5-24-2013

Can We Assess Formative Measurement using Item Weights? A Monte Carlo Simulation Analysis

Jun He
University of Michigan - Dearborn, junhe@umich.edu

Can We Assess Formative Measurement using Item Weights? A Monte Carlo Simulation Analysis

Jun He
University of Michigan-Dearborn
junhe@umich.edu

ABSTRACT

This study questions a common practice of using item weights for construct validity tests in the application of formative measurement. The practice does not confirm to the theoretical formation of formative constructs. A Monte Carlo simulation analysis is conducted to examine the practice. The results clearly demonstrate that item weights do not reflect the true design of the focal formative construct; using item weights may mislead the development and the application of formative instruments in empirical research.

Keywords

Formative measurement, item weights, construct validity.

INTRODUCTION

In recent years, formative measurement has received much attention among researchers. Some researchers suggest rethinking the theoretical origin of constructs and adopt measurement models accordingly (Chin, 1998; Marakas et al., 2007). Procedures and methods have been proposed for developing and utilizing formative measurement (e.g., Diamantopoulos and Winklhofer, 2001; Jarvis et al., 2003; Podsakoff et al., 2003; Diamantopoulos, 2006; Diamantopoulos et al., 2008). But contradictory recommendations (Howell et al., 2007) and unsolved statistical issues (Diamantopoulos et al., 2008) make the use of formative measurement an art rather than a science.

A central issue in the development of formative measurements is the test of construct validity. The validity of a construct should be established before the research model can be tested for any empirical inference (Straub, 1989). For reflective measurement, which is derived from the traditional Factor Analysis and Classical Test Theory (Bollen and Lennox, 1991), various validity test techniques have been developed and validated in the literature. For formative measurement, however, there is no widely-accepted technique that is both theoretically profound and empirically validated; existing guidelines in the literature are ambiguous in general and contradictory on certain issues.

This study attempts to question one common practice in the validation of formatively measured constructs: using item weights for the decision of retaining or dropping measurement items. The practice has been suggested and applied by many researchers in their endeavor of using formative measurement (e.g., Loch et al., 2003; Marakas et al., 2007; Urbach and Ahlemann, 2010), but has not been formally validated with careful examination on its methodological grounds and statistical implications. The study argues that using item weights violates the underlying assumptions of formative measurement. To investigate the query, Monte Carlo simulation analysis is employed as the main research method.

The paper is organized as follows. First, the importance of construct validity tests is discussed, and the properties of formative and reflective measurements are examined. Then, Monte Carlo simulation analysis is introduced as the main research method. The design of the simulation analysis is explained, and the results are summarized. The paper ends with a discussion of the results and their implications for the application of formative measurement in empirical research.

CONSTRUCT VALIDITY

In his seminal work, Straub (1989) discussed the importance of instrument validation in the research of information systems (IS), and provided guidelines on how to test instrument validity in empirical research. Figure 1 summarizes the recommended procedure of validating instruments.

Table 1. Procedures and Tests of Instrument Validity, Adapted from Straub (1989)

Procedure of Validity Tests	Purpose	Questions to Ask	Common Methods	Techniques
1. Content	The representatives of	Are instrument measures drawn from all	Review process	

Validity	measures	possible measures of the properties under investigation?		
2. Construct Validity	The meaningfulness of constructs as measured	Do measures show stability across methodologies? Are the data a reflection of true scores or artifacts of the kind of instrument chosen?	Convergent validity test and Discriminant validity test	• MTMM analysis • Factor analysis • Factor loadings and AVE in SEM
3. Reliability	Stability of measures	Do measures show stability across the units of observation? That is, could measurement error be so high as to discredit the findings?	Measurement reliability tests	• Cronbach alpha • Composite reliability
4. Internal Validity	Properties of the hypotheses, soundness of the research model	Are there untested rival hypotheses for the observed effects?	Grounding the research model in established theories	Literature review and hypothesis development
5. Statistical Conclusion Validity	Testing results of the research model	Do the variables demonstrate relationships not explainable by chance or some other standard of comparison?	Examination of the path coefficients and model fit indexes	• Regression • MANCOVA • SEM using PLS, LISREL etc.

According to Straub (1989), construct validity covers convergent validity and discriminant validity. A construct measurement demonstrates convergent validity if its measures correlates strongly (both in significance and in magnitude) with and converges on the designated construct; the measurement demonstrates discriminant validity if the correlations between the measures and other constructs are not as strong as that to the focal construct, i.e., of smaller magnitudes. Straub (1989) further recommended multitrait-multimethod (MTMM) techniques, and confirmatory or principal components factor analysis, for the assessment of convergent validity. If the method of structure equation modeling (SEM) is employed, researchers can examine factor loadings and average variance extracted (AVE) for the test of construct validity (Chin, 1998; Gefen et al., 2000; Urbach and Ahlemann, 2010).

Struab's recommendations for instrument validation have profound influence in the field of IS research. Construct validity test has been accepted as an essential component of empirical research. However, one should note that Straub's recommendations were designed for reflective measurements. The question that Straub has asked for the examination of construct validity is that "Are the data a reflection of true scores or artifacts of the kind of instrument chosen?" (Straub, 1989; p. 150); the question suggests a reflective nature of the construct measurement.

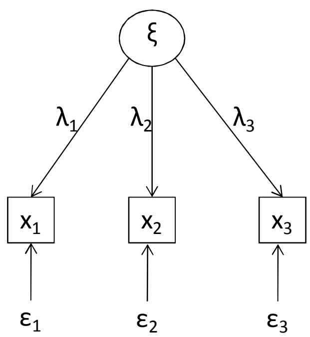
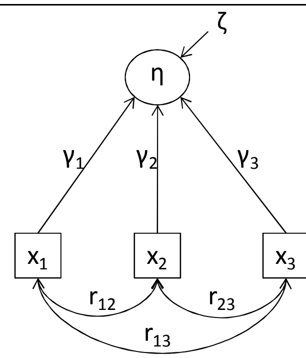
REFLECTIVE VS. FORMATIVE MEASUREMENT

Construct operationalization is an important part of the empirical research process in social science and management research. Abstract concepts are assessed by certain sets of measures before their hypothesized effects can be empirically tested within a theoretical network. The statistical test method of structure equation modeling crystallizes the issue by separating measurement models from structural models: measurement models depict the nature and direction of relationships between constructs (also labeled as latent variables for that they cannot be directly observed) and their measures or indicators, while structural models depict the relationships among the constructs themselves (Bollen, 1989; Byrne, 1998). Conventionally, constructs are modeled as causes of measures, meaning that variation in a construct leads to variation in its measures (Bollen, 1989). The mathematical modeling of reflective measurement conveys such design by treating each measure as a function of the designated construct plus error (Edwards and Bagozzi, 2000).

In their seminal work, Bollen and Lennox (1991) discussed the limitations of conventional construct measurement models and presented an alternative model in which indicators are modeled as causes, rather than effects as suggested by the classical test theory, of a latent variable. The model is labeled as formative measurement for that the meaning of a construct is formed by rather than reflected from its measures (Diamantopoulos et al., 2008; Kim et al., 2010). The mathematical modeling accordingly formulates the construct as a function of its measures plus error (Edwards and Bagozzi, 2000). The work led to great interest among researchers on the use of formative measurement in social science and management research (e.g., Diamantopoulos and Winklhofer, 2001; Jarvis et al., 2003; Podsakoff et al., 2003; Diamantopoulos, 2006). In a recent issue of MIS Quarterly, much of the designated Special Research Commentary Series on Quantitative Research was devoted to the difference between formative and reflective measurements (Gefen et al., 2011).

Many statistical issues remain unsolved for formative measurement (Diamantopoulos et al., 2008). However, the discussion has reached a common agreement among researchers regarding the properties of each measurement. These properties are briefly summarized in Table 2.

Table 2. Properties of Reflective Measurement and Formative Measurement

	Reflective Measurement	Formative Measurement
Theoretical Foundation	Based on Factor Analysis (Spearman, 1904) and Classical Test Theory (Lord & Novick, 1968; Spearman, 1910) with a common assumption that a construct (i.e., the latent variable) determines its indicators.	Alternative approach from the traditional reflective measurement with the assumption that indicators cause the focal construct (i.e., the latent variable) (Blalock, 1964; Bollen and Lennox, 1991)
Mathematical Model	$x_i = \lambda_i \xi + \varepsilon_i$ in which, x_i is the ith indicator of the latent variable ξ, ε_i is the measurement error for the ith indicator, and λ_i is a coefficient (loading) capturing the effect of ξ on x_i.	$\eta = \sum_{i=1}^{n} \gamma_i x_i + \zeta$ in which, γ_i is a coefficient capturing the effect of indicator x_i on the latent variable η, and ζ is a disturbance term.
Graphical Representation		
Source of Variance	The latent variable ξ represents the common cause shared by a set of indicators.	The latent variable η represents a combined variance supplied by a set of indicators, including the interactions among them.
Measurement errors:	Measurement error is assumed for each indicator. The measurement error is fully independent, i.e., cov(ε_i, ξ)=0, and cov(ε_i, ε_j)=0 for $i \neq j$.	No measurement errors. In other words, all indicators are assumed to be accurate measures of η.
Relationships among indicators	All indicators (including potential measures) equally reflect the value of the underlying construct ξ after controlling measurement errors. Dropping or adding an indicator does not affect the value of ξ.	Each indicator represents a unique information source of the focal construct η. Dropping or adding an indicator will change the value of η (MacKenzie et al., 2005).
Relationships with other constructs in a structural model	The value of the construct is self-sufficiently explained by its indicators, therefore largely independent from the structural model. Correlations between the construct's indicators and other exogenous variables (i.e., cross-effect-relationships) should not exceed the correlations between indicators and their designated variable.	The value of the construct is contingent on its relationships with other constructs, especially the outcome variables, in the structural model. Correlations between the construct's indicators and other exogenous variables are freely estimated.

QUESTION THE USE OF ITEM WEIGHTS IN FORMATIVE MEASUREMENT

When developing a formative instrument, each item should be examined regarding its contribution to the focal construct: items that provide significant contribution should be retained and items with trivial influence to the focal construct should be dropped (Loch et al., 2003). A common practice in the examination is the use of item weight. Significance of a weight suggests a substantial contribution while insignificance suggests a negligible contribution from the investigated item (Marakas et al., 2007). Often, a p-value of 0.05 is employed as the threshold for such examination. As articulated in Urbach and Ahlemann (2010), "a significance level of at least .050 suggests that an indicator is relevant for the construction of the formative index and, thus, demonstrates a sufficient level of validity" (p. 20).

However, the suggestion in fact violates the theoretical origin of formative measurement. In the mathematical model of formative measurement (Table 2), the construct is formed as a function of its measures plus error (Edwards and Bagozzi, 2000); there is no assumption regarding internal relationships among the measures. In fact, the correlations among the measures should be freely estimated in a SEM test. As such, if items are highly correlated with one another, the significance of an item weight may not truly reflect the contribution of the item.

The argument raises the question of indicator collinearity (large amount of shared variances among indicators, or measurement items), which is common in formative measures (Diamantopoulos and Winklhofer, 2001). With the presence of collinearity, the influence of one item on its designated construct cannot be separated from the influences of the other indicators in the formative measurement model. Indicator collinearity in formative measurement has the same statistical properties with multicollinearity in multiple regression, an issue that does not reduce the predictive power or reliability of the model as a whole but distorts the coefficient estimates of predictors in an erratic way. The presence of indicator collinearity does not violate the statistical assumptions of formative measurement (e.g., the correlations among indicators are freely estimated in the formative measurement model), however, the estimated item weights will not reflect the true unique contributions of indicators. As such, these item weights should not be associated with meanings.

To further assess the use of item weights in the validity test of formative measurement, a Monte Carlo simulation analysis is designed.

MONTE CARLO SIMULATION ANALYSIS

Design

In the simulation analysis, data were generated to conform to an underlying population model where the predictor of X has a predefined impact, i.e., beta = 0.0, 0.1, 0.3, 0.5, on the dependent variable of Y. X was designed as a formative construct that was measured with five items; the contributions (i.e., item weights) of the five items are also predefined with certain weights.

Each factor is randomly generated with certain properties of (1) being an integer within the range of 0-10, (2) presenting the property of normal distribution, (3) having a predefined relationship with other factors (the correlation matrix are predefined; four patterns of correlations are designed, including correlation coefficient $r = 0, 0.1, 0.3, 0.5$ among the five factors).

More specifically, the data are randomly generated with the following properties:

- Formative measurement of X:

  X = 0.1 * Item 1 + 0.15 * Item 2 + 0.2 * Item 3 + 0.25 * Item 4 + 0.3 * Item 5

- Correlation matrix for the five items is defined as:

	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1				
Item 2	*r*	1			
Item 3	*r*	*r*	1		
Item 4	*r*	*r*	*r*	1	
Item 5	*r*	*r*	*r*	*r*	1

  Where $r = 0, 0.1, 0.3, 0.5$

- Structural model:

  $Y = \beta X + \text{error}$, where $\beta = 0, 0.1, 0.3, 0.5$

Results

In the IS empirical research, PLS and LISREL are the two most popular SEM techniques. The two techniques often generate very similar, if not identical, results (Gefen et al., 2000). For the study, PLS was selected as the statistical tool for the test of the simulated data. The decision was made because of a unique feature of PLS that allows the inclusion of a variable to "stand alone" in the model without any specified relationship to other variables. In this study, the true value of X (calculated by the aforementioned formula) was included in the test, so that its relationship with the estimate (hereinafter referred to as X') could be assessed.

The data were generated with Excel. Each data set had 1000 data points. PLS-Graph 3.0 was used to test the simulated data. To get reliable results, the test were repeated about 20 times, each time with a differently simulated sample data. The results of the simulation are summarized in Table 3 and 4: Table 3 reports the pattern of simulated tests that have been conducted; Table 4 reports the results regarding the item weights and their associated significance (T-test values) for the simulated formative measures.

Table 3. Simulation Patterns

True β	Average Estimate of β	Average T-tests of the β	r	Average of the Simulated r	Counts of Simulation Tests
0	-0.01	1.65	0	0.00	21
0	-0.03	1.45	0.1	0.08	19
0	0.02	1.77	0.3	0.25	20
0	-0.01	1.41	0.5	0.40	20
0.1	0.14	3.39	0	0.00	21
0.1	0.14	3.13	0.1	0.08	19
0.1	0.14	3.22	0.3	0.25	20
0.1	0.12	2.48	0.5	0.40	20
0.3	0.31	7.93	0	0.00	21
0.3	0.31	7.95	0.1	0.08	19
0.3	0.30	7.95	0.3	0.25	20
0.3	0.30	7.78	0.5	0.40	20
0.5	0.48	14.43	0	0.00	21
0.5	0.49	15.11	0.1	0.08	19
0.5	0.48	14.47	0.3	0.25	20
0.5	0.49	14.30	0.5	0.40	20
Total simulation tests					320

One should note that for a true $\beta = 0$, which suggests the nonexistence of relationship between X and Y, all simulation tests have concluded β with insignificance (i.e., the associated T-values are less than 1.96) regardless of the correlation patterns among the five formative measures; for other βs, the simulation tests have concluded significance on the relationship of X→Y; such significance is not affected by the internal relationships (i.e., *r*'s) among the five formative items.

Table 4. Results of Simulation Tests on Item Weights

Simulation Pattern		Average Item Weights					Average T-tests of Item Weights					*R* (X - X')
β	r	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	
0	0	0.30	0.15	0.16	-0.13	0.05	1.20	1.20	1.27	1.11	1.14	0.14
0	0.1	0.23	0.10	0.09	0.23	-0.05	1.13	1.33	1.12	1.03	0.97	0.28
0	0.3	0.13	0.25	-0.03	-0.05	-0.01	1.65	0.94	1.10	1.56	1.28	0.15
0	0.5	0.09	0.13	-0.14	0.01	0.14	1.01	1.21	1.31	1.02	1.47	0.17
0.1	0	0.15	0.31	0.30	0.37	0.47	1.14	1.48	1.51	1.52	2.24	0.73
0.1	0.1	0.03	0.20	0.22	0.43	0.40	0.88	1.26	1.18	1.73	1.73	0.71
0.1	0.3	0.22	0.18	0.27	0.30	0.19	1.22	1.19	1.18	1.31	1.11	0.73
0.1	0.5	-0.01	0.23	0.19	0.23	0.30	0.96	1.11	0.80	0.94	1.27	0.72
0.3	0	0.18	0.29	0.42	0.47	0.64	1.30	2.32	3.50	3.91	6.04	0.93
0.3	0.1	0.18	0.20	0.38	0.49	0.46	1.43	1.70	2.92	4.20	3.79	0.92

0.3	0.3	0.11	0.25	0.33	0.31	0.42	1.03	1.68	2.28	2.16	3.02	0.93
0.3	0.5	0.14	0.10	0.25	0.34	0.42	1.07	0.88	1.59	2.11	2.57	0.95
0.5	0	0.20	0.30	0.44	0.50	0.63	2.53	3.80	5.96	6.89	9.69	0.96
0.5	0.1	0.19	0.29	0.34	0.44	0.54	2.36	3.76	4.37	6.02	7.41	0.96
0.5	0.3	0.13	0.23	0.30	0.40	0.41	1.51	2.50	3.41	4.60	4.93	0.97
0.5	0.5	0.12	0.22	0.26	0.31	0.37	1.25	2.25	2.57	3.11	3.84	0.97

Note:

1. F1-5 are the five formative items formulated for X.
2. The actual weights assigned to the five items are 0.1, 0.15, 0.2, 0.25, and 0.3 respectively.

The results clearly demonstrate that item weights and their associated significance (i.e., T-values) significantly deviate from the designed formation of the construct of X. If only significant weights are included in the measurement, as marked by the shaded area with T-values > 1.96, many items will be dropped and the resulting instrument may not reflect the true value of X.

The correlation between X (the true value of the formative construct) and X' (the estimated value of the construct) can be used as an indicator of the quality of the measurement. Close examination of the correlation (R(X - X') in the Table) suggests that the correlation increases to a very high level (i.e., R>0.92) when X serves as a strong predictor of Y (β = 0.3 and 0.5 in the simulation tests). The pattern suggests that for the application of formative measurement, the theoretical relationship of the formative construct X to other variables (i.e., the hypothesized effects) plays a crucial role in achieving quality assessment of the focal construct. The stronger the theoretical relationship, more likely will the formative measurement reveal the true value of the underlying construct.

CONCLUSION

This study questions a common practice of using item weights for construct validity tests in the application of formative measurement. The practice does not confirm to the theoretical formation of formative constructs. A Monte Carlo simulation analysis is conducted to examine the practice. The results clearly demonstrate that item weights do not reflect the true design of the focal formative construct; using item weights may mislead the development and the application of formative instruments in empirical research.

REFERENCES

1. Bollen, K.A. (1989) *Structural Equations with Latent Variables*, New York: Wiley.
2. Bollen, K., and Lennox, R. (1991) Conventional wisdom on measurement: A structural equation perspective, *Psychological Methods*, 110, 2, 305-314.
3. Byrne, B.M. (1998) *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming*, Mahwah, New Jersey: Lawrence Erlbaum Associates.
4. Chin, W.W. (1998) Issues and opinion on structural equation modeling, *MIS Quarterly*, 22, 1, vii-xvi.
5. Diamantopoulos, A. (2006) The error term in formative measurement models: interpretations and modelling implications, *Journal of Modelling in Management*, 1, 1, 7–17.
6. Diamantopoulos, A., and Winklhofer, H.M. (2001) Index construction with formative indicators, *Journal of Marketing Research*, 38, 269-277.
7. Diamantopoulos, A., Riefler, P., and Roth, K.P. (2008) Advancing formative measurement models, *Journal of Business Research*, 61, 12, 1203-1218.
8. Edwards, J.R., and Bagozzi, R.P. (2000) On the nature and direction of relationships between constructs and measures, *Psychological Methods*, 5, 2, 155-174.
9. Gefen, D., Rigdon, E.E., and Straub, D. (2011) Editor's comments: An update and extension to SEM guidelines for administrative and social science research, *MIS Quarterly*, 35, 2, iii-xiv.
10. Gefen, D., Straub, D.W., and Boudreau, M.C. (2000) structural equation modeling and regression: Guidelines for research practice, Communications of the Association for Information Systems, 4, Article 7, 1-77.

11. Jarvis, C.B., MacKenzie, S. B., and Podsakoff, P. (2003) A critical review of construct indicators and measurement model mis-specification in marketing and consumer research, *Journal of Consumer Research*, 30, 2, 199-218.
12. Howell, R.D., Breivik, E., and Wilcox, J.B. (2007) Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007), *Psychological Methods*, 12, 2, 238-245.
13. Kim, G., Shin, B., and Grover, V. (2010) Investigating two contradictory views of formative measurement in information systems research, *MIS Quarterly*, 34, 2, 345-365.
14. Loch, K.D., Straub, D.W. and Kamel, S. (2003) Diffusing the Internet in the Arab world: The role of social norms and technological culturation, IEEE Transactions on Engineering Management, 50, 1, 45-63.
15. Marakas, G.M., Johnson, R.D., and Clay P.F. (2007) The evolving nature of the computer self-efficacy construct: An empirical investigation of measurement construction, validity, reliability and stability over time, *Journal of the Association for Information Systems*, 8, 1, 15–46.
16. Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., and Podsakoff, N.P. (2003) Common method biases in behavioral research: A critical review of the literature and recommended remedies, *Journal of Applied Psychology*, 88, 5, 879-903.
17. Straub, D.W. (1989) Validating instruments in MIS research, *MIS Quarterly*, 13, 2, 147-169.
18. Urbach, N., Ahlemann, F. (2010) Structural equation modeling in information systems research using partial least squares, *Journal of Information Technology Theory and Application*, 11, 2, 5-40.