_1 and η_1 in Model 1A are 5.700 and 1.013, respectively. Since the structural coefficient linking both latent variables is 0.300 (here, all numbers are expressed in unstandardized form), the variance explained in η_1 by ξ_1 is $5.700 \times 0.300^2 = 0.513$ or $0.513 / 1.013 = 0.5064$ or 50.64 percent. Following the same logic and drawing from the calculations in Appendix B, the proportion of variance explained is 62.60 percent in Model 1B and 69.90 percent in Model 1C.

Recall that when coefficients are standardized they are expressed to show a relationship when the variances of the variables involved are one; for example, standardizing a covariance results in a correlation between two variables. In Table C1, the standardized coefficients for this relationship are 0.712, 0.791, and 0.836 for Models 1A, 1B, and 1C, respectively. Since standardized coefficients involve variances equal to one, the proportions of variance explained for each model are $0.712^2 = 0.5069$ or 50.69 percent, $0.791^2 = 0.6257$ or 62.57 percent, and $0.836^2 = 0.6989$ or 69.89 percent. Minor differences with the percentages shown above are due to rounding of the standardized coefficients to three decimal places.

Therefore, although the structural coefficients, in unstandardized form, remain the same across all three variations of each model shown in Figure 1, the variances of the variables involved do not, and thus the same unstandardized coefficient represents a different relationship in each instance. When expressed in standardized form, these coefficients are different in each case so that the relationship can stay the same, as shown above.

Appendix D

Results from Simulations with N = 500

Table D1. Average Standardized and Unstandardized Paths, and Deviations from Population Values (Models 1 and 2, over 1,500 replications)

	Correctly Specified Models				Misspecified Models			
Path	Unstd.	% Dev.	Std.	% Dev.	Unstd.	% Dev.	Std.	% Dev.
Model 1A (formative indicators correlated at 0.10)								
K → E1	0.299	-0.3%	0.712	0.0%	1.762	487.3%	0.763	7.2%
K → E3	0.299	-0.3%	0.712	0.0%	1.754	484.7%	0.759	6.6%
E1 → E2	0.600	0.0%	0.649	0.0%	0.600	0.0%	0.649	0.0%
E1 → E4	0.600	0.0%	0.649	0.0%	0.600	0.0%	0.649	0.0%
Model 1B (formative indicators correlated at 0.40)								
K → E1	0.299	-0.3%	0.792	0.1%	1.456	385.3%	0.827	4.6%
K → E3	0.299	-0.3%	0.792	0.1%	1.454	384.7%	0.827	4.3%
E1 → E2	0.600	0.0%	0.700	0.0%	0.600	0.0%	0.700	0.0%
E1 → E4	0.600	0.0%	0.700	0.0%	0.600	0.0%	0.700	0.0%
Model 1C (formative indicators correlated at 0.70)								
K → E1	0.298	-0.7%	0.836	0.0%	1.310	336.7%	0.852	1.9%
K → E3	0.298	-0.7%	0.836	0.0%	1.310	336.7%	0.852	1.9%
E1 → E2	0.600	0.0%	0.738	0.3%	0.600	0.0%	0.738	0.3%
E1 → E4	0.600	0.0%	0.738	0.3%	0.600	0.0%	0.738	0.3%
Model 2A (formative indicators correlated at 0.10)								
K → E1	3.071	2.4%	0.377	0.3%	0.350	-88.3%	0.265	-29.5%
K → E3	1.012	1.2%	0.416	0.2%	1.013	1.3%	0.416	0.2%
E1 → E2	0.299	-0.3%	0.737	-0.1%	1.950	550.0%	0.776	5.1%
E1 → E4	0.298	-0.7%	0.737	-0.1%	1.949	549.7%	0.777	5.3%
Model 2B (formative indicators correlated at 0.40)								
K → E1	3.089	3.0%	0.303	0.3%	0.331	-89.0%	0.164	-45.7%
K → E3	1.012	1.2%	0.416	0.2%	1.013	1.3%	0.416	0.2%
E1 → E2	0.298	-0.7%	0.804	-0.1%	1.531	410.3%	0.828	2.9%
E1 → E4	0.298	-0.7%	0.805	0.0%	1.528	409.3%	0.828	2.9%
Model 2C (formative indicators correlated at 0.70)								
K → E1	3.135	4.5%	0.261	0.4%	0.209	-93.0%	0.080	-69.2%
K → E3	1.012	1.2%	0.416	0.2%	1.013	1.3%	0.416	0.2%
E1 → E2	0.297	-1.0%	0.844	-0.1%	1.335	345.0%	0.843	-0.2%
E1 → E4	0.297	-1.0%	0.844	-0.1%	1.331	343.7%	0.843	-0.2%

Note: Unstd. = average of unstandardized paths over replications.
Std. = average of standardized paths over replications.
% Dev. = average deviation from population value, calculated as (average of paths – population value) / population value.

Table D2. Average Standardized and Unstandardized Paths, and Deviations from Population Values (Models 3 and 4, over 1,500 replications)

	Correctly Specified Models				Misspecified Models			
Path	Unstd.	% Dev.	Std.	% Dev.	Unstd.	% Dev.	Std.	% Dev.
Model 3A (formative indicators correlated at 0.10)								
K → E1	0.299	-0.3%	0.712	0.0%	1.764	488.0%	0.764	7.3%
K → E3	0.299	-0.3%	0.712	0.0%	1.752	484.0%	0.759	6.6%
E1 → E2	0.600	0.0%	0.649	0.0%	0.600	0.0%	0.649	0.0%
E1 → E4	0.599	-0.2%	0.648	-0.2%	0.599	-0.2%	0.647	-0.3%
K → E4	0.001		0.002		0.005		0.002	
Model 3B (formative indicators correlated at 0.40)								
K → E1	0.299	-0.3%	0.792	0.1%	1.456	385.3%	0.826	4.4%
K → E3	0.298	-0.7%	0.792	0.1%	1.454	384.7%	0.825	4.3%
E1 → E2	0.600	0.0%	0.700	0.0%	0.600	0.0%	0.700	0.0%
E1 → E4	0.599	-0.2%	0.698	-0.3%	0.598	-0.3%	0.698	-0.3%
K → E4	0.001		0.002		0.004		0.003	
Model 3C (formative indicators correlated at 0.70)								
K → E1	0.298	-0.7%	0.836	0.0%	1.310	336.7%	0.852	1.9%
K → E3	0.298	-0.7%	0.836	0.0%	1.310	336.7%	0.852	1.9%
E1 → E2	0.600	0.0%	0.738	0.3%	0.600	0.0%	0.738	0.3%
E1 → E4	0.600	0.0%	0.737	0.1%	0.599	-0.2%	0.737	0.1%
K → E4	0.000		0.001		0.002		0.002	
Model 4A (formative indicators correlated at 0.10)								
K → E1	3.070	2.3%	0.376	0.0%	0.349	-88.4%	0.265	-29.5%
K → E3	1.010	1.0%	0.415	0.0%	1.014	1.4%	0.416	0.2%
E1 → E2	0.299	-0.3%	0.737	-0.1%	1.950	550.0%	0.776	5.1%
E1 → E4	0.298	-0.7%	0.737	-0.1%	1.948	549.3%	0.777	5.3%
E1 → E3	0.000		0.001		0.002		0.001	
Model 4B (formative indicators correlated at 0.40)								
K → E1	3.088	2.9%	0.303	0.3%	0.331	-89.0%	0.164	-45.7%
K → E3	1.011	1.1%	0.415	0.0%	1.014	1.4%	0.416	0.2%
E1 → E2	0.298	-0.7%	0.805	0.0%	1.531	410.3%	0.828	2.9%
E1 → E4	0.298	-0.7%	0.805	0.0%	1.527	409.0%	0.828	2.9%
E1 → E3	0.000		0.001		0.002		0.001	
Model 4C (formative indicators correlated at 0.70)								
K → E1	3.135	4.5%	0.261	0.4%	0.208	-93.1%	0.080	-69.2%
K → E3	1.011	1.1%	0.415	0.0%	1.013	1.3%	0.416	0.2%
E1 → E2	0.297	-1.0%	0.844	-0.1%	1.335	345.0%	0.843	-0.2%
E1 → E4	0.297	-1.0%	0.844	-0.1%	1.331	343.7%	0.843	-0.2%
E1 → E3	0.000		0.001		0.001		0.001	

Note: Unstd. = average of unstandardized paths over replications.
Std. = average of standardized paths over replications.
% Dev. = average deviation from population value, calculated as (average of paths – population value) / population value. Not calculated for the K → E4 path, which is zero at the population value.

Table D3. Statistical Power Results ((Models 1 and 2, over 1,500 replications)				
Model	**K → E1**	**K → E3**	**E1 → E2**	**E1 → E4**
Model 1A (indicators correlated at 0.10)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	100.0%	100.0%	100.0%	100.0%
Model 1B (indicators correlated at 0.40)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	100.0%	100.0%	100.0%	100.0%
Model 1C (indicators correlated at 0.70)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	100.0%	100.0%	100.0%	100.0%
Model 2A (indicators correlated at 0.10)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	97.6%	100.0%	100.0%	100.0%
Model 2B (indicators correlated at 0.40)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	72.1%	100.0%	100.0%	100.0%
Model 2C (indicators correlated at 0.70)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	27.1%	100.0%	100.0%	100.0%

Table D4. Statistical Power and Type I Error Results (Model 3, over 1,500 replications)					
Model	**K → E1**	**K → E3**	**E1 → E2**	**E1 → E4**	**K → E4**
Model 3A (indicators correlated at 0.10)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	4.5%
Misspecified	100.0%	100.0%	100.0%	100.0%	4.6%
Model 3B (indicators correlated at 0.40)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	4.0%
Misspecified	100.0%	100.0%	100.0%	100.0%	4.8%
Model 3C (indicators correlated at 0.70)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	3.7%
Misspecified	100.0%	100.0%	100.0%	100.0%	4.8%

Note: For the K → E4 path, reported values represent Type I error occurrence.

Table D5. Statistical Power and Type I Error Results (Model 4, over 1,500 replications)					
Model	**K → E1**	**K → E3**	**E1 → E2**	**E1 → E4**	**E1 → E3**
Model 4A (indicators correlated at 0.10)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	5.7%
Misspecified	96.9%	100.0%	100.0%	100.0%	6.2%
Model 4B (indicators correlated at 0.40)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	5.5%
Misspecified	70.9%	100.0%	100.0%	100.0%	6.8%
Model 4C (indicators correlated at 0.70)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	4.8%
Misspecified	25.7%	100.0%	100.0%	100.0%	6.1%

Note: For the E1 → E3 path, reported values represent Type I error occurrence.

Appendix E

Results from Simulations with N = 250

Table E1. Average Standardized and Unstandardized Paths, and Deviations from Population Values (Models 1 and 2, over 1,500 replications)

	Correctly Specified Models				Misspecified Models			
Path	Unstd.	% Dev.	Std.	% Dev.	Unstd.	% Dev.	Std.	% Dev.
Model 1A (formative indicators correlated at 0.10)								
K → E1	0.300	0.0%	0.711	-0.1%	1.769	489.7%	0.759	6.6%
K → E3	0.300	0.0%	0.713	0.1%	1.774	491.3%	0.761	6.9%
E1 → E2	0.601	0.2%	0.647	-0.3%	0.601	0.2%	0.647	-0.3%
E1 → E4	0.603	0.5%	0.649	0.0%	0.603	0.5%	0.649	0.0%
Model 1B (formative indicators correlated at 0.40)								
K → E1	0.301	0.3%	0.791	0.0%	1.456	385.3%	0.825	4.3%
K → E3	0.301	0.3%	0.792	0.1%	1.456	385.3%	0.825	4.3%
E1 → E2	0.600	0.0%	0.698	-0.3%	0.600	0.0%	0.698	-0.3%
E1 → E4	0.602	0.3%	0.700	0.0%	0.602	0.3%	0.700	0.0%
Model 1C (formative indicators correlated at 0.70)								
K → E1	0.302	0.7%	0.836	0.0%	1.311	337.0%	0.851	1.8%
K → E3	0.302	0.7%	0.836	0.0%	1.310	336.7%	0.852	1.9%
E1 → E2	0.600	0.0%	0.736	0.0%	0.602	0.3%	0.736	0.0%
E1 → E4	0.602	0.3%	0.738	0.3%	0.600	0.0%	0.738	0.3%
Model 2A (formative indicators correlated at 0.10)								
K → E1	3.187	6.2%	0.377	0.3%	0.361	-88.0%	0.267	-29.0%
K → E3	1.030	3.0%	0.414	-0.2%	1.035	3.5%	0.415	0.0%
E1 → E2	0.300	0.0%	0.740	0.3%	1.978	559.3%	0.777	5.3%
E1 → E4	0.300	0.0%	0.740	0.3%	1.976	558.7%	0.777	5.3%
Model 2B (formative indicators correlated at 0.40)								
K → E1	3.200	6.7%	0.304	0.7%	0.344	-88.5%	0.165	-45.4%
K → E3	1.030	3.0%	0.414	-0.2%	1.036	3.6%	0.415	0.0%
E1 → E2	0.301	0.3%	0.807	0.2%	1.538	412.7%	0.828	2.9%
E1 → E4	0.301	0.3%	0.807	0.2%	1.537	412.3%	0.828	2.9%
Model 2C (formative indicators correlated at 0.70)								
K → E1	3.274	9.1%	0.262	0.8%	0.220	-92.7%	0.082	-68.5%
K → E3	1.031	3.1%	0.414	-0.2%	1.036	3.6%	0.415	0.0%
E1 → E2	0.302	0.7%	0.846	0.1%	1.338	346.0%	0.843	-0.2%
E1 → E4	0.302	0.7%	0.846	0.1%	1.337	345.7%	0.843	-0.2%

Note: Unstd. = average of unstandardized paths over replications.
Std. = average of standardized paths over replications.
% Dev. = average deviation from population value, calculated as (average of paths – population value) / population value.

Table E2. Average Standardized and Unstandardized Paths, and Deviations from Population Values (Models 3 and 4, over 1,500 replications)

Path	Correctly Specified Models				Misspecified Models			
	Unstd.	% Dev.	Std.	% Dev.	Unstd.	% Dev.	Std.	% Dev.
Model 3A (formative indicators correlated at 0.10)								
K → E1	0.300	0.0%	0.711	-0.1%	1.771	490.3%	0.760	6.7%
K → E3	0.300	0.0%	0.713	0.1%	1.772	490.7%	0.760	6.7%
E1 → E2	0.601	0.2%	0.647	-0.3%	0.601	0.2%	0.647	-0.3%
E1 → E4	0.606	1.0%	0.653	0.6%	0.607	1.2%	0.654	0.8%
K → E4	-0.002		-0.005		-0.010		-0.006	
Model 3B (formative indicators correlated at 0.40)								
K → E1	0.301	0.3%	0.791	0.0%	1.456	385.3%	0.825	4.3%
K → E3	0.301	0.3%	0.792	0.1%	1.456	385.3%	0.825	4.3%
E1 → E2	0.600	0.0%	0.698	-0.3%	0.600	0.0%	0.698	-0.3%
E1 → E4	0.605	0.8%	0.704	0.6%	0.606	1.0%	0.705	0.7%
K → E4	-0.002		-0.004		-0.007		-0.005	
Model 3C (formative indicators correlated at 0.70)								
K → E1	0.302	0.7%	0.836	0.0%	1.311	337.0%	0.851	1.8%
K → E3	0.302	0.7%	0.836	0.0%	1.310	336.7%	0.852	1.9%
E1 → E2	0.600	0.0%	0.736	0.0%	0.600	0.0%	0.736	0.0%
E1 → E4	0.606	1.0%	0.742	0.8%	0.606	1.0%	0.743	1.0%
K → E4	-0.002		-0.005		-0.006		-0.005	
Model 4A (formative indicators correlated at 0.10)								
K → E1	3.185	6.2%	0.377	0.3%	0.362	-87.9%	0.268	-28.7%
K → E3	1.030	3.0%	0.414	-0.2%	1.043	4.3%	0.418	0.7%
E1 → E2	0.300	0.0%	0.740	0.3%	1.978	559.3%	0.777	5.3%
E1 → E4	0.300	0.0%	0.740	0.3%	1.976	558.7%	0.777	5.3%
E1 → E3	-0.001		-0.002		-0.009		-0.005	
Model 4B (formative indicators correlated at 0.40)								
K → E1	3.193	6.4%	0.304	0.7%	0.345	-88.5%	0.166	-45.0%
K → E3	1.030	3.0%	0.415	0.0%	1.040	4.0%	0.417	0.5%
E1 → E2	0.301	0.3%	0.807	0.2%	1.538	412.7%	0.828	2.9%
E1 → E4	0.301	0.3%	0.807	0.2%	1.537	412.3%	0.828	2.9%
E1 → E3	-0.001		-0.003		-0.005		-0.004	
Model 4C (formative indicators correlated at 0.70)								
K → E1	3.270	9.0%	0.262	0.8%	0.222	-92.6%	0.083	-68.1%
K → E3	1.031	3.1%	0.415	0.0%	1.039	3.9%	0.417	0.5%
E1 → E2	0.302	0.7%	0.846	0.1%	1.338	346.0%	0.843	-0.2%
E1 → E4	0.302	0.7%	0.846	0.1%	1.337	345.7%	0.843	-0.2%
E1 → E3	-0.001		-0.003		-0.003		-0.003	

Note: Unstd. = average of unstandardized paths over replications.
Std. = average of standardized paths over replications.
% Dev. = average deviation from population value, calculated as (average of paths – population value) / population value. Not calculated for the K → E4 path, which is zero at the population value.

Table E3. Statistical Power Results (Models 1 and 2, over 1,500 replications)				
Model	**K → E1**	**K → E3**	**E1 → E2**	**E1 → E4**
Model 1A (indicators correlated at 0.10)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	100.0%	100.0%	100.0%	100.0%
Model 1B (indicators correlated at 0.40)				
Correctly specified	100.0%	100.0%	100.0%	100.0%
Misspecified	100.0%	100.0%	100.0%	100.0%
Model 1C (indicators correlated at 0.70)				
Correctly specified	99.9%	99.9%	100.0%	100.0%
Misspecified	100.0%	100.0%	100.0%	100.0%
Model 2A (indicators correlated at 0.10)				
Correctly specified	100.0%	99.9%	100.0%	100.0%
Misspecified	75.4%	99.7%	100.0%	100.0%
Model 2B (indicators correlated at 0.40)				
Correctly specified	99.9%	99.9%	100.0%	100.0%
Misspecified	44.1%	99.7%	100.0%	100.0%
Model 2C (indicators correlated at 0.70)				
Correctly specified	98.9%	99.8%	99.5%	99.5%
Misspecified	14.9%	99.7%	100.0%	100.0%

Table E4. Statistical Power and Type I Error Results (Model 3, over 1,500 replications)					
Model	**K → E1**	**K → E3**	**E1 → E2**	**E1 → E4**	**K → E4**
Model 3A (indicators correlated at 0.10)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	4.1%
Misspecified	99.1%	100.0%	100.0%	100.0%	3.9%
Model 3B (indicators correlated at 0.40)					
Correctly specified	100.0%	100.0%	100.0%	100.0%	3.7%
Misspecified	100.0%	100.0%	100.0%	100.0%	4.3%
Model 3C (indicators correlated at 0.70)					
Correctly specified	100.0%	100.0%	99.9%	99.9%	3.0%
Misspecified	100.0%	100.0%	100.0%	100.0%	4.2%

Note: For the K → E4 path, reported values represent Type I error occurrence.

Table E5. Statistical Power and Type I Error Results (Model 4, over 1,500 replications)					
Model	**K → E1**	**K → E3**	**E1 → E2**	**E1 → E4**	**E1 → E3**
Model 4A (indicators correlated at 0.10)					
Correctly specified	100.0%	99.6%	100.0%	100.0%	4.1%
Misspecified	73.4%	99.5%	100.0%	100.0%	3.7%
Model 4B (indicators correlated at 0.40)					
Correctly specified	99.9%	99.8%	100.0%	100.0%	4.0%
Misspecified	41.7%	99.7%	100.0%	100.0%	4.4%
Model 4C (indicators correlated at 0.70)					
Correctly specified	98.8%	99.7%	99.6%	99.6%	3.1%
Misspecified	14.3%	99.7%	100.0%	100.0%	4.9%

Note: For the E1 → E3 path, reported values represent Type I error occurrence.

References

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), pp. 199-218.

Marsh, H., Wen, Z., and Hau, K. 2004. "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction," *Psychological Methods* (9:3), pp. 275-3000.

Wright, S. 1934. "The Method of Path Coefficients," *Annals of Mathematical Statistics* (5:3), pp. 161-215.