

Not Merely a Matter of Drawing Arrows: The Empirical Consequences of Measurement Model Specification and Recommendations for Practice

Cornell Hospitality Quarterly
2017, Vol. 58(3) 272–292
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1938965516686115
journals.sagepub.com/home/cqx


Elisa K. Chan¹, Michael Sturman², Sanghee Park³,
and Chelsea Vanderpool⁴

Abstract

Understanding measurement model specification is especially important for hospitality research due to its cross-disciplinary nature and the prevalence of measures used in the field which are often central to the formative versus reflective debate (e.g., SERVQUAL, socioeconomic status). The current study contributes to this topic by providing empirically based prescriptive advice to drive better measurement model specification. Specifically, the decision-making procedures developed by this study can complement theoretical reasons for a model choice as well as help determine a correct model choice when theories are equivocal or non-existent. This study combines actual and simulated data to show that model fit statistics alone cannot determine which model specification is correct, but also that a correct measurement model will generate more accurate predictions within a model which in turn will offer more accurate managerial recommendations.

Keywords

marketing constructs, formative measures, reflective measures, simulations, structural equation modeling

To know how to measure a construct accurately, one must understand how the specific measures relate to the focal construct (Bagozzi, 1984). This is the fundamental question when considering whether a construct is best represented as a formative (aka causal) or reflective (aka effect) measure. How the process of measuring a construct is conceptualized is not merely a matter of how arrows should be drawn between a construct and its measures in a figure—it has empirical consequences in resultant research (Bagozzi, 2007; Howell, Breivik, & Wilcox, 2007; Jarvis, MacKenzie, & Podsakoff, 2003; Law & Wong, 1999; MacKenzie, Podsakoff, & Jarvis, 2005). While it is recognized that the nature of the measure–construct relationship is important (Bagozzi, 1984), the empirical consequences still need to be addressed more thoroughly (Howell et al., 2007; Jarvis et al., 2003). This study gauges the type and extent of consequences associated with how a measure is specified. More importantly, this study aims to contribute to the literature by providing guidelines as to how to make better decisions regarding the nature of measures. Our main objective is to provide recommendations of how to weigh conceptual and empirical considerations for deciding whether a measure should be formative or reflective.

Reflective and formative measurements differ fundamentally in how they conceptualize the relationship between

observed indicators and the unobserved latent construct: The two types of measurement treat the construct–indicators relationship in an opposite manner. For reflective measures, the indicators are viewed as a manifestation of the underlying construct (i.e., the latent construct causes the measure; Schwab, 1980); for formative measures, indicators are seen as a cause of the underlying variable (MacCallum & Browne, 1993). While the nature of this construct–indicators relationship should be determined by theory, there is plentiful evidence across fields where researchers have expressed different opinions as to whether even some well-established measures of theoretical constructs should have formative or reflective indicators (Alvarez & Asugman, 2009; Diamantopoulos & Winklhofer, 2001; Howell et al., 2007; Jarvis et al., 2003; Murphy & Hofacker, 2009).

¹Ecole hôtelière de Lausanne, HES-SO // University of Applied Sciences Western Switzerland, Switzerland

²Cornell University, Ithaca, NY, USA

³Rutgers University, Piscataway, NJ, USA

⁴Governors State University, University Park, IL, USA

Corresponding Author:

Elisa K. Chan, Assistant Professor, Ecole hôtelière de Lausanne, Route de Cojonex 18, 1000 Lausanne 25, Switzerland.
Email: elisa.chan@ehl.ch

The issue of formative versus reflective measures is a particularly salient issue in hospitality research for two key reasons. First, because hospitality research is focused on a context (i.e., the hospitality industry) rather than a particular discipline (e.g., management, marketing, finance, etc.), research in this area is inherently more cross-disciplinary. Research in hospitality journals thus brings to bear a wide variety of methodological approaches and perspectives. Methodological papers focused on, for example, management research may not necessarily inform hospitality research on a marketing topic. Second, constructs often used in hospitality research seem to be some of the most controversial constructs in the formative versus reflective measurement debate (Alvarez & Asugman, 2009; Murphy & Hofacker, 2009). Measures like SERVQUAL (Parasuraman, Zeithaml, & Berry, 1988) have been the focus of attention as to whether they are best represented as formative or reflective (cf. Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003; Rossiter, 2002), as have the measure of perceived service quality (cf. Prasnjar, Rajkovic, & Zabkar, 2010). Other debated measures relevant to the wide array of hospitality research include corporate reputation (e.g., Helm, 2005), perceived quality of a destination (e.g., Nadeau, Heslop, O'Reilly, & Luk, 2009; Žabkar, Brenčič, & Dmitrović, 2010), socioeconomic status (SES; e.g., Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003), and website performance (e.g., Dickinger & Stangl, 2013).

In view of this, the current study contributes to the hospitality literature by proposing a new method to aid measurement decisions, especially when theory is ambiguous or equivocal. Our main objective is to generate decision rules to help decide whether the measure of a construct should have formative or reflective indicators. To accomplish this goal, we first provide a brief discussion of the two measurement models. We then implement a computer simulation to evaluate each measurement type and develop decision rules to guide measurement model choices. Finally, we illustrate the use of the decision rules with two examples from published academic papers. Ultimately, the combination of actual and simulated data allows us to (a) provide prescriptive advice on how to weigh conceptual and empirical considerations in deciding whether indicators are formative or reflective, and (b) illustrate how these prescriptions can be implemented in actual research. This article includes two studies each deploying multiple methods and data sources. A summary of the methodology can be found in Appendix A.

Literature Overview

The conceptualization of the reflective model is in line with classical test theory (Bollen & Bauldry, 2011; Bollen & Lennox, 1991; Fornell & Bookstein, 1982). That is, a reflective measure is viewed as a manifestation of the

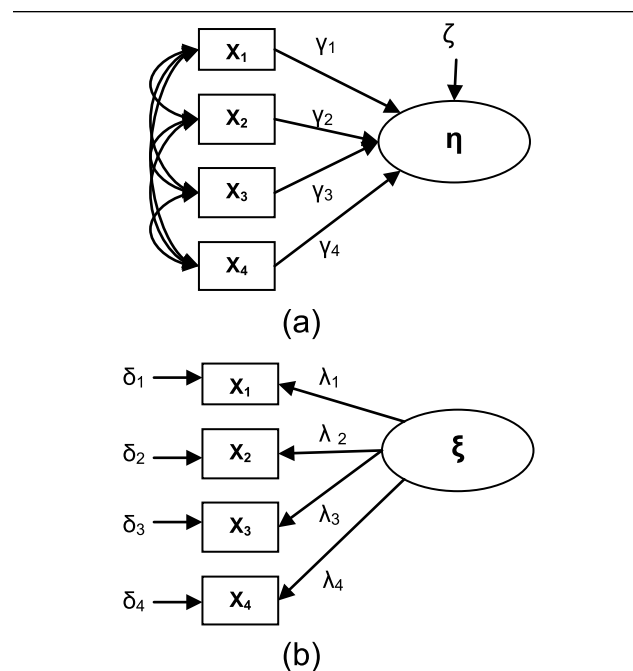


Figure 1. Measurement Models: (a) Formative Measurement Model and (b) Reflective Measurement Model.

underlying construct (Schwab, 1980); or put another way, the latent construct causes the indicators. Alternatively, an observed item can be seen as a cause of the underlying construct, rendering it a formative measure (MacCallum & Browne, 1993).

Formative and reflective measurement models also differ in their assumptions regarding measurement error. Whereas reflective measures assume that each indicator is measured with error, formative measures assume that measurement and other types of error manifest in a disturbance term at the construct level. Moreover, the objectives for the items in reflective and formative measures are different. The purpose of items in reflective measures is to achieve internal consistency by taking into consideration the observed variances and covariances for all measures of a focal construct. Alternatively, the ideal for items in a formative measure is to increase explanatory power by minimizing the amount of measurement error associated with the construct (Fornell & Bookstein, 1982; MacKenzie et al., 2005). Figure 1 illustrates both a formative (Figure 1a) and Reflective (Figure 1b) measure.

Formative Measurement

Figure 1a illustrates a formative measurement model. The construct η is a function of the indicators x_i and their structural coefficients (γ_i). The disturbance term ζ refers to the unexplained variance in η after taking into consideration all

the x_i indicators, which may also be interpreted as measurement error associated with the construct η . Scholars, however, have pointed out that it is inappropriate in any case to assume that measurement errors in the indicators are manifested in the disturbance term (Edwards, 2011, p. 377). The proper interpretation should be that the x_i indicators for a formative measure are assumed to possess no error (Diamantopoulos, 2006). Moreover, the x_i indicators, as depicted in Figure 1a, are allowed to freely correlate. As such, item-intercorrelation is not captured by the construct but is independently estimated by the model. This marks an important difference between a construct with formative versus reflective indicators, the implications of which will become clearer in later discussion. The following is the mathematical expression of a formative measurement model:

$$\eta = \Sigma \gamma_i x_i + \zeta. \quad (1)$$

Ultimately, the accuracy of a formative measure depends on the ability of the x_i indicators to capture the variance of the construct η . Because the construct η is formed by the collection of X s, ideal indicators for the same formative construct should tap into different aspects and should not be highly correlated with one another. If the X s are correlated, this indicates that certain components of the variance of η are being incorporated into the formative measure more than once. Ultimately, the smaller the disturbance term ζ , the better the explanatory power of the specific x_i indicators, and therefore the more accurate the measurement of construct η .

Reflective Measurement

Figure 1b illustrates a reflective measurement model. The construct ξ has a direct influence on each of the y_i indicators. Each indicator is associated with its own measurement error δ_i . In this model, each measure possesses two types of variance: variance explained by the latent construct and variance due to measurement error. Moreover, covariance among measures is attributed to their common latent construct (Edwards & Bagozzi, 2000). The reflective model is expressed by the following equation:

$$y_i = \lambda_i \xi + \delta_i, \quad (2)$$

where y_i is the indicator, λ_i is the factor loading, ξ is the construct, and δ_i is the unique error associated with y_i .

In contrast to a formative measure, the precision of a reflective model not only emphasizes minimizing the unique error associated with the individual indicator y_i , but also maximizing the magnitude of the effects of the latent construct ξ on the y_i s (i.e., λ_i). This idea of developing a measure that maximizes the λ_i terms while minimizing the

δ_i terms is captured in the calculations for reliability (cf. Kline, 2011; Nunnally & Bernstein, 1994). Furthermore, the higher the λ_i s, the higher the intercorrelations among indicators y_i s because the correlation between pairs of reflective measures is a function of the respective λ_i s (Bollen, 1989). Therefore, ideal indicators of the same reflective construct should be highly correlated.

Mathematical Differences and Measurement Model Misspecification

To summarize the key mathematical differences in the measurement models, while a formative measure expresses a construct as a function of its indicators, for a reflective measure, each indicator is a function of its reflective construct. Consequently, for a formative construct, highly correlated indicators will result in multicollinearity which will undermine the accuracy of the estimated relationship between the construct and the group of indicators (i.e., γ_i s), while the opposite is true for a reflective construct. It is also worth noting that the residual of a formative measure manifests the amount of unexplained variance for the construct after all indicators are accounted for. On the contrary, the residual in the reflective measure is essentially the measurement error for a particular indicator.

Past studies have demonstrated that when formative indicators are misspecified as reflective, there is a bias in the estimates of substantive relationships with the construct (Jarvis et al., 2003; Law & Wong, 1999; MacKenzie et al., 2005). For example, when the intercorrelation among items was at .10, the strength of the substantive relationship between a misspecified reflective independent variable and a dependent variable was inflated by an average of 490% (Jarvis et al., 2003). Other studies revealed that when independent variable's indicators are misconstrued as formative, the relationship between the construct and its indicators (i.e., γ_i s) becomes a function of the dependent variable (Bagozzi, 2007; Howell et al., 2007). This creates an interpretational confound for the specific independent variable and results in ambiguity in the meaning of the variable (Burt, 1976). It is clear that formative versus reflective model specification has important theoretical and empirical consequences. That said, there are limitations in this past research.

While past research showed that there are significant consequences to misspecification, there is no clear guidance as to how researchers, if lacking clear and precise theory, can better determine the appropriate measurement model to employ. We know that fit indices alone are inadequate to determine or verify the appropriate model (Jarvis et al., 2003; MacKenzie et al., 2005), but we have no clear guidance on how to make such a determination. Consequently, the current study seeks to provide empirically based prescriptive advice to drive better choices regarding the nature of construct indicators. Specifically, the decision-making

procedures developed by this study can complement theoretical reasons for a model choice as well as help determine a correct model choice when theories are equivocal or non-existent.

For hospitality research, the decision-making procedure we develop will generate better predictions as well as offer more accurate managerial recommendations based on empirical results. Using SERVQUAL as an example, if results of a particular study were to indicate that SERVQUAL affects customer satisfaction, arguably, managerial results would differ according to measurement model specification. If SERVQUAL was specified as a formative measure, the appropriate implication would be that changes in assurance, empathy, reliability, responsiveness, and tangibility cause changes in service quality and in turn lead to higher customer satisfaction. This would imply that companies need to manage or improve the five service aspects to increase customer satisfaction which is a concern for employee and service experience management. If, however, SERVQUAL was specified as a reflective measure, the appropriate implication would be that changes in perception of service quality would influence perceptions of assurance, empathy, reliability, responsiveness, and tangibility which then affect customer satisfaction. This would suggest that companies need to manage or improve customer's perceptions of the service quality to enhance the service aspects and customer satisfaction which is a concern for promotion and branding.

Study 1: Using Simulated Data to Demonstrate the Consequences of Measurement Model Choice

Simulation Procedure and Data (Step 1)

Two groups of data sets were generated resembling the conceptual models depicted in Figure 2. In particular, for both groups, there is an independent variable (i.e., η for formative and ξ for reflective) with four indicators X_1 to X_4 . This independent variable is related to two dependent variables (i.e., Y_1 and Y_2) each with three reflective indicators. The two groups varied in the true construct type of its independent variable. In one group, the formative construct type was true (Data Group 1) and in the other group, the reflective construct type was true (Data Group 2).

Based on the conceptual models in Figure 2, we generated covariance matrices which vary by the following parameters for both groups: (a) the item-intercorrelation of the four indicators of the independent variable (r_{xx} : .2, .7, .9), (b) the relationships between the four indicators and independent variable (i.e., factor weights γ for formative and factor loadings λ for reflective, at .2, .7, .9), (c) the correlation between the two dependent variables (r_{y_1,y_2} : -.6, -.25, 0, .25, .6), and (d) the relationships between the independent

variable and each of the two dependent variables (i.e., β_1 and β_2 : .4, .4; .4, .25; .25, .25; .25, -.25; .4, -.25; .4, -.4).

We chose these specific levels partly because they are consistent with prior simulations of measurement models (Howell et al., 2007; Jarvis et al., 2003; MacKenzie et al., 2005). More importantly, they represent meaningful categories derived from conventional beliefs concerning formative and reflective measures. For example, measures having low covariation are often taken as empirical evidence that they are formative in nature, whereas those having high covariation are thought to be reflective (Podsakoff, Shen, & Podsakoff, 2006). The three levels we tested—namely, .2, .7, and .9—represent situations where items-shared variance is minimal ($.2^2 = 4\%$), moderate ($.7^2 = 49\%$), and large ($.9^2 = 81\%$). Therefore, the three levels actually cover low covariation presumably associated with a formative measure, moderate covariation which could be interpreted either way, and high covariation presumably representing a reflective measure. The same applies to the specific levels for the relationship between the dependent variables: We tested negatively correlated (i.e., -.6 and -.25), uncorrelated (i.e., 0), and positively correlated (i.e., .25 and .6) circumstances. By building the correlation between the dependent variables into our simulation, we can control for its potential effects on important empirical outcomes, thus allowing us to be more confident that any observed conceptual model misfits can only be attributed to the misspecification of the focal independent variable. In sum, the specific levels for each of our parameters allow us to examine the most significant range of each variable without using so many parameters as to make the simulation overly complex. Finally, as in past research (Howell et al., 2007; Jarvis et al., 2003), factor loadings of all reflective indicators for the two dependent variables were fixed at 0.95.

The two groups of simulated data sets combined to a total of 540 simulation conditions: this derives from 2 (true construct type: reflective vs. formative) \times 3 (levels of item-intercorrelation) \times 3 (levels of factor loading/weight) \times 5 (levels of correlation between the two dependent variables) \times 6 (pairs of IV–DV relationship) = 540 simulation conditions. For each condition, 100 covariance matrices were created based on the corresponding combination of parameters. Consequently, the simulation consists $540 \times 100 = 54,000$ cases (i.e., $n = 54,000$). Each covariance matrix is based on a hypothetical sample size of 500, which is sufficiently large to make comparison of parameters at two decimal places (Bedeian, Sturman, & Streiner, 2009). In essence, each case is a covariance matrix within its respective simulation condition which adheres to the conceptual model depicted in Figure 2 with one independent variable and two dependent variables. A summary of the steps followed to generate the data is described in Appendix B.

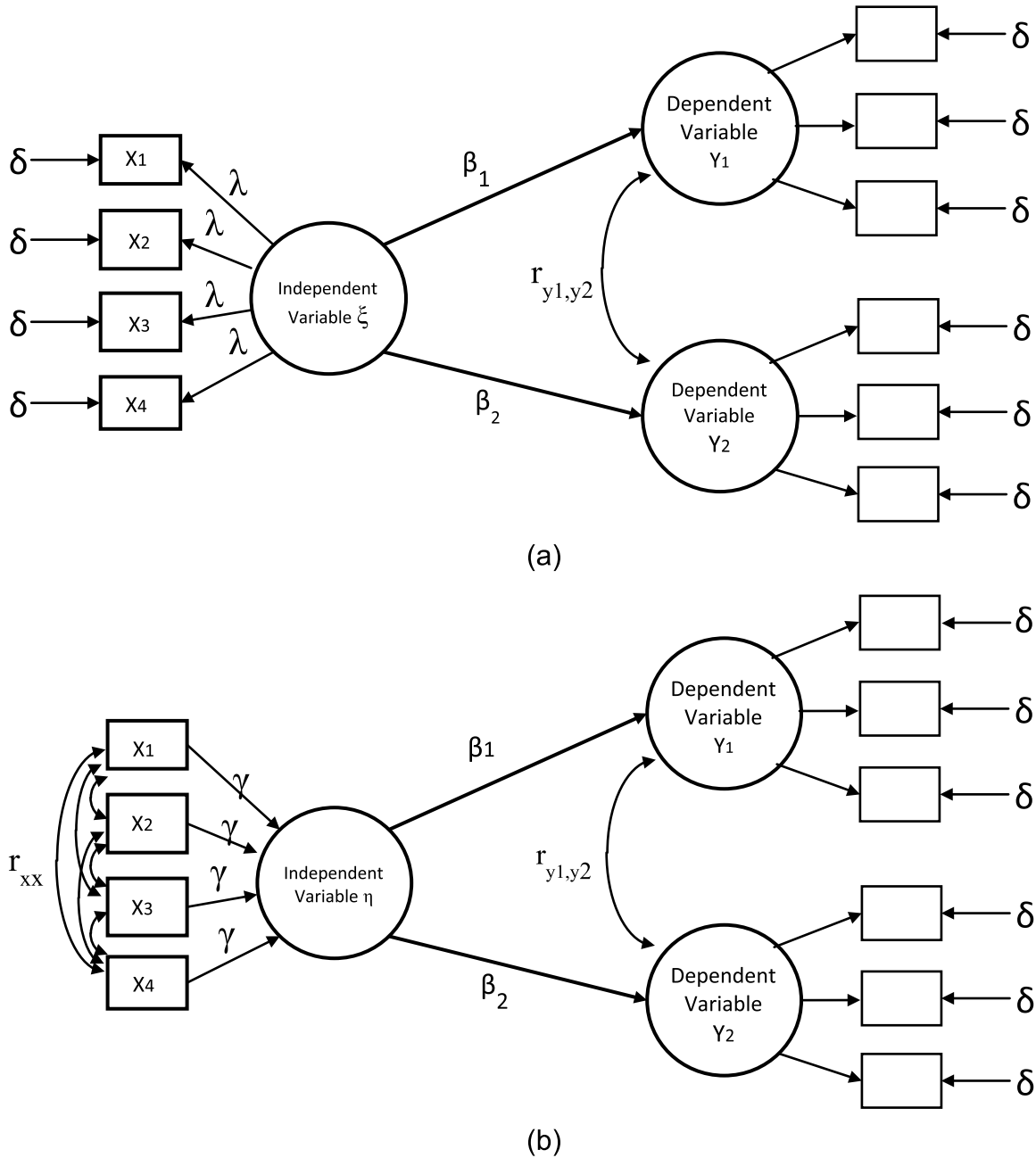


Figure 2. Simulation Conceptual Models: (a) Reflective as True Construct Type Condition and (b) Formative as True Construct Type Condition.

Structural Equation Modeling (SEM) Analyses of Simulated Data (Step 2)

We submitted all 54,000 cases (i.e., covariance matrices) to an SEM analysis with LISREL. Each case is analyzed in two different ways. Set 1 involved taking the 27,000 cases where the true independent variable type was a formative measure, and conducting 27,000 SEM analyses assuming

(correctly) the independent variable was a formative measure, and 27,000 analyses assuming (incorrectly) that the independent variable was a reflective measure. Set 2 involved taking the 27,000 cases where the true independent variable type was a reflective measure, and conducting 27,000 SEM analyses assuming (incorrectly) the independent variable was a formative measure, and 27,000 analyses assuming (correctly) that the independent variable was a

Table 1.
Regression Analyses: Coding the Correct Model Predictor.

Step 1: Simulation ($n = 54,000$ cases)	Step 2: SEM Analyses ($n = 108,000$ cases)	Step 3: Regression Analyses (Coding Correct Model Predictor)
Data Group 1: Set of analyses where the true construct type is a formative measure ($n = 27,000$)	Set 1 (first half): Analyzing Data Group 1 using formative model specification (27,000 cases)	Coded as 1 = correct model used
	Set 1 (second half): Analyzing Data Group 1 using reflective model specification (27,000 cases)	Coded as 0 = wrong model used
Data Group 2: Set of analyses where the true construct type is a reflective measure ($n = 27,000$)	Set 2 (first half): Analyzing Data Group 2 using formative model specification (27,000 cases)	Coded as 0 = wrong model used
	Set 2: (second half): Analyzing Data Group 2 using reflective model specification (27,000 cases)	Coded as 1 = correct model used

Note. SEM = structural equation modeling.

reflective measure. Thus, we conducted a total of 108,000 (i.e., $27,000 \times 4$) SEM analyses. As a result of this procedure, the model is correctly specified in half of the analyses, and incorrectly in the other half (see Table 1).

Regression Analyses of SEM Results (Step 3)

To examine the consequences of measurement model choice, we conducted regression analyses of the SEM results to determine the impacts of key characteristics on key outputs. Two sets of regression analyses were run. The first set of analyses are for the cases where the true construct type was a formative measure. Of the 54,000 potential cases, 52,466 provided results (i.e., 1,554 did not converge to a solution). Summary statistics from these analyses are provided in Table 2. The second set of analyses is for the cases where the true construct type was a reflective measure. Of the 54,000 potential cases, 49,028 provided results (i.e., 4,972 did not converge to a solution). Summary statistics from these analyses are provided in Table 3.

Key outputs (i.e., dependent variables of regression). From each of the SEM analyses, we obtained the following outputs: (a) beta estimates (i.e., the IV–DV relationships); (b) commonly reported fit indices (i.e., chi-square, root mean square error approximation [RMSEA], standardized root mean square residual [SRMR], Normed Fit Index [NFI], Comparative Fit Index [CFI], Incremental Fit Index [IFI], Goodness of Fit Index [GFI], and Adjusted Goodness-of-Fit Index [AGFI]); (c) factor loadings/weights (i.e., factor weights γ for formative measures and factor loadings λ for reflective measures). For examining beta estimates and factor loadings, the dependent variables in the regression analyses were the absolute error in their estimations (i.e., the absolute value of the true values minus the estimated

value). In the analyses of the fit statistics, the fit statistics were regressed as is on the independent variables.

Key characteristics (i.e., independent variables of regression). To assess the consequences of measurement model choice, the following predictor variables were included in the regression analyses: correct model (1 = correct model used vs. 0 = wrong model used; refer to Table 1 for coding method), item-intercorrelation, the relationship between the two dependent variables, correct model \times item-intercorrelation, and correct model \times the relationship between the two dependent variables.

Discussion of Regression Analyses Results

For both sets of analyses (i.e., where the true construct type is formative, and where the true construct type is reflective), having specified the correct model (i.e., simulated true construct type matches the SEM measurement model specification) is negatively correlated with χ^2 , RMSEA, and SRMR. As a result, not surprisingly, a correctly specified measurement model will, on average, produce lower χ^2 , RMSEA, and SRMR. Moreover, the results reveal that a correctly specified measurement model has only small influences on the estimation of substantive relationships (i.e., betas). Accordingly, based on our more diverse set of simulated data, measurement model specification (alone) does not seem to be as consequential as has been postulated in past research. Despite the discrepant results with past studies, however, our regression analyses show that these seemingly insubstantial effects are masked by other factors: item-intercorrelation and the correlation between dependent variables. This indeed provides an explanation for the difference between our findings and prior simulations.

Table 2.
Correlation Matrix (Set 1: True Construct Type Is Formative).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Correct model	1																		
2. Intercorrelation	-.001	1																	
3. r_{y1y2}	.001	.004	1																
4. Correct \times Intercorrelation	.826	.392	.005	1															
5. Correct \times r_{y1y2}	.003	.007	.697	.009	1														
6. Errors in Beta 1	.024	.010	-.202	.125	-.097	1													
7. Errors in Beta 2	.034	.092	.328	.091	.225	.219	1												
8. χ^2	-.105	-.043	-.675	-.130	-.510	.116	-.259	1											
9. RMSEA	-.012	-.073	-.677	-.046	-.541	.198	-.239	.932	1										
10. SRMR	-.050	-.060	-.676	-.117	-.532	-.093	-.390	.866	.851	1									
11. Normed Fit Index (NFI)	.068	.399	.618	.232	.471	.030	.355	-.850	-.816	-.841	1								
12. Comparative Fit Index (CFI)	.037	.278	.636	.162	.494	-.046	.295	-.905	-.857	-.853	.934	1							
13. Incremental Fit Index (IFI)	.037	.277	.636	.162	.494	-.046	.294	-.905	-.858	-.853	.934	.999	1						
14. Goodness of Fit Index (GFI)	.098	.046	.685	.118	.519	-.180	.251	-.979	-.955	-.861	.836	.887	.887	1					
15. Adjusted Goodness-of-Fit Index (AGFI)	-.025	.050	.691	.021	.563	-.184	.250	-.971	-.961	-.862	.839	.894	.894	.978	1				
16. Errors in λ_1	.220	.104	-.037	.250	-.013	-.078	-.027	-.050	-.022	.011	.032	.029	.029	.050	.026	1			
17. Errors in λ_2	.272	.117	-.051	.303	-.032	.074	-.056	.023	.580	.107	-.021	-.028	-.028	-.037	-.075	.709	1		
18. Errors in λ_3	.271	.116	-.052	.302	-.034	.076	-.058	.024	.590	.108	-.022	-.027	-.027	-.037	-.076	.707	.872	1	
19. Errors in λ_4	.270	.119	-.052	.304	-.035	.077	-.055	.024	.590	.107	-.021	-.027	-.027	-.038	-.076	.706	.865	.868	1

Note. $n = 52,446$ (1,554 cases did not provide a solution). Correlations $< -.001$ or $> .004$ are significant at $p = .05$ level. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

Table 3.
Correlation Matrix (Set 2: True Construct Type Is Reflective).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Correct model	1																		
2. Intercorrelation	.050	1																	
3. r_{y1y2}	.040	-.041	1																
4. Correct \times intercorrelation	.853	.405	-.008	1															
5. Correct \times r_{y1y2}	.050	-.076	.706	-.016	1														
6. Errors in Beta 1	-.065	-.348	-.162	-.254	-.159	1													
7. Errors in Beta 2	.019	-.015	.121	-.060	.129	.086	1												
8. χ^2	-.022	.145	-.557	.119	-.397	-.031	-.063	1											
9. RMSEA	-.168	.177	-.492	.011	-.333	-.002	-.043	.934	1										
10. SRMR	-.061	.125	-.531	.119	-.385	-.094	-.132	.939	.919	1									
11. Normed Fit Index (NFI)	.089	.123	.520	.014	.353	-.009	.089	-.919	-.870	-.893	1								
12. Comparative Fit Index (CFI)	.119	.012	.526	-.001	.357	.009	.081	-.948	-.908	-.923	.953	1							
13. Incremental Fit Index (IFI)	.119	.007	.526	-.003	.358	.010	.081	-.948	-.909	-.923	.952	.998	1						
14. Goodness of Fit Index (GFI)	.048	-.158	.538	-.098	.370	.006	.042	-.982	-.955	-.919	.903	.934	.934	1					
15. Adjusted Goodness-of-Fit Index (AGFI)	.153	-.134	.539	.006	.342	-.017	.035	-.937	-.963	-.909	.908	.943	.934	.981	1				
16. Errors in λ_1	-.114	-.130	-.019	-.192	.009	.152	.007	-.061	-.063	-.047	.038	.047	.048	-.070	.054	1			
17. Errors in λ_2	-.271	-.190	-.006	-.311	-.003	.318	.048	.008	.057	-.010	-.051	-.045	-.044	-.026	-.064	.615	1		
18. Errors in λ_3	-.272	-.187	-.006	-.312	-.003	.316	.049	.009	.058	-.010	-.051	-.045	-.045	-.027	-.065	.616	.824	1	
19. Errors in λ_4	-.270	-.189	-.003	-.311	-.001	.317	.052	.007	.055	-.013	-.049	-.043	-.043	-.025	-.062	.617	.825	.825	1

Note. $n = 49,028$ (4972 cases did not provide a solution). Correlations $< -.008$ or $> .008$ are significant at $p = .05$ level. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

Accuracy of Substantive Relationships Estimations Is Contingent on Item-Intercorrelation

Findings of the current study provide a more complete picture of the influence of formative versus reflective measurement model specification on substantive relationships (i.e., beta estimates) compared with previous research (Table 4). Unlike past research which suggests that a formative measure

always gives more accurate estimations of the substantive relationships than a reflective measure (Jarvis et al., 2003; Law & Wong, 1999; MacKenzie et al., 2005), the current study shows that the amount of error in beta estimations is contingent on item-intercorrelation. The results also show that, in some cases, the amount of error in beta estimation is also contingent on the correlation between the dependent variables (i.e., r_{y1y2}).

Table 4.
Regression Analyses: Results for Absolute Errors in Betas Estimation.

Predictors	True Independent Variables Measure Type Is			
	Formative (Set 1) (n = 52,446)		Reflective (Set 2) (n = 49,028)	
	Errors in Beta 1	Errors in Beta 2	Errors in Beta 1	Errors in Beta 2
Correct model	-0.217 (0.004)***	-0.068 (0.006)***	0.245 (0.006)***	0.229 (0.006)***
Intercorrelation	-0.176 (0.004)***	0.031 (0.007)***	-0.153 (0.006)***	0.163 (0.006)***
r_{y1y2}	-0.140 (0.003)***	0.273 (0.005)***	-0.065 (0.004)***	0.038 (0.004)***
Correct × Intercorrelation	0.379 (0.006)***	0.150 (0.009)***	-0.423 (0.008)***	-0.354 (0.009)***
Correct × r_{y1y2}	0.064 (0.004)***	-0.009 (0.007)	-0.130 (0.006)***	0.072 (0.006)***

Note. The regression coefficients reported are unstandardized. Standardized errors are in parentheses.

* $p < .10$. ** $p < .05$. *** $p < .001$.

Specifically, the regression results reveal that the reduction in error associated with specifying the correct model is mitigated by the interaction effects of the correct model × item-intercorrelation. In particular, the interaction effects of correct model × item-intercorrelation on errors in beta estimation are significant for both true construct type formative and true construct type reflective groups. Consequently, for a correct model where true construct type is formative, beta estimation errors actually increase with higher intercorrelations ($\beta_1 = 0.379$, $\beta_2 = 0.150$, both at $p < .001$). On the contrary, for a correct model where true construct type is reflective, beta estimation errors are reduced as item-intercorrelation becomes larger ($\beta_1 = -0.423$, $\beta_2 = -0.354$, both at $p < .001$).

Some Model Fit Indices Can Effectively Imply a Correctly Specified Measurement Model

In general, these results corroborate past findings (Jarvis et al., 2003; MacKenzie et al., 2005) which point to the ineffectiveness of global fit indices in signaling misspecification (Table 5). In particular, given that the common rules of thumb for minimal indications of good fit range from 0.90 to 0.95 (Hu & Bentler, 1999), the majority of the effects of misspecification on NFI, CFI, IFI, GFI, and AGFI for both true construct type groups are very small. Results of the current study, however, suggest that there are differences across the reflective and formative construct type groups with respect to model fit indices.

For the true construct type as reflective group, the correct model specification decreases χ^2 (-58.449), RMSEA (-0.042), and SRMR (-0.086) which all indicate better fit ($ps < .001$). While the implications of the magnitude of the effect on χ^2 is unclear, the effects for RMSEA and SRMR are quite large with respect to the common rules of thumb for both RMSEA and SRMR (that they should be less than .08) as indications of good fit (Kline, 2011). As item-intercorrelation and the correlation between the dependent variables increase, χ^2 , RMSEA, and SRMR decrease further ($ps < .001$);

similarly χ^2 , RMSEA, and SRMR decrease ($ps < .001$) with positive and larger correlations between the dependent variables. Note also that the interactive effects of correct model × item-intercorrelation counteract the individual negative main effects. Given the magnitudes of the main effects of using a correctly specified model, the effects of the item-intercorrelation, and the effects of the correlations between dependent variables, a correctly specified reflective measurement model still has lower χ^2 , RMSEA, and SRMR (on average) when compared with a misspecified one.

For the true construct type as formative group, the effect of a correctly specified measurement model actually increases χ^2 (2.719), RMSEA (0.001), and SRMR (0.022) ($ps < .05$). Increasing correlations between the dependent variables, on the contrary, reduces χ^2 (-105.711), RMSEA (-0.048), and SRMR (-0.076) ($ps < .0001$). The interaction of correct model × item-intercorrelation has significant negative effects on χ^2 (-28.558), RMSEA (-0.003), and SRMR (-0.045). The interaction of correct model × correlations between the dependent variables decreases χ^2 (-17.997) ($ps < .001$). Taken together, when a formative measurement model is specified in the analysis, χ^2 , RMSEA, and SRMR are inflated (i.e., indicating worse fit) even if it matches the true construct type, but these upward biases are overcome by positive and larger correlations between dependent variables and item-intercorrelation. Thus, after considering the magnitudes of the main and interactive effects in the regression results, a correctly specified formative measurement model still generally has lower χ^2 , RMSEA, and SRMR when compared with a misspecified one.

A Reflective Measurement Model More Accurately Predicts Factor Loadings

In general, when the true construct type was a formative measure, errors in factor loadings/weights were larger (Table 6). The average effects on errors in factor loadings/weights is 0.036 and ranges from -0.032

Table 5.
Regression Analyses: Results for Fit Indices.

Predictor	χ^2	RMSEA	SRMR	Normed Fit Index (NFI)	Comparative Fit Index (CFI)	Incremental Fit Index (IFI)	Goodness of Fit Index (GFI)	Adjusted Goodness-of-Fit Index (AGFI)
Set 1: True construct types is a formative measure (n = 52,446)								
Correct model	2.719 (1.004)**	0.001 (0.000)**	0.022 (0.001)***	-0.001 (0.000)**	-0.002 (0.000)***	-0.002 (0.000)***	0.000 (0.000)	-0.010 (0.001)***
Intercorrelation	4.187 (1.045)***	-0.006 (0.001)***	0.011 (0.001)***	0.017 (0.000)***	0.010 (0.000)***	0.010 (0.000)***	0.000 (0.000)	-0.000 (0.001)
r_{y1y2}	-105.711 (0.749)***	-0.048 (0.000)***	-0.076 (0.001)***	0.020 (0.000)***	0.019 (0.000)***	0.019 (0.000)***	0.033 (0.000)***	0.059 (0.000)***
Correct × Intercorrelation	-28.558 (1.501)***	-0.003 (0.001)***	-0.045 (0.001)***	0.004 (0.000)***	0.005 (0.000)***	0.005 (0.000)***	0.006 (0.000)***	0.013 (0.001)***
Correct × r_{y1y2}	-17.997 (1.074)***	-0.016 (0.001)***	-0.021 (0.001)***	0.004 (0.000)***	0.005 (0.000)***	0.005 (0.000)***	0.006 (0.000)***	0.021 (0.001)***
Set 2: True construct type is a reflective measure (n = 49,028)								
Correct model	-58.449 (1.216)***	-0.042 (0.001)***	-0.086 (0.001)***	0.017 (0.000)***	0.019 (0.000)***	0.019 (0.000)***	0.019 (0.000)***	0.040 (0.001)***
Intercorrelation	-14.691 (1.222)**	-0.003 (0.001)***	-0.039 (0.001)***	0.019 (0.000)***	0.014 (0.000)***	0.014 (0.000)***	0.002 (0.000)***	0.005 (0.001)***
r_{y1y2}	-92.042 (0.855)***	-0.044 (0.000)***	-0.076 (0.001)***	0.018 (0.000)***	0.020 (0.000)***	0.020 (0.000)***	0.029 (0.000)***	0.057 (0.000)***
Correct × Intercorrelation	91.524 (1.764)***	0.050 (0.001)***	0.128 (0.002)***	-0.025 (0.000)***	-0.026 (0.000)***	-0.026 (0.000)***	-0.028 (0.001)***	-0.046 (0.001)***
Correct × r_{y1y2}	3.744 (1.215)**	0.008 (0.001)***	0.002 (0.001)**	-0.002 (0.000)***	-0.002 (0.000)***	-0.002 (0.000)***	-0.004 (0.000)***	-0.014 (0.001)***

Note. The regression coefficients reported are unstandardized estimates. Standardized errors are in parentheses. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.
* $p < .10$. ** $p < .05$. *** $p < .001$.

Table 6.
Regression Analyses: Results for Absolute Errors in Factor Loading/Weight Estimation.

Predictor	Errors in λ_1	Errors in λ_2	Errors in λ_3	Errors in λ_4	Average
Set 1: True construct type is a formative measure (n = 52,466)					
Correct model	0.039 (0.005)***	0.050 (0.004)***	0.052 (0.004)***	0.046 (0.004)***	0.05
Intercorrelation	0.031 (0.005)***	0.031 (0.005)***	0.032 (0.005)***	0.031 (0.005)***	0.03
r_{y1y2}	-0.032 (0.003)***	-0.032 (0.003)***	-0.031 (0.003)***	-0.031 (0.003)***	-0.03
Correct × Intercorrelation	0.111 (0.007)***	0.130 (0.007)***	0.128 (0.007)***	0.136 (0.007)***	0.13
Correct × r_{y1y2}	0.019 (0.005)***	0.003 (0.005)	0.001 (0.005)	0.000 (0.005)	0.01
Set 2: True construct type is a reflective measure (n = 49,028)					
Correct model	0.099 (0.006)***	-0.110 (0.007)***	-0.102 (0.007)***	-0.105 (0.007)***	-0.05
Intercorrelation	0.004 (0.006)	-0.158 (0.007)***	-0.148 (0.007)***	-0.152 (0.007)***	-0.11
r_{y1y2}	-0.032 (0.004)***	-0.003 (0.005)	-0.003 (0.005)	0.001 (0.005)	-0.01
Correct × Intercorrelation	-0.248 (0.008)***	-0.099 (0.010)***	-0.111 (0.010)***	-0.106 (0.010)***	-0.14
Correct × r_{y1y2}	0.027 (0.006)***	-0.004 (0.007)	-0.004 (0.007)	-0.007 (0.007)	0.003

Note. The regression coefficients reported are unstandardized. Standardized errors are in parentheses.
* $p < .10$. ** $p < .05$. *** $p < .001$.

to 0.136. These effects are generally small when compared with factor loadings/weights which were set at 0.2, 0.7, and 0.9. The most notable biases in errors are

observed as item-intercorrelation increase, which results in an upward bias (average of 0.13 across the four loadings/weights).

The opposite results occur when the true construct type was a reflective measure. The majority of observed errors is negative with an average of -0.063 and a range from -0.248 to 0.099 . The magnitude of the errors is slightly larger than those when the true construct type was a formative measure. These results imply that when the true construct type was reflective, more accurate factor loadings/weights estimations can be expected especially when item-intercorrelations are high ($\text{Mean}_{\text{reflective}} = -0.14$).

Conclusion and Recommendations

Results of this study corroborate findings in extant literature that under the true construct type as reflective condition, regardless of whether or not the measurement model is correctly specified, betas (i.e., IV–DV relationship) would be overestimated. The inflations ranged in magnitude from 0.229 to 0.245 . These represent 91.6% to 98.0% for a moderate (i.e., correct betas = 0.25) and 57.3% to 61.3% for a strong (i.e., correct betas = 0.4) substantive relationship. These inflations, however, are mitigated as item-intercorrelation increases. The reduction of the upward biases ranged from -0.423 to -0.354 . These represent 141.2% to 169.2% for a moderate (i.e., correct betas = 0.25) and 88.5% to 105.8% for a strong (i.e., correct betas = 0.4) substantive relationship. On the contrary, results for the true construct type as formative condition were opposite. In this condition, while a correctly specified measurement model will reduce absolute errors in beta estimates, this reduction will be compromised as item-intercorrelations increase. In particular, while a correct specification of a true formative model can reduce errors in beta estimates by 0.068 to 0.217 (i.e., 27.2% to 86.8% for a moderate and 17.0% to 54.3% for a strong relationship), these reductions will be lessened by 0.150 to 0.379 (i.e., 60.0% to 151.2% and 37.5% to 94.8% , respectively), as items are more strongly correlated.

The caveat to take away from these results is that the effect of measurement model specification on substantive relationships must be considered alongside the effect of item-intercorrelation. The effects of item-intercorrelation, however, are different for each type of measurement model. On the one hand, when a measure is theorized to be reflective, the goal is to minimize the potential upward bias in beta estimates by selecting indicators which are highly correlated. On the other hand, when theory suggests a formative measure, the goal is to avoid the potential hindering effects of high item-intercorrelation by using indicators which are only slightly correlated.

The finding that item-intercorrelation has an impact on beta and factor loading/weight estimates is also important. First and foremost, it corroborates with scale development literature that it is very important when developing reflective measures that there are strong

inter-item correlations (Hinkin, 1995) among the items, but that very different steps should be taken to develop formative measures (Diamantopoulos & Sigauw, 2006). For a reflective measure, higher item-intercorrelation not only corrects for the upward bias in beta estimates but also reduces errors in factor loading estimates. On the contrary, increases in item-intercorrelation seem to be detrimental to formative measures in that the higher these correlations, the larger the errors in beta and factor weight estimations. It is important to note, however, that item-intercorrelation alone cannot determine whether or not a measure should be reflective or formative (indeed, our simulation has low and high intercorrelations for each true construct type). The theoretical underpinnings of the nature of a construct and the model fit indices—namely, χ^2 , RMSEA, and SRMR—must be considered together to make a more informed decision on measurement model choice.

Study 2: Application of Regression Analyses Results With Published Data

Prescriptive Guidance to Make Measurement Model Choice

The results of the simulation provide useful information on the implications of measurement model misspecification. Perhaps of greater utility, it can also be used to develop specific guidelines and an evaluative tool to make an informed measurement model decision. Using the regression analyses presented in Table 5 on the influences of correct model, items-intercorrelation, and correlation between dependent variables on global fit statistics (i.e., χ^2 , RMSEA, and SRMR) for both true construct type conditions, we establish two decision trees providing step-by-step instructions to help choose the appropriate measurement model for a construct. These decision trees are applicable for situations where theory is applicable, as well as where theory is lacking.

Using logistic regression, we also offer a formula to calculate the probability of a particular measurement model being correct with respect to each true construct type condition (see Table 7). This probability formula can be used as an evaluative tool when fit statistics are ambiguous or as additional empirical support for a theoretically derived measurement model. Finally, we consider two examples to demonstrate the application of the decision trees and probability formula.

It is worth noting that both the decision trees and probability formula are developed based on the regression analyses conducted on SEM results of global fit statistics for the simulated conceptual model (Figure 2). When we conducted the SEM analyses, the only variation across conceptual models is how the measurement model is specified (i.e.,

Table 7.
Study 2: Correlation Matrix of Correct Model and Model Fit Differences (Formative – Reflective) of Fit.

	1	2	3	4	5	6	7	8	9
1. Formative model	1.00								
2. $\Delta\chi^2$.15	1.00							
3. Δ RMSEA	.25	.88	1.00						
4. Δ SRMR	.08	.89	.82	1.00					
5. Δ Normed Fit Index (NFI)	-.13	.38	.42	.43	1.00				
6. Δ Comparative Fit Index (CFI)	-.14	.38	.42	.43	1.00	1.00			
7. Δ Incremental Fit Index (IFI)	-.14	.38	.42	.43	1.00	1.00	1.00		
8. Δ Goodness of Fit Index (GFI)	-.19	.37	.41	.42	1.00	1.00	1.00	1.00	
9. Δ Adjusted Goodness-of-Fit Index (AGFI)	-.17	.32	.35	.38	1.00	1.00	1.00	1.00	1.00

Note. "Formative Model" is coded as 0 if the formative model is the "true" model, and 1 if the reflective model is the "true" model. Model fit differences were calculated as the fit statistic from the formative model minus the fit statistic from the reflective model. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

either as formative or reflective). As a result, any observed changes in the regression analyses on global fit statistics can only be attributed to how the measurement model is specified.

Calculating the Probability of Choosing the Correct Measurement Model

We developed a formula to calculate the probability of choosing the correct measurement model. This formula is based on differences in fit statistics obtained from the results based on first assuming a formative structure, and then running the same model except with assuming a reflective model structure (i.e., statistics from the formative model minus statistics from the reflective model). A correlation matrix of the differences for all fit statistics can be found in Table 7. We conducted a logistic regression to calculate the probability of the formative measurement being true, and obtained the following logistic regression (see Table 8):

$$\text{Logit (Formative is true construct type)} = \pi = 0.19 + 0.0089 \Delta\chi^2 + 83.4 \Delta\text{RMSEA} - 13.86 \Delta\text{SRMR} + 44.62 \Delta\text{AGFI} \tag{3}$$

Thus, based on this finding, one may simply obtain χ^2 , RMSEA, SRMR, and AGFI and then subtract the respective fit indices of the reflective measurement model from the formative measurement model to get the differences, and then use the formula to calculate the expected value (i.e., π). To convert π from equation (3) to a probability, one applies the following equation:

$$\begin{aligned} &\text{Probability of formative is true} \\ &\text{construct type} = \exp[\pi]/[1 + \exp(\pi)], \end{aligned} \tag{4}$$

$$\begin{aligned} &\text{Probability of reflective is true} \\ &\text{construct type} = 1 - [\exp[\pi]/[1 + \exp(\pi)]] \end{aligned} \tag{5}$$

When there is an a priori theoretical measurement model, this formula can be used to calculate the probability that the theoretical measurement model is correct and use it as further support for the choice. When no a priori theory is present, this formula can help quantify the chances that a measurement model choice is correct based on fit statistics. In addition, should fit statistics point to different measurement model choices (e.g., χ^2 and RMSEA indicate different model superiority), this formula can be useful for suggesting which measurement model is more likely to be superior by taking into consideration a set of fit indices.

Introducing Two Decision Trees to Drive Measurement Model Choice

Based on regression results in Table 5, we derived two decision-making processes for determining whether a construct should be formative or reflective (Figures 3 and 4). For each true construct type condition, cutoffs for item-intercorrelation and the correlation between dependent variables were identified to help make a measurement model decision. The two processes provide the general guidelines for two conditions: (a) when an a priori theory exists for one measurement model (Figure 3) and (b) when no theory exists to support either measurement model (Figure 4).

When comparing an a priori theoretical measurement model and an alternative one, the assumption is that the theoretical measurement model is correct even when empirical results are inconclusive; however, one must also consider the potential for falsification. Therefore, our goal is to examine χ^2 , RMSEA, and SRMR to explain why the theoretical measurement model may not appear to be better fitting than the alternative one.

Table 8.
Study 2: Logistic Regression Results of Fit Statistics on
Formative Model Being the “True” Model.

	B (SE)	Wald
Intercept	0.19 (0.013)	213.37
$\Delta\chi^2$	0.0089 (0.00059)	226.27
Δ RMSEA	83.40 (1.48)	3,177.20
Δ SRMR	-13.86 (0.44)	996.69
Δ Adjusted Goodness-of-Fit Index (AGFI)	44.62 (1.32)	1,143.10

Note. $n = 47,791$; Pseudo R^2 (i.e., $R_{C\&S}^2$) for the model is .09. Since $R_{C\&S}^2 = 1 - (L_0 / LM)^{2/n}$, the low R^2 obtained is attributable to the large sample size of the simulated data set. Consequently, we further calculated that this logistic model correctly classifies mode type 61% of the time to support the utility of this logit in assessing probability of a correct formative model. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

When comparing two measurement models with no a priori theory—essentially an exploratory analysis—a common assumption is that the one with the lower χ^2 , RMSEA, and SRMR is correct. Yet our results show that better fit is not always associated with the correct choice. Our goal is to find out whether or not the apparently superior measurement model is truly better or likely to be an artifact of different levels of item-intercorrelation and the correlation between dependent variables.

Steps for each type of decision are depicted in Figures 3 and 4. Moreover, one can use the regression results presented in Table 5 to assess whether or not the preferred measurement model, be it based on theory or empirical evidence, is correct at specific levels of item-intercorrelations and correlation between dependent variables. Elaboration on the steps and explanation are provided in Appendices C and D.

Application 1: Comparing a Theoretically Derived Measurement Model With the Alternative

We used an analysis of the SERVQUAL measure to demonstrate the use of our decision tree in a situation with a theoretically based measurement model (Appendix C and Figure 3). SERVQUAL is a widely used measure (i.e., a google scholar count for all time yields 45,900 articles of which 22,200 were published between 2006 and 2016), and is particularly heavily used in hospitality research (Tanford, 2016).

For our example here, we used the correlation matrix published in Cenfetelli and Bassellier (2009). This correlation matrix provided us with the correlations of the five dimensions of SERVQUAL—assurance, empathy, reliability, responsive, and tangible—as well as correlations of the

measures for two dependent variables: satisfaction and purchase intention. This is an appropriate construct to illustrate this decision tree because SERVQUAL was developed to measure service quality with reflective indicators (e.g., Parasuraman et al., 1988), although some had suggested the possibility of it having formative indicators (e.g., Rossiter, 2002). Thus, while there is some debate about the nature of the measure, it begins with a theoretical basis to expect it to operate as a reflective measure. Using the specified correlation matrix, we analyzed the data as if it were a reflective measure, and then as a formative measure. Results of the analyses are reported in Tables 9 to 11.

As shown in Table 9, factor loadings/weights of the indicators for SERVQUAL are different for the reflective and formative measurement models. The factor loadings for the dependent variables—satisfaction and purchase intention, which are both specified as reflective measures—do not vary across measurement specification for the SERVQUAL construct. The estimated nature of the relationships of SERVQUAL with satisfaction and purchase intention are presented in Table 10. For both measures, the directions of the relationships remain the same, but the estimated effect sizes differ. A reflective model yields lower absolute path estimates (.79 for SERVQUAL–satisfaction; .77 for SERVQUAL–purchase intention) than a formative model (.86 for SERVQUAL–satisfaction; .84 for SERVQUAL–purchase intention). Table 11 provides a comparison of the fit indices. Note that all fit indices (e.g., RMSEA, SRMR, CFI, and AGFI) are almost identical, although the formative indices are marginally better.

Based on these results, we can demonstrate how to apply the decision tree depicted in Figure 3. First, fit statistics show that the formative measurement model has better fit which means that our theoretical measurement model seems (at least upon initial examination) to be inferior to the alternative one. We then turn to item-intercorrelation. The average correlation among the five SERVQUAL indicators is .73, which is greater than .70. We next look at the correlation between the two dependent variables—satisfaction and purchase intention. These two dependent measures are positively and strongly correlated at .66. According to the decision tree, for intercorrelations $> .7$, if the true (i.e., correct) measurement model were reflective, it would show worse fit than an incorrect one at all levels of r_{y1y2} . As this is the case, it suggests that the superior formative measurement model may appear to be superior due to misspecification biases; consequently, the theoretical, reflective measurement model should be used. Specifically, the SERVQUAL measure seems more appropriately specified as having reflective indicators. Incidentally, if we submit this to the decision tree without an a priori theory (Figure 4), we would arrive at the same conclusion that SERVQUAL is best operationalized as reflective. In addition, the

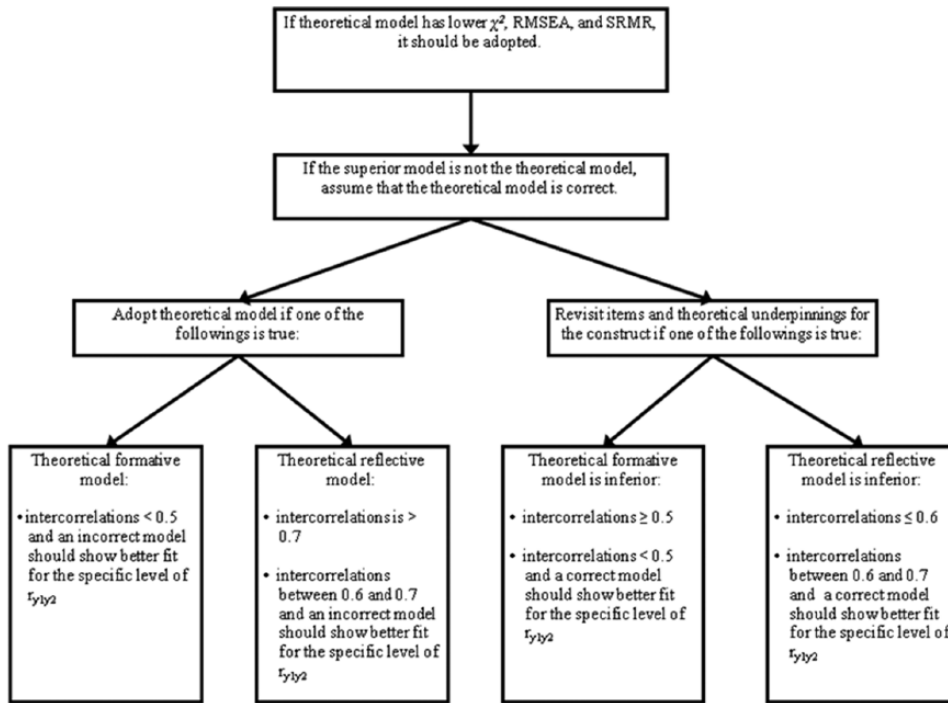


Figure 3.
Comparing a Theoretically Derived and an Alternative Measurement Model.
 Note. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

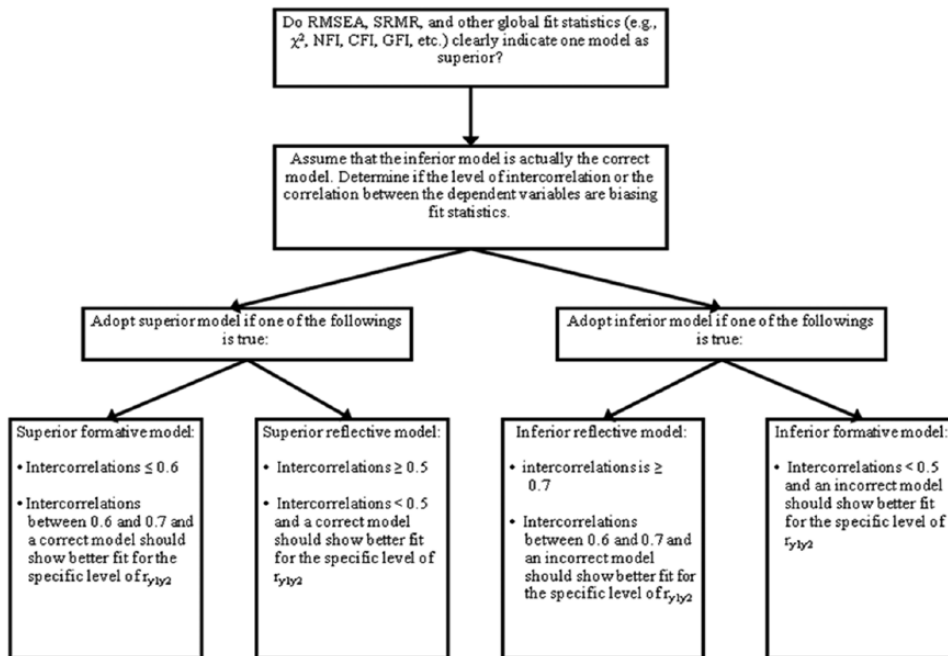


Figure 4.
Comparing Two Measurement Models Without an A Priori Theory.
 Note. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

Table 9.
Study 2: SEM Results of SERVQUAL—Factor Loadings/Weights.

	Reflective		Formative	
	Standardized Factor Loadings (t-Value)	SE	Standardized Weights (t-Value)	SE
SERVQUAL				
Assurance	0.88	0.22	0.43	
Empathy	0.80 (23.41)	0.36	-0.09 (-2.05)	
Reliability	0.92 (31.11)	0.15	0.38 (5.64)	
Responsive	0.86 (27.08)	0.26	0.13 (2.41)	
Trust	0.80 (23.18)	0.37	0.11 (2.40)	
Satisfaction				
Sat1	0.86	0.26	0.87	0.25
Sat2	0.87 (26.07)	0.25	0.87 (26.34)	0.25
Sat3	0.90 (28.00)	0.19	0.90 (28.45)	0.19
Sat4	0.92 (29.39)	0.15	0.92 (29.94)	0.15
Purchase intention				
Int1	0.93	0.14	0.94	0.12
Int2	0.80 (23.79)	0.37	0.80 (24.19)	0.37
Int3	0.88 (28.56)	0.23	0.87 (29.21)	0.24

Note. All t-values are significant at $p < .05$. SEM = structural equation modeling.

Table 10.
Study 2: SEM Results of SERVQUAL—Path Estimates.

	Reflective	Formative
	Standardized Path Loadings (t-Value)	Standardized Path Loadings (t-Value)
SERVQUAL to		
Satisfaction	0.79 (19.03)	0.86 (12.66)
Purchase intention	0.77 (19.58)	0.84 (13.08)

Note. All t-values are significant at $p < .05$. SEM = structural equation modeling.

probability equations (3) and (4) indicate a 72.5% chance that the reflective measurement model choice is correct based on fit statistics.

Application 2: Comparing Two Measurement Models Without an A Priori Theory

As a second example, we wanted to illustrate how empirically derived results can inform measurement model choice when existing theory is equivocal or contradictory (Appendix D and Figure 4). For this example, we examined SES. The effects of SES is often examined in the hospitality context under a wide range of topics including cultural or country differences in service perceptions, consumer behaviors in service settings, and service employees’ experience and management. Incidentally, *Cornell Hospitality Quarterly* alone had 10 publications with SES as a variable

Table 11.
Study 2: SEM Results of SERVQUAL—Model Fit Indices.

	Reflective	Formative
χ^2 (df)	288.43 (52)	216.97 (43)
RMSEA	.09	.09
90% CI for RMSEA	[0.08, 0.10]	[0.08, 0.10]
SRMR	.06	.04
Comparative Fit Index (CFI)	.98	.99
Adjusted Goodness-of-Fit Index (AGFI)	.87	.88

Note. SEM = structural equation modeling; RMSEA = root mean square error approximation; CI = confidence interval; SRMR = standardized root mean square residual.

between 2008 and 2015. Theories exist in support of both formative and reflective SES measures (e.g., Borsboom, 2008; Edwards & Bagozzi, 2000; Nunnally & Bernstein, 1994; Podsakoff et al., 2006). Empirical data are also inconclusive, with some measurement models seeming to fit better as formative, whereas others fit better as reflective. It is thus difficult to reconcile the contradiction between theory and empirical evidence, and so this is an example where a research may indeed not have a clear a priori theoretical basis for an appropriate measurement model. For this example, we examined the results reported by Christie and Barling (2009).

Christie and Barling (2009) examined the direct impact of SES on personal control and work stress with a sample of 3,419 people. As in other studies concerning SES, three

Table 12.
Study 2: SEM Results SES—Factor Loadings/Weights.

	Reflective		Formative	
	Standardized Factor Loadings (t-Value)	SE	Standardized Weights (t-Value)	SE
SES				
Income	0.58	0.67	0.52	1
Prestige	0.65 (20.68)	0.58	0.37 (45.49)	1
Education	0.55 (19.98)	0.70	0.36 (5.78)	1
Personal control				
PC1	0.68	0.54	0.68	0.54
PC2	0.87 (40.13)	0.25	0.87 (40.01)	0.25
PC3	0.84 (40.16)	0.30	0.84 (40.07)	0.30
Work stress				
WS1	0.58	0.66	0.58	0.66
WS2	0.73 (24.08)	0.47	0.73 (23.83)	0.47
WS3	0.65 (24.19)	0.57	0.65 (24.05)	0.57

Note. All t-values are significant at $p < .05$. SEM = structural equation modeling; SES = socioeconomic status.

Table 13.
Study 2: SEM Results SES—Path Estimates.

	Reflective	Formative
	Standardized Path Loadings (t-Value)	Standardized Path Loadings (t-Value)
SES to		
Personal control	0.35 (13.44)	0.26 (9.24)
Work stress	-0.46 (-14.49)	-0.35 (-9.76)

Note. All t-values are significant at $p < .05$. SEM = structural equation modeling; SES = socioeconomic status.

indicators—income, occupational prestige, and education—were used to operationalize SES. Using SEM, we examined the direct impact of SES on both personal control and work stress with SES specified as both a reflective and formative measurement model. Results are summarized in Tables 12 to 14.

As shown in Table 12, factor loadings of the three indicators for SES are different for the reflective and formative measurement models. Note again that the factor loadings for the dependent variables—namely, personal control and work stress, which are both specified as reflective measures—do not vary across measurement specification for the SES construct. The estimated nature of the relationships of SES with personal control and work stress are presented in Table 13. For both measures, the directions of the relationships remain the same, but the estimated effect sizes differ. Consistent with findings in Study 1, a reflective measurement model yields higher absolute path estimates (.35 for SES–personal control; -.46 for SES–work stress) than a formative measurement model (.26 for SES–personal

Table 14.
Study 2: SEM Results SES—Model Fit Indices.

	Reflective	Formative
χ^2 (df)	442.90 (25)	480.92 (23)
RMSEA	0.07	0.08
90% CI for RMSEA	[0.06, 0.08]	[0.07, 0.08]
SRMR	0.06	0.07
Comparative Fit Index (CFI)	0.96	0.95
Adjusted Goodness-of-Fit Index (AGFI)	0.95	0.94

Note. SEM = structural equation modeling; SES = socioeconomic status; RMSEA = root mean square error approximation; CI = confidence interval; SRMR = standardized root mean square residual.

control; -.35 for SES–work stress). Table 14 provides a comparison of the fit indices. Note that all fit indices (e.g., RMSEA, SRMR, CFI, and AGFI) are almost identical, although the reflective indices are marginally better.

We use these empirical outcomes to demonstrate how to apply the decision tree depicted in Figure 4. This decision tree is used to compare two measurement models without an a priori theory. First, fit statistics show that the reflective measurement model yields better fit, although the slight advantage may not be so strong as to make someone feel it is clearly superior. We thus calculated the probability based on equations (3) and (4) and found that there is a 77.3% chance that the reflective measurement model is most accurate.

The results thus far indicate that the reflective measurement model is correct. We then examine the average correlation among the three SES indicators, finding them to be .36 (i.e., income–prestige = .39, income–education = .30, prestige–education = .38), which is smaller than .5. For intercorrelations $< .5$, whether or not the correct measurement model has better fit than an incorrect measurement model depends on the level of $r_{y_1y_2}$. The correlation between the two dependent variables ($r_{y_1y_2}$)—personal control and work stress—are negatively and slightly correlated at $-.16$. Using the regression results in Table 5, we see that if the formative measurement model were the correct, it would have shown better fit for intercorrelations = .36 and $r_{y_1y_2} = -.16$. As this is not the case, it suggests that the superior reflective measurement model is not likely to be an artifact. We therefore conclude that the reflective measurement model should be adopted. Note that the same conclusion is derived if we submit the SES results to the decision tree with an a priori theory (Figure 3).

Conclusions and Future Research

The use of multiple-item measures in hospitality and service research is ubiquitous, and that is why greater consideration needs to be devoted to the relationship between a focal construct and its measures. One must be cognizant of

the ample evidence provided in past literature as well as in the current study that using multiple-item measures can only enhance accuracy of measurement and predictions if these measures are specified correctly (Diamantopoulos & Winklhofer, 2001; Edwards & Bagozzi, 2000). Extant model specification literature has informed us of some of the empirical consequences of misspecification, such as errors in beta estimations (Jarvis et al., 2003; MacKenzie et al., 2005) and factor loadings (Howell et al., 2007). Scholars have stressed the importance of model specification on the basis of theory, but in some cases, theories may support different measurement models. In other cases, especially for newly developed constructs, measurement model theories may not yet exist. These situations highlight a gap in the literature regarding how to identify the correct model when confronted with contradictory, vague, or non-existent theory. The current study contributes to hospitality and service research by providing concrete procedures for identifying the correct measurement model using empirical outcomes when theories are unclear or contradicting, as well as when theories are not yet developed. In addition, we developed a formula for calculating the probability of a particular model being correct. This formula quantifies the degree to which one model specification is superior. It also helps determine a clear model choice when empirical outcomes alone are equivocal.

Using the decision processes proposed in this article, we were able to resolve the conflict between theory and empirical evidence for two sets of actual data. It is, however, only two examples. To expand research in this area, we believe that the decision processes need to be tested with other constructs. Future research may try to replicate our results with other constructs. When the procedures have undergone more tests with various constructs, it may be worthwhile to expand the level of the relevant parameters, such as varying item-intercorrelations, to allow for a more fine-grained estimation of their influence on empirical outcomes. Another

way to extend the current study is to include more variables (e.g., more than one exogenous variable) and to test more complex structural models (e.g., with mediating variables).

The findings of this study open a new avenue for model specification research toward making more informed decisions regarding the nature of constructs. The decision-making processes and probability formula could also be applied as a scale and construct validation procedure. In a recent discussion on alternative models testing, Vandenberg and Grelle (2009) found that the bulk of confirmatory factor analyses in studies published in notable organizational science journals were merely conducted to demonstrate measurement invariance and/or discriminant validity (Vandenberg & Grelle, 2009, p. 185). Only around 38% of studies deployed the method in a manner adhering to Popper’s (1959) disconfirmatory strategy by pitting a theoretical measurement model against competing ones. It is worth noting that these competing measurement models encompassed only alternatives with different dimensional structure (e.g., pitting a theoretical two-factor structure against a single factor structure). No study that we know of compared a theoretical reflective measurement model with an alternative formative measurement model or vice versa, even when a competing theory exists. This may very well be attributed to the lack of clear guidelines especially for comparing reflective and formative measures. In this regard, we believe that this current study will facilitate the use of SEM to rule out alternative theoretical measurement models, thereby providing a more stringent validation test for constructs of interest.

In conclusion, we hope that the results and actionable recommendations of this study will encourage the use of our decision-making processes to complement theoretical reasons in driving measurement model choice. We also hope that this study will stimulate more research in this direction of measurement model specification.

Appendix A

Summary of Methodology.

	Procedure	Data Source	Variables	Supplements/ Results
Study I	Step I: Simulation 540 scenarios, each simulated 100 time, yielding $n = 54,000$ cases	N/A	True construct types (2): Formative vs. Reflective Item-intercorrelation (3): .2 vs. .7 vs. .9 Factor loadings/weights (3): 0.2 vs. 0.7 vs. 0.9 Correlation between dependent variables (5): -.6 vs. -.25 vs. 0 vs. .25 vs. .6 Beta (IV–DV relationship) (6): 0.4, 0.4 vs. 0.4, 0.25 vs. 0.25, 0.25 vs. 0.25, -0.25 vs. 0.4, -0.25 vs. 0.4, -0.4	Appendix B

(continued)

Appendix A (continued)

	Procedure	Data Source	Variables	Supplements/ Results
	Step 2: SEM analyses Set 1: 54,000 analyses assuming a formative measure Set 2: 54,000 analyses assuming a reflective measure	Simulated data in Step 1, analyzed first assuming a formative structure, and then reanalyzed assuming a reflective structure	SEM analyses on 54,000 × 2 (<i>Formative specification and Reflective specification</i>)	
	Step 3: Regression analyses	SEM analysis results in Step 2	<p>Predictors:</p> <ul style="list-style-type: none"> • Correct model: <i>True Construct Type matches SEM specification vs. True Construct Type mismatches SEM specification</i> (Refer to Table 1) • Item-intercorrelation • Correlation between dependent variables ($r_{y_1y_2}$) <p>Dependent variables:</p> <ol style="list-style-type: none"> Errors in beta estimation (IV–DV relationship) Fit Indices Errors in factor loadings/weights estimation 	Tables 1 to 6
Study 2	Step 1: Logistic regression	Regression analysis results for the dependent variable fit indices in Step 3	<p>Predictors:</p> <ul style="list-style-type: none"> • Change in Chi-square • Change in RMSEA • Change in SRMR • Change in Adjusted Goodness-of-Fit Index (AGFI) <p>Dependent variable: Dichotomized where 1 = Formative model is true and 0 = Reflective model is true (π)</p>	Tables 7 and 8
	Step 2: Decision trees	Regression analysis results for the dependent variable fit indices in Step 3.	<p>Example: $\chi^2 = 2.719(\text{correct model}) + 4.187(\text{intercorrelation}) - 105.711(r_{y_1y_2}) - 28.558(\text{correct model} \times \text{intercorrelation}) - 17.997(\text{correct model} \times r_{y_1y_2})$</p> <p>The correct model variable is dummy coded as 1 = correct model used and 0 = incorrect model used. As such, the interaction terms will only change (in this case, reduce) χ^2 when the measurement model was incorrectly specified. It follows then that there are conditions when a measurement model is incorrectly specified but generates lower χ^2 which could be misinterpreted as the accurate measurement model for a construct. The same reasoning can be applied to other fit statistics. Cutoffs and ranges specified in the decision trees are developed following this logic.</p>	Appendix C and D Table 5
	Step 3: Illustrative examples	Cenfetelli and Bassellier (2009) Christie and Barling (2009)	<ol style="list-style-type: none"> Predictor: Servqual; dependent variables: Satisfaction and Purchase Intention Predictor: socioeconomic status; dependent variables: Personal Control and Work Stress 	Tables 9 to 14

Note. SEM = structural equation modeling; RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

Appendix B

Simulation Data Generation Steps

1. Determine simulation parameters (from 540 total possible parameters combinations):
 - Item-intercorrelations of indicators of the independent variable (r_{xx})
 - Relationships between the indicator and independent variables (γ for the formative model and factor loadings λ for the reflective model)
 - Correlation between the two dependent variables ($r_{y1,y2}$)
 - Set of relationships between the independent and dependent variables (β_1 and β_2)
2. Given $r_{y1,y2}$, β_1 , and β_2 , generate variables that conform to the specified relationships (create 500 observations).
3. Create indicators of the independent variable, based on specified relationship (γ or λ) (for each of the 500 observations).
4. Create indicators of the dependent variables, as three-item reflective measures, based on fixed λ of 0.95) (for each of the 500 observations).
5. Create covariance matrix based on the 500 observations of indices.
6. Analyze data as both a reflective (see Figure 2a) and formative (see Figure 2b) model.
7. Record parameters and fit statistics from analysis.
8. Repeat steps 2 to 8 100 times.
9. Repeat steps 1 to 9 until all 540 parameter combinations have been simulated.

Appendix C

Elaboration on Decision-Making Process.

Comparing a Theoretically Derived Model Versus an Alternative Model (Figure 3)

1. Are χ^2 , RMSEA, and SRMR lower (numerically) for the theoretical model?
 - i. If theoretical model has lower χ^2 , RMSEA, and SRMR, it should be adopted.
 - ii. If not and the theoretical model is formative, go to 2.
 - iii. If not and the theoretical model is reflective, go to 3.
 2. When a superior model is reflective but the theoretical model is formative, we'll assume that the theoretical model is correct. We'll apply the regression models for the condition where a formative model is the correct model (i.e., upper half of Table 4) to establish that the superior model is an artifact.
Decision rules:
 - i. For intercorrelations $\geq .5$, if the correct model were a formative model, it would have shown better fit than an incorrect model at all levels of $r_{y1,y2}$. Since this is not the case, it suggests that the superior reflective model is not an artifact. Individual items and theory to support the theoretical formative model should be examined and the alternative reflective model should be given careful consideration.
 - ii. For intercorrelations $< .5$, whether or not the correct model has better fit than an incorrect model depends on the level of $r_{y1,y2}$. Use the regression model in Table 4 to calculate the estimates of each fit index for a correct and an incorrect model.
 - If a correct model shows better fit for the specific level of $r_{y1,y2}$, then it suggests that the superior reflective model is not an artifact; therefore, individual items and theory to support the theoretical formative model should be examined and the alternative reflective model should be given careful consideration.
 - If an incorrect model shows better fit for the specific level of $r_{y1,y2}$, then it suggests that the superior reflective model may be an artifact; therefore, the theoretical formative model should be adopted.
 3. When a superior model is formative but the theoretical model is reflective, we'll assume that the theoretical model is correct. We'll apply the regression models for the condition where a reflective model is the correct model (i.e., lower half of Table 4) to establish that the superior model is an artifact.
Decision rules:
 - i. For intercorrelations ≤ 0.6 , if the correct model were a reflective model, it would have shown better fit than an incorrect model at all levels of $r_{y1,y2}$. Since this is not the case, it suggests that the superior formative model is not an artifact; Individual items and theory to support the theoretical formative model should be examined and the alternative reflective model should be given careful consideration.
-

(continued)

Appendix C (continued)

Comparing a Theoretically Derived Model Versus an Alternative Model (Figure 3)

- ii. For intercorrelations between .6 and .7, whether or not the correct model has better fit than an incorrect model depends on the level of $r_{y_1y_2}$. Use the regression model in Table 4 to calculate the estimates of each fit index for a correct and an incorrect model.
 - If a correct model shows better fit for the specific level of $r_{y_1y_2}$, then it suggests that the superior formative model is not an artifact; therefore, individual items and theory to support the theoretical formative model should be examined and the alternative reflective model should be given careful consideration.
 - If an incorrect model shows better fit for the specific level of $r_{y_1y_2}$, then it suggests that the superior formative model may be an artifact; therefore, the theoretical reflective model should be adopted.
- iii. For intercorrelations > 0.7 , if the correct model were a reflective model, it would have shown worse fit than an incorrect model at all levels of $r_{y_1y_2}$. Since this is the case, it suggests that the superior formative model is an artifact; therefore, the theoretical reflective model should be adopted.

Note. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

Appendix D

Elaboration on Decision-Making Process.

Comparing Two Measurement Models With No a Priori Theory (Figure 4)

1. Do χ^2 , RMSEA, and SRMR clearly indicate one model as superior (i.e., fit statistics are in agreement)?
 - i. If not, calculate the probability for a correct model with equation 4 and 5 to determine a superior model. Then proceed.
 - ii. If superior model is formative, go to 2.
 - iii. If superior model is reflective, go to 3.
2. When a superior model is formative, we'll assume that the correct model is a reflective model. We'll apply the regression models for the condition where a reflective model is the correct model (i.e., lower half of Table 4) to help us rule out the possibility that the superior formative model is actually incorrect but is biased to show better fit statistics.
Decision rules:
 - i. For intercorrelations $\leq .6$, if the correct model were a reflective model, it would have shown better fit than an incorrect model at all levels of $r_{y_1y_2}$. Since this is not the case, it suggests that the superior formative model is not an artifact; therefore, it should be adopted.
 - ii. For intercorrelations between .6 and .7, whether or not the correct model has better fit than an incorrect model depends on the level of $r_{y_1y_2}$. Use the regression model in Table 4 to calculate the estimates of each fit index for a correct and an incorrect model.
 - If a correct model shows better fit for the specific level of $r_{y_1y_2}$, then it suggests that the superior formative model is not an artifact; therefore, the superior formative model should be adopted.
 - If an incorrect model shows better fit for the specific level of $r_{y_1y_2}$, then it suggests that the superior formative model may be an artifact; therefore, the inferior reflective model should be adopted.
 - iii. For intercorrelations $\geq .7$, if the correct model were a reflective model, it would have shown worse fit than an incorrect model at all levels of $r_{y_1y_2}$. Since this is the case, it suggests that the superior formative model is an artifact; therefore, the inferior reflective model should be adopted.
3. When a superior model is reflective, we'll assume that the correct model is a formative model. We'll apply the regression models for the condition where a formative model is the correct model (i.e., upper half of Table 4) to help us rule out the possibility that the superior reflective model is actually incorrect but is biased to show better fit statistics.
Decision rules:
 - i. For intercorrelations $\geq .5$, if the correct model were a formative model, it would have shown better fit than an incorrect model at all levels of $r_{y_1y_2}$. Since this is not the case, it suggests that the superior reflective model is not an artifact; therefore, it should be adopted.
 - ii. For intercorrelations $< .5$, whether or not the correct model has better fit than an incorrect model depends on the level of $r_{y_1y_2}$. Use the regression model in Table 4 to calculate the estimates of each fit index for a correct and an incorrect model.
 - If a correct model shows better fit for the specific level of $r_{y_1y_2}$, then it suggests that the superior reflective model is not an artifact; therefore, the superior reflective model should be adopted.
 - If an incorrect model shows better fit for the specific level of $r_{y_1y_2}$, then it suggests that the superior reflective model may be an artifact; therefore, the inferior formative model should be adopted.

Note. RMSEA = root mean square error approximation; SRMR = standardized root mean square residual.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, or publication of this article.

References

- Alvarez, M., & Asugman, G. (2009). Issues in formative measurement: Response to Murphy. *Annals of Tourism Research, 36*, 738-742.
- Bagozzi, R. P. (1984). A prospectus for theory construction in marketing. *Journal of Marketing, 48*, 11-29.
- Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox. *Psychological Methods, 12*, 229-237.
- Bedeian, A. G., Sturman, M., & Streiner, D. (2009). Decimal dust, significant digits, and the search for stars. *Organizational Research Methods, 12*, 687-694.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley.
- Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods, 16*, 265-284.
- Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305-314.
- Borsboom, D. (2008). Latent variable theory. *Measurement, 6*, 25-53.
- Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research, 5*, 3-52.
- Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly, 33*, 689-707.
- Christie, A. M., & Barling, J. (2009). Disentangling the indirect links between socioeconomic status and health: The dynamic roles of work stressors and personal control. *Journal of Applied Psychology, 94*, 1466-1478.
- Diamantopoulos, A. (2006). The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modelling in Management, 1*, 7-17.
- Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*, 263-282.
- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*, 269-277.
- Dickinger, A., & Stangl, B. (2013). Website performance and behavioral consequences: A formative measurement approach. *Journal of Business Research, 66*, 771-777.
- Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Method, 14*, 370-388.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*, 155-174.
- Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research, 19*, 440-452.
- Helm, S. (2005). Designing a formative measure for corporate reputation. *Corporate Reputation Review, 8*, 95-109.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*, 967-988.
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods, 12*, 205-217.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariances structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55.
- Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model mis-specification in marketing and consumer research. *Journal of Consumer Research, 30*, 199-218.
- Kline, R. B. (2011). *Principle and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Law, K., & Wong, C. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management, 25*, 143-160.
- MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114*, 533-541.
- MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model mis-specification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*, 710-730.
- Murphy, J., & Hofacker, C. F. (2009). Rigor in tourism research: Formative and reflective constructs. *Annals of Tourism Research, 36*, 730-734.
- Nadeau, J., Heslop, L., O'Reilly, N., & Luk, P. (2009). Modeling place image: Specifying indicator direction. *Annals of Tourism Research, 36*, 735-738.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). Servqual. *Journal of Retailing, 64*, 12-40.
- Podsakoff, N. M., Shen, W., & Podsakoff, P. M. (2006). The role of formative measurement models in strategic management research: Review, critique, and implications for future research. In D. J. Ketchen & D. D. Bergh (Eds.), *Research methodology in strategy and management*, (Vol. 3., pp. 197-252). Oxford, UK: Emerald.
- Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.
- Prasnikar, J., Rajkovic, T., & Zabkar, V. (2010). Summer tourist perceptions of service quality. *Annals of Tourism Research, 27*, 1181-1185.
- Rosser, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing, 19*, 305-335.

- Schwab, D. P. (1980). Construct validity in organizational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (pp. 3-43). Greenwich, CT: JAI Press.
- Tanford, S. (2016). Antecedents and outcomes of hospitality loyalty: A meta-analysis. *Cornell Hospitality Quarterly*, 57, 122-137.
- Vandenberg, R. J., & Grelle, D. M. (2009). Alternative model specifications in structural equation modeling: Facts, fictions, and truth. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 165-191). New York, NY: Routledge.
- Žabkar, V., Brenčič, M. M., & Dmitrović, T. (2010). Modelling perceived quality, visitor satisfaction and behavioural intentions at the destination level. *Tourism Management*, 31, 537-546.

Author Biographies

Elisa K. Chan (PhD, Cornell University) is an assistant professor at Ecole hôtelière de Lausanne, Switzerland. Her research

focuses on consumer value and consumer experience in service settings.

Michael Sturman (PhD, Cornell University) is the Kenneth and Marjorie Blanchard professor of Human Resources and a professor of management at Cornell University's School of Hotel Administration. His current research focuses on the prediction of individual job performance over time and the influence of compensation systems.

Sanghee Park (PhD, Cornell University) is an assistant professor at Rutgers University's School of Management and Labor Relations and a Louis O. Kelso Fellow. Her primary research interests are on compensation systems, pay and employee motivation, dyadic relationships, and impression management.

Chealsea Vanderpool (PhD, Cornell University) is an assistant professor at the Governor's State University's College of Business. Her research focuses on work-family issues, gender differences and attitudes, and research methods and statistics.