5-1-2015

Convergent and Discriminant Validity with Formative Measurement: A Mediator Perspective

Xuequn Wang
Murdoch University, Perth, AUS, xuequnwang1600@gmail.com

Brian F. French
Washington State University, frenchb@wsu.edu

Paul F. Clay
Fort Lewis College, pfclay@fortlewis.edu

Convergent and Discriminant Validity with Formative Measurement: A Mediator Perspective

Xuequn Wang
Murdoch University
Perth, Australia

Brian F. French
Washington State University
Pullman, WA

Paul F. Clay
Fort Lewis College
Durango, CO

The ability to validate formative measurement has increased in importance as it is used to develop and test theoretical models. A method is proposed to gather convergent and discriminant validity evidence of formative measurement. Survey data is used to test the proposed method.

Keywords: Causal indicators, formative measurement, construct validity, convergent validity, discriminant validity, mediator

Introduction

There has been a vigorous debate and discussion about the issues surrounding the application of formative measurement (Bollen, 2007; Howell et al., 2007a, 2007b; Petter et al., 2007) and how to validate this specific kind of measurement model (Hardin et al. 2011). Because procedures used to validate reflective measurement are not appropriate for formative measurement, there is a need to develop measurement theory to validate formative measurement (Hardin et al., 2011).

Formative measurement has been applied in multiple disciplines, including Marketing (e.g., Chandon et al., 2000), Entrepreneurship (e.g., Brettel et al., 2011), and Information Systems (IS) (e.g., Pavlou & Gefen, 2005). For example, Pavlou and Gefen (2005) measured perceived effectiveness of institutional structures with formative measurement, which included four dimensions: feedback technologies, escrow services, credit card guarantees and trust in intermediary.

Although some researchers question the appropriateness of such models (e.g., Edwards, 2011), others have shown that formative measurement can be appropriate in certain contexts. For example, for multidimensional constructs, causal indicators can be developed to "comprise all essential aspects of the focal construct's definition" (MacKenzie et al., 2011, p. 304).

Using only global reflective indicators may, however, "diminish the correspondence between the empirical meaning of the construct and its nominal meaning, because there is no way to know whether the respondent is considering all of the subdimensions (facets) of the focal construct that are part of the nominal definition when responding to the global question" (MacKenzie et al., 2011, p. 327). Therefore, though there remain several issues related to the adoption of formative measurement, given that formative measurement can be appropriate in many contexts (Cadogan & Lee, 2013; Diamantopoulos et al., 2008; Jarvis et al., 2003; MacKenzie et al., 2011), developing corresponding methods is necessary so that researchers can validate formative measurement.

There are multiple aspects of construct validity that require evaluation using various methods to develop and maintain a strong validity argument. Having such evidence does not and cannot rely on a single method. According to Messick (1995), there are six aspects of construct validity: content, substantive, structural, generalizability, external, and consequential aspects of construct validity. In this paper, external aspect of validity evidence is focused upon, which deals with "convergent and discriminant evidence" (Messick, 1995, p. 745). More recently, Cizek et al. (2008) examined various aspects of validity from previously published indicators. They discussed validity including the traditional division of construct validity evidence (convergent and discriminant evidence), criterion-related evidence, content evidence, evidence based on response process, evidence based on consequences, face validity evidence and evidence based on internal structure, supporting the need for various forms of evidence. In this study associations with other variables (convergent and discriminant evidence) rather than all possible sources of validity evidence is focused on. Note that this is only one step toward developing a comprehensive validity argument to support inferences from formative measurement.

Previous studies have paid little attention to convergent and discriminant validity of formative measurement (Bollen, 2011). This may be attributed to the fact that formative measurement is quite different from reflective measurement. Although there are relatively mature and sophisticated methods to gather convergent and discriminant validity evidence for reflective measurement based on classical test theory (CTT) (Kane, 2006), there lacks an agreed method or set

of procedures to gather convergent and discriminant validity evidence for formative measurement (Barki et al., 2007; Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003; Petter et al., 2007). Thus, a researcher and practitioner can often faces difficulty in dealing with convergent and discriminant validity when one moves from reflective measurement to formative measurement (Diamantopoulos et al., 2008).

In this study, constructs are used to refer to "a conceptual term used to describe a phenomenon of theoretical interest" (Edwards & Bagozzi, 2000, p. 156-157), and latent variable is used to refer to the representation of a certain construct in a model. Indicators are used to refer to "observed variables that measure a latent variable" (Bollen, 2011, p. 360). The kind of indicators depends on "whether the indicator is influenced by the latent variable or vice versa" (Bollen, 2011, p.360). Reflective indicators are used to refer to those influenced by the latent variable, and causal indicators are used to refer to those influencing the latent variable.

The focus in this study is on formative measurement with causal indicators. As Bollen (2011) illustrated, formative measurement may include causal indicators or formative indicators. The key difference between these two types of indicators is that "causal indicators should have conceptual unity in that all the variables should correspond to the definition of the concept whereas formative indicators are largely variables that define a convenient composite variable where conceptual unity is not a requirement" (Bollen, 2011, p. 360). Variables consisting of formative indicators may not have any meaningful conceptualization. Therefore, formative measurement with causal indicators is focused upon in this study (Bollen, 2011).

Although formative measurement have been recognized in the literature (Diamantopoulos et al., 2008); there are no agreed upon methods to provide convergent and discriminant validity evidence for formative measurement. Because construct validity is "a necessary condition for theory development and testing" (Jarvis et al., 2003, p. 199), it is important to gain validity evidence before one tests theory. This paper adds to the current validity literature by proposing and testing a method to gain validity evidence (convergent and discriminant evidence) for formative measurement. Note that the proposed method does not aim to challenge or replace CTT when testing reflective measurement. After testing our method with real data for formative measurement, construct validity for reflective measurement is also examined following our new method. The results from our method and those from Confirmatory Factor Analysis (CFA) are consistent.

Reflective vs. Formative Measurement

A. Reflective Measurement   **B.** Formative Measurement
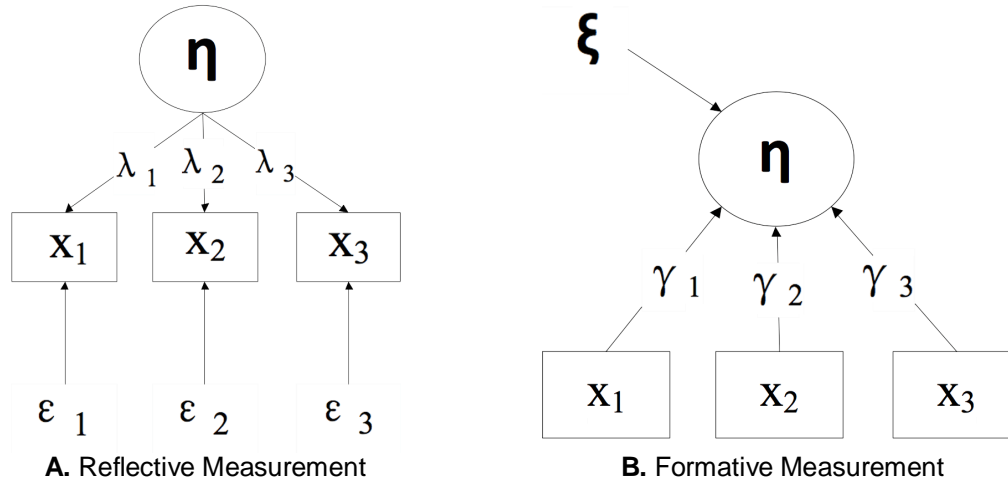
Figure 1. Two kinds of measurement models.

Many measurement models that social science deals with are reflective (Panel A from Figure 1; Diamantopoulos et al., 2008, Petter et al., 2007). For reflective measurement, the direction of causality is from the latent variable to the indicators. Because all indicators are the effects of the same latent variable, they are expected to be highly correlated (internal consistency reliability) (Bollen, 1984). The deletion of an indicator will probably not alter the meaning of the latent variable given that there are sufficient and similar functioning indicators to represent the latent variable. Ideally the indicators are interchangeable. Measurement errors are taken into account at the indicator level (c.f. Edwards and Bagozzi (2000), Jarvis et al. (2003), MacKenzie et al. (2005), for a more detailed description). Thus, the equation for a measurement model with reflective indicators is given as (Bollen & Lennox, 1991):

$$x_i = \lambda_i \eta + \varepsilon_i \tag{1}$$

where η is the latent variable, x_i is the ith reflective indicator for the latent variable η, λ_i represents the effect of η on that indicator (coefficient) and ε_i is the measurement error for x_i.

In contrast, for formative measurement the latent variable is influenced by these causal indicators (Bollen, 1984; Chin, 1998). Thus, deleting an indicator will alter the meaning of the latent variable (Bagozzi, 2007; Bollen, 2007; Diamantopoulos et al., 2008; Howell et al., 2007b; Jarvis et al., 2003). Additionally, there is no reason to expect that these causal indicators are necessarily highly correlated with each other, which makes internal consistency reliability inappropriate. Unlike reflective indicators, causal indicators are assumed to be error free (c.f. Edwards and Bagozzi (2000), Jarvis et al. (2003), and MacKenzie et al. (2005)) and that there may be a disturbance term representing "non-modeled causes" (Diamantopoulos, 2006, p. 7). Thus, the equation for a measurement model with causal indicators is (Bollen & Lennox, 1991):

$$\eta = \gamma_1 x_1 + \ldots + \gamma_i x_i + \zeta \qquad (2)$$

where η represents the latent variable, x_i is the ith causal indicator for latent variable η, γ_i represents the path weights for indicators x_i and ζ is the disturbance term which includes other variance not accounted for by the indicators (MacKenzie et al., 2005). For example, job satisfaction can be measured with indicators such as "I am very satisfied with my pay", "I am very satisfied with the nature of my work", and "I am very satisfied with my opportunities for promotion", and so on, and these three indicators influences one's job satisfaction level (MacKenzie et al., 2011). Because the covariance between causal indicators could be any value, the way to examine construct validity (convergent validity and discriminant validity) for reflective measurement based on CTT (e.g., CFA) cannot be used. Therefore, a new method is required to validate formative measurement.

For reflective measurement, convergent evidence is provided when "different indicators of theoretically similar or overlapping constructs are strongly interrelated" (Brown, 2006, p. 2), and discriminant evidence is provided when "indicators of theoretically distinct constructs are not highly intercorrelated" (Brown, 2006, p. 3). In other words, convergent validity essentially refers to whether indicators from a latent variable do belong to that latent variable, and discriminant validity essentially refers to whether indicators from a latent variable do not belong to other latent variables.

However, for formative measurement, high correlations are *not* required between its indicators (Jarvis et al., 2003). Furthermore, correlations among causal indicators within a measurement model need not be higher compared to correlations between them and indicators from other measurement models (Bollen, 2011; Bollen & Lennox, 1991). Therefore, the traditional approach toward establishing convergent and discriminate validity from CTT is not appropriate. In this study, an adaptation of the definition of convergent and discriminant validity is proposed to accommodate the context of formative measurement. Convergent validity is used to specify that causal indicators from a measurement model should explain a significant proportion of variance from the latent variable that they measure; discriminant validity is used to specify that these same indicators should explain a much lower proportion of variance from other latent variables. That is, indicators that are associated with the target latent variable will explain much more variance of that latent variable and those indicators should not explain a large amount of variance of other latent variables relative to the target latent variable.

These definitions adapt Brown (2006)'s definition by reversing the direction of relationship between the latent variable and the indicators. Discriminant evidence is particularly important because it indicates that these indicators do not belong to other latent variables.

The Context of Validation

Identification is always an issue for structural equation models with latent variables, and there are two general identification rules: First, each latent variable must be assigned a scale; Second, the number of free parameters estimated in a model must be no more than the number of unique pieces of information in the covariance matrix of manifest variables (Bollen & Davis, 2009). Thus, for a reflective measurement model, the minimum number of indicators should be at least three. However, there is one more identification requirement raised by formative measurement. MacCallum and Browne (1993) showed that an additional requirement for the identification of the disturbance from formative measurement was that the latent variable measured by causal indicators must emit two paths to its reflective indicators or other latent variables. Therefore, a model is proposed in which the latent variable measured by causal indicators predicts two or more outcome variables measured by reflective indicators as the context in which to gather convergent and discriminant validity evidence (Bollen & Davis,

2009). Our model is consistent with the circumstances identified by Bagozzi (2011) under which formative measurement are appropriate to be used.

The example model proposed is shown in Figure 2, where latent variable η_1 is measured by causal indicators and its convergent and discriminant validity evidence is to be examined. Note that the actual research model may be different from this test model: The model is used to gather convergent and discriminant validity evidence only; and its structural paths may differ widely from those of the research model. What the model is trying to do is to examine the indicators from latent variable η_1 in terms of convergent and discriminant validity.

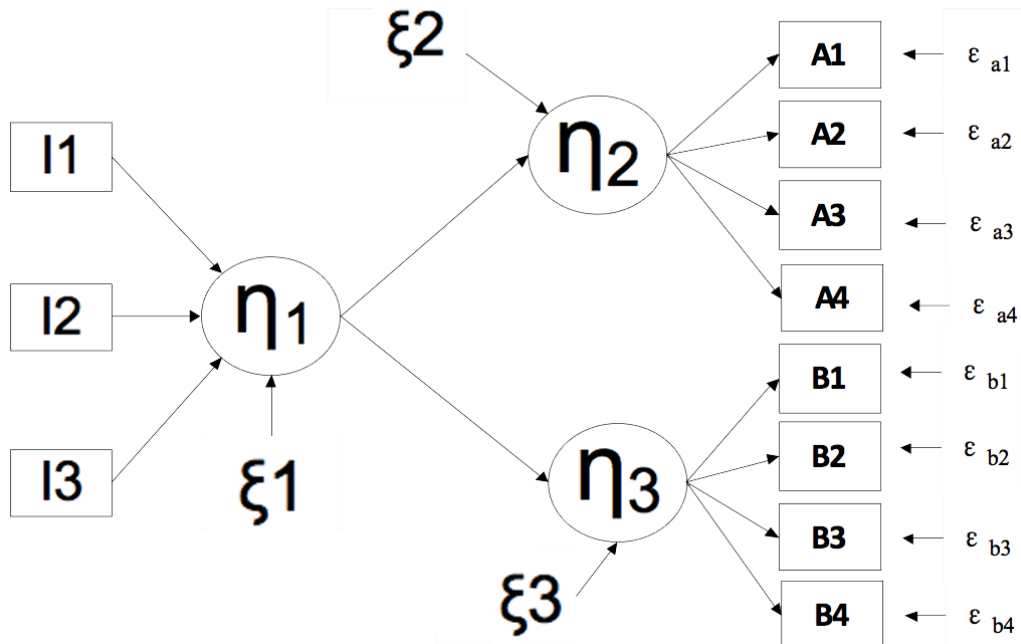**Figure 2.** An example model of the proposed method.

A Mediator Perspective

Psychologists have recognized the concept of a mediator for quite a long time (e.g., Woodworth, 1928). Furthermore, Baron and Kenney (1986) clarified the nature of a mediator: a given variable functioned as a mediator if it accounted for the relationship between an independent variable and a dependent variable. To be

a mediator, a variable needs to meet three conditions: (a) Variance of independent variable *A* significantly accounts for variance of mediator *B*. In other words, the path coefficient of Path A is significant. (b) Variance of mediator *B* significantly accounts for variance of the dependent variable *C*. In other words, the path coefficient of Path B is significant. (c) When Paths A and B are controlled, the previous significant relation (Path C) between the independent variable *A* and dependent variable *B* significantly decreases (or even becomes zero).

By applying the mediator perspective, the relevant latent variable η_1 can be seen as a mediator which accounts the influence of causal indicators I1-I3 on the other latent variables (e.g., η_2; Panel A from Figure 3) (Bollen, 2007; Bollen & Davis, 2009; Howell et al., 2007b). Then, latent variable η_1's construct validity (i.e., convergent and discriminant evidence) can be examined. Note that our method is justified based on previous literature. Bollen (2007), for example, argued that the latent variables measured by causal indicators mediated "the effect of causal indicators on these other variables" (p. 222). MacKenzie et al. (2011) also argued that "the adequacy of the hypothesized multidimensional structure can be assessed by testing whether the sub-dimensions of the multidimensional focal construct have significant direct effects on a consequence construct, over and above the direct effect that the focal construct has on the consequence" (p. 323). Specifically, the causal indicators "*must* share the latent variable η as a common consequence and, moreover, η must fully mediate the effects of" their indicators "on other observed or latent variables that are modeled as outcomes of η" (Diamantopoulos, 2011, p. 340). Also as Franke et al. (2008, p. 1230) argued, the latent variables measured by causal indicators "mediate the effects of their indicators on other variables, constraining their indicators to have the same proportional influence on the outcome variables….If the formative indicators could have direct as well as mediated effects on the outcome variables, then the proportionality constraint would not necessarily hold". (Here formative indicators refer to causal indicators in Bollen (2011)'s terminology.)

In the proposed method, the validity of formative measurement is supported even if causal indicators have direct influence on the outcomes variables, as long as "the magnitude of the effect of the focal construct on the consequence construct is substantially larger than the combined magnitudes of the direct effects" of its indicators on the outcome variables (MacKenzie et al., 2011, p. 323). In other words, the latent variable can fully or partially mediate the influence of causal indicators I1-I3 on latent variable η_2. It is similar to the context in which the research model only contains reflective measurement and construct validity is

supported even if cross-loadings exist as long as these cross-loadings are much less then loadings between reflective indicators and the focal latent variables.

Therefore, to gather η_1's convergent evidence, if indicator I1 indeed belongs to η_1, the influence of I1 on η_2 should be mediated by η_1 (Panel A from Figure 3). In other words, I1 should explain a significant amount of variance of η_1. That is consistent with the definition of formative measurement: Indicator I1 influences η_1, and then η_1 influences η_2. Following Baron and Kenny's instruction, we can examine convergent validity in three steps. See Table 1 for each step. Especially, significant indicator weight is the first step. If indicator weights (Path A) are not significant, there is no need to go further, given that the strength of indicator weight is the statistical metric used to judge indicator retention (Bollen & Lennox, 1991; Chin, 1998; Diamantopoulos et al., 2008; Diamantopoulos & Winklhofer, 2001).
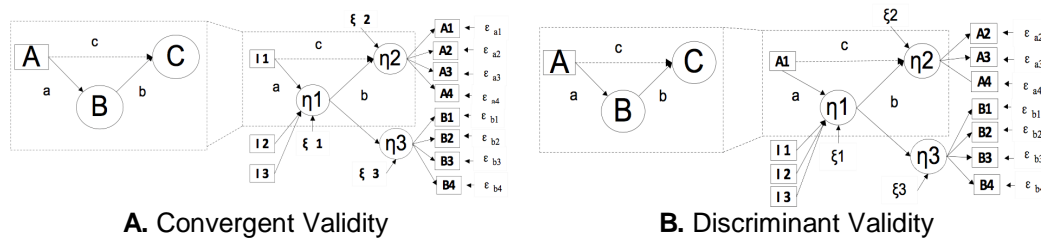

A. Convergent Validity   **B.** Discriminant Validity

Figure 3. A mediator perspective.

Table 1. A mediator perspective to gather validity evidence for formative measurement.

Step	Description
Step 1	Examine if path coefficient for Path A is significant • If path coefficient for Path A is not significant, then I1 does not significantly cause η_1. There is no need to go further. • If path coefficient for Path A is significant, then
Step 2	Examine the coefficient for Path C (without controlling B) • If path coefficient for Path C is not significant, then I1 and η_2 do not share a significant amount of variance. There is no need to go further. • If path coefficient for Path C is significant, then
Step 3	Examine the coefficient for Path C by controlling A and B • If path coefficient for Path C becomes less or insignificant, then η_1 mediates the influence of I1 on η_2. Therefore I1 probably belongs to η_1. • If path coefficient for Path C remains the same or changes little, then η_1 does not mediate the influence of I1 on η_2. Therefore I1 may not belong to Y_1.

To gather η_1's discriminant evidence, the same process is gone through by examining if η_1 mediates indicators from other measurement models. For example, indicators A1-A4 from latent variable η_2 can be examined and confirmed that η_1 cannot mediate these indicators' influences on η_2 (Panel B from Figure 3). Indicators from η_2 should explain a much less amount of variance of η_1 than I1 - I3. The same process in Table 1 is followed. When path coefficient for Path C is tested controlling for Path A and Path B, if path coefficient for Path C does not change significantly, then the influences of indicator A1- A4 are not mediated by η_1. Therefore, indicators A1- A4 do not belong to η_1. In contrast, if the path coefficient for Path C reduces significantly or even becomes insignificant, A1- A4 may belong to η_1. Here content analysis is needed to further examine these indicators, and indicators A1- A4 are problematic in the sense that the results are not consistent with developed theory.

Methodology

Participants

Participants ($N = 337$) from an entry level business class at a large state university in the Northwest of the U.S. completed the scales described below. The demographic information collected includes age and gender. The mean age of the participants was 20.35, with the range between 18 and 36 years. The percentage of male students was 62.00%.

Measures

Perceived Effectiveness of Institutional Structures (PE) (Pavlou & Gefen, 2005), a correctly modeled formative measurement (Petter et al., 2007), was selected as our example of formative measurement. Two other constructs (Trust and Trust Propensity (TP), where Trust is Trust in the Community of Sellers, and TP is Trust Propensity). For a detailed description of PE, Trust and TP and their indicators, please refer to Pavlou and Gefen (2005).) were chosen to form the model to test in Figure 2. The instruments from original studies were adapted to fit the new study environment. The indicators of PE and Trust were reworded to focus on online shopping behaviors.

Procedures

Participants were given class credit to participate in the study (less than 1% of their final grade) with other options if they selected not to participate. Data collection occurred in laboratories for the business class. After participants arrived in the laboratories, the administrator read aloud the purpose and procedures for the study. Then participants accessed a website to complete the questionnaire. The questionnaire contained a randomized sequence of indicators from PE, Trust, TP and other constructs from Pavlou and Gefen (2005) as well as demographic information questions. Once the questionnaire was completed (about 10 mins), participants were thanked and exited the laboratory.

Data Analysis

M*plus* (Muthén & Muthén, 1998-2012) was used to analyze the data. Our analysis had two components. First, our proposed method was tested with the model including PE, Trust and TP. Second, the proposed method was applied to gain convergent and discriminant evidence for Trust, to show that the proposed method is consistent with CTT when examining measurement models with reflective indicators.

For the first component of the analysis, CFA was first performed to gather the convergent and discriminant evidence of the two latent variables measured by reflective indicators: Trust and TP (Brown, 2006). The global fit was assessed and the following fit indices were used: chi-square statistic (χ^2), Comparative Fit Index (*CFI*), and the Standardized Root Mean Squared Residual (*SRMR*). The χ^2 test is significant when *p* value is less than 0.05. In such contexts, the model may not represent data reasonably well. *CFI* equal to or greater than .90 indicates reasonable global fit (Rigdon, 1996). The *SRMR* less than .05 indicates acceptable fit (Byrne, 1998). Because the result of chi-square test is likely inflated by sample size, the result of χ^2 test is routinely significant with large sample size, even if the differences between *S* and Σ are negligible (Brown, 2006). Therefore, other fit indices were used in combination with the chi-square test. Standardized loadings were then used to gather the convergent evidence and cross loadings were used to gather the discriminant evidence. For the size of item loadings, suggestions given by Straub et al. (2004) were followed, who suggest that loadings should be "above .707 so that over half of the variance is captured by the latent construct" (p. 410).

Next the model including PE, TP and Trust was examined to gather convergent and discriminant validity evidence for PE, which is measured by

causal indicators. The global fit of the model was first examined. Here acceptable overall goodness of model fit is important to show that the baseline model can fit the data well (Brown, 2006). The convergent and discriminant validity evidence for PE was then gathered following the method proposed above (refer to Table 1).

For convergent evidence, proposed indicators for PE should converge on PE. From a mediator perspective, PE should mediate the influence of its indicators on the other two latent variables (Figure 4). For discriminant evidence, indicators from other measurement models should not belong to PE. From a mediator perspective, PE should not mediate the influence of indicators from other latent variables on these two latent variables.
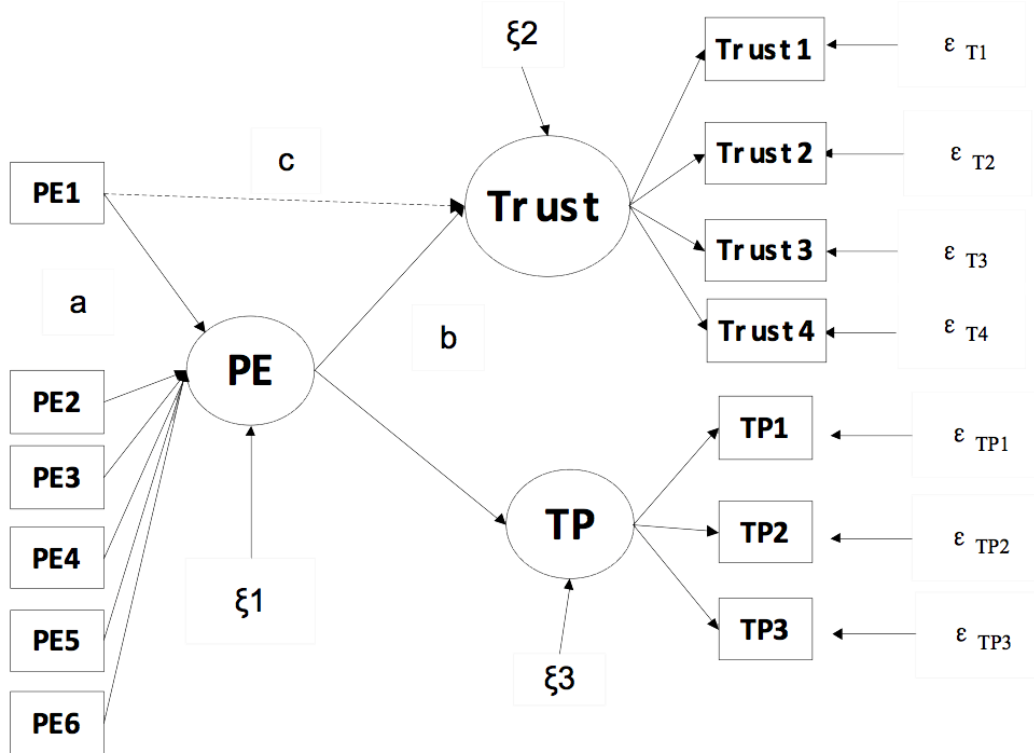

Figure 4. Model to gather convergent and discriminant evidence for PE.

In the second component of the analysis, the convergent and discriminant validity evidence of Trust were gathered with the method proposed in this study.

These analyses demonstrated that our proposed method was consistent with CTT when gathering convergent and discriminant evidence from reflective measurement as well. First convergent validity of Trust was examined to check if Trust1-Trust4 belonged to Trust (Figure 5). Next discriminant validity was examined to check if TP1-TP3 belonged to Trust.
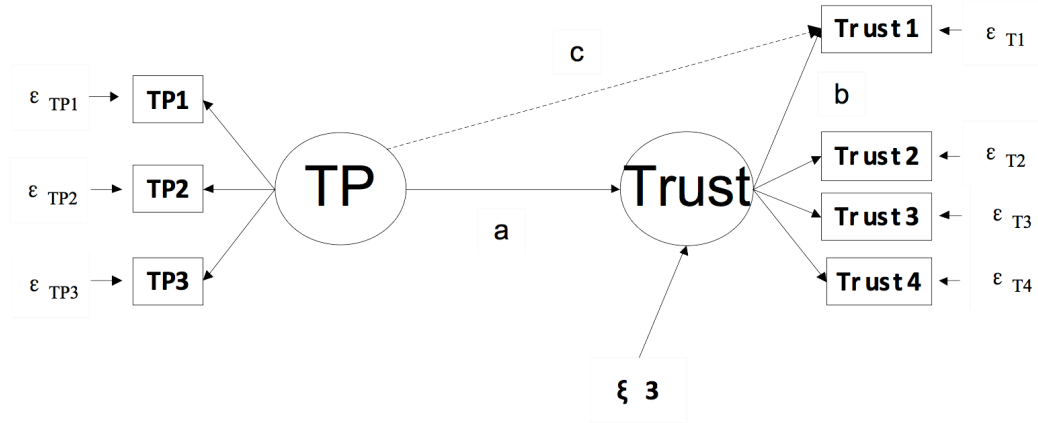

Figure 5. A mediator method to gather convergent and discriminant evidence for trust.

Results

CFA

The global fit of the model was acceptable ($\chi^2(13) = 85.779$, $NC = 6.60$, $p < 0.0001$, $CFI = 0.943$, $SRMR$ is 0.040). Although the result of χ^2 test was significant, it was largely due to the large sample size (337). Other fit indices met stated criteria.

For convergent evidence, indicators' standardized loadings were examined. The standardized loadings for all indicators are shown in Table 2: all loadings were significant and most loadings were above 0.707 (except for Trust2 and TP2), which indicates that the latent variables explain more than 50% of variance for most indicators. This indicated reasonable convergent evidence. For discriminant evidence, the cross loadings between indicators and other latent variables were examined, requiring that indicators load much higher on the latent variables they measure than on other latent variables (Gefen & Straub, 2005). From the results of Modification Indices (M.I.), no M.I.s for cross loading are significant,

indicating good discriminant evidence. (In M*plus*, M.I. is the amount chi-square which would drop if the parameter is estimated as part of the model. 3.84 is the chi-square value which is significant at the .05 level for one degree of freedom. When the M.I. is significant, we also want to examine the size of completely standardized expected parameter change. Usually, values more than 0.300 are considered large and should be included in the model. Value less than 0.200 indicates a trivial change of parameter, and we may not include it into the model, even if M.I. is significant.) To summarize, Trust and TP have good convergent and discriminant evidence.

Table 2. Loadings.

	Trust		**TP**
Trust1	0.786	TP1	0.750
Trust2	0.687	TP2	0.595
Trust3	0.907	TP3	0.803
Trust4	0.928		

Construct Validity (Convergent and Discriminant Evidence): Formative Measurement

The fit for baseline model was first examined. The model met fit criteria ($\chi^2(48) = 145.439$, $p < 0.0001$, $NC = 3.03$, $CFI = .92$, SRMR is 0.039). Therefore, the global fit of baseline model was reasonable.

The method outlined in Table 1 was followed. For convergent validity, PE1-PE6 were considered as independent variable, PE as the mediator, and Trust (or TP) as the dependent variable. In the first model (Trust as the dependent variable, refer to Table 3), the path coefficient for Path A was first examined. According to the second column, the path coefficients from PE1 and PE6 to PE were significant, indicating that PE1 and PE6 significantly influenced PE in this context. Next, the path coefficient for Path C was examined, without controlling Path A. According to the forth column, path coefficients from PE1 and PE6 to Trust were significant, indicating that the PE1 and PE6 explained a significant amount of variation of Trust. Finally, the path coefficient for Path C was examined, controlling Path A and B. According to the third column in Table 3, the path coefficient for Path B (from PE to Trust) was significant. According to the last column, when controlling Path A and Path B, all path coefficients were insignificant, indicating that there were no direct effects from PE1 and PE6 to Trust. Therefore, PE fully

mediated the influence of PE1 and PE6 on Trust. In the second model (TP as the dependent variable, refer to Table 4), the same procedures were followed, and the results also indicated full mediation. Specially, path coefficients for Path C were not significant according to the forth column, indicating that PE1 and PE6 could not explain a significant amount of variance of TP even before controlling Path A and Path B. Therefore, PE1 and PE6 belonged to PE, indicating good convergent evidence.

Table 3. Path coefficient between PE, PE's indicators and Trust.

	Path A	Path B	Path C (before controlling Path A)	Path C (after controlling Path A)
PE1	0.239*	0.764*	0.148*	0.082
PE2	0.173	0.764*	-	0.098
PE3	0.142	0.764*	-	-0.131
PE4	0.046	0.764*	-	-0.136
PE5	-0.020	0.764*	-	0.007
PE6	0.355*	0.764*	0.163*	0.000

*Note: $p < 0.05$

Table 4. Path coefficient between PE, PE's indicators and TP.

	Path A	Path B	Path C (before controlling Path A)	Path C (after controlling Path A)
PE1	0.239*	0.629*	0.011	-0.069
PE2	0.173	0. 629*	-	-0.094
PE3	0.142	0. 629*	-	0.097
PE4	0.046	0. 629*	-	0.099
PE5	-0.020	0. 629*	-	-0.004
PE6	0.355*	0. 629*	0.091	0.001

*Note: $p < 0.05$

For discriminant validity, Trust1-Trust4 were considered as independent variable, PE as the mediator, and Trust as the dependent variable (refer to Table 5). First, the path coefficient for Path A was examined. According to the second column, path coefficients from Trust1-Trust4 to PE were significant, indicating that Trust1-Trust4 significantly influenced PE. Next, the path coefficient for Path

C was examined, without controlling Path A. According to forth column, Trust1-Trust4 significantly influenced Trust.

Finally, the path coefficient for Path C was examined, controlling Path A and Path B. According to third column, the path coefficient for Path B (from PE to Trust) was significant. According to the last column, path coefficient for Path C (from Trust1-Trust4 to Trust) was still significant and decreased little after controlling for Path B, indicating that PE did not mediate the influence of Trust1-Trust4 on Trust. Therefore, indicators Trust1-Trust4 did not belong to PE, and discriminant evidence was supported.

Table 5. Path coefficient between PE, Trust and Trust's indicators.

	Path A	Path B	Path C (before controlling Path A)	Path C (after controlling Path A)
Trust1	0.755*	0.967*	0.715*	0.636*
Trust2	0.633*	0. 931*	0.575*	0.445*
Trust3	0.867*	0. 964*	0.837*	0.620*
Trust4	0.883*	0. 985*	0.868*	0.678*

*Note: $p < 0.05$

Another evidence of discriminant validity was that after adding Trust1 (to Trust4) to PE, the path coefficient from PE to Trust was more than 0.900, indicating bad discriminant validity (Now PE and Trust cannot discriminate from each other). Therefore, to keep PE as a meaningful and separate latent variable, Trust1 (to Trust4) should be removed from PE. However, this argument should be based on the previous step in that PE could mediate several indicators' influence on Trust and TP. If PE could not function as mediator in previous steps, then indicators could be problematic.

Construct Validity (Convergent and Discriminant Evidence): Reflective Measurement

In this section the proposed method was applied to gather convergent and discriminant evidence of reflective measurement (Trust), to confirm that Trust1-Trust4 belonged to Trust and TP1-TP3 did not belong to Trust. To gather convergent evidence, TP was considered as the independent variable, Trust as the mediator and Trust1-Turst4 as the dependent variable (refer to Table 6).

Table 6. Path coefficient between Trust, Trust's indicators and TP.

	Path A	Path B	Path C (before controlling Path A)	Path C (after controlling Path A)
Trust1	0.473*	0.786*	0.375*	0.004
Trust2	0.473*	0.687*	0.321*	-0.006
Trust3	0.473*	0.907*	0.435*	0.012
Trust4	0.473*	0.928*	0.435*	-0.011

*Note: $p < 0.05$

The path coefficient for Path A was first examined. According to the second columns in Table 6, the path coefficients were significant and not more than 0.800, which indicated that TP explained a significant amount of variance of Trust, and TP and Trust were discriminant from each other. Next the path coefficient for Path C was examined, without controlling Path A. According to the forth column, path coefficients for Path C were significant, indicating that Trust1-Trust4 loaded on TP significantly. Finally, the path coefficient for Path C was examined, controlling Path A and Path B. According to the third column, path coefficients for Path B were significant and more than 0.707 (except for Trust2). According to the last column, all path coefficients for Path C were insignificant, which indicated that Trust fully mediated TP's effect on Trust1-Trust4. Therefore, good convergent evidence was supported.

To gather discriminant evidence, TP was considered as the independent variable, Trust as the mediator and TP1-TP3 as the dependent variable (refer to Table 7). The path coefficient for Path A was first examined. According to the second column, the path coefficient was significant and less than 0.800, indicating that TP explained a significant amount of variance from Trust, and they were discriminant from each other. Next, the path coefficients for Path C were examined, without controlling Path A. According to the forth column, path coefficients for Path C were all significant, indicating that TP1-TP3 loaded on TP significantly. Finally, the path coefficients for Path C was examined, controlling Path A and Path B. According to the third column, the path coefficients for Path B (from Trust to TP1-TP3) were significant. However, no path coefficients (loading) were more than 0.707. According to the last column, all path coefficients for Path C were significant and decreased little, indicating Trust could not mediate TP's effect on TP1-TP3. Therefore, TP1-TP3 did not belong to Trust. Thus, good discriminant evidence was supported.

Table 7. Path coefficient between Trust, TP and TP's indicators.

	Path A	Path B	Path C (before controlling Path A)	Path C (after controlling Path A)
TP1	0.437*	0.432*	0.750*	0.642*
TP2	0.500*	0.269*	0.595*	0.625*
TP3	0.525*	0.366*	0.803*	0.920*

*Note: $p < 0.05$

To summarize, our results showed that Trust1-Trust4 are indicators of Trust but TP1-TP3 were not. These conclusions are consistent with the results of CFA in the framework of CTT. Therefore, the method proposed is consistent with CTT when we gather convergent and discriminant evidence for reflective measurement.

Discussion

Formative measurement has been recognized in previous literature (Bollen, 1984; Bollen, 2011; Petter et al., 2007; Wang, Jessup, & Clay, 2015). However, there has not been an agreed method to gain convergent and discriminant validity evidence for formative measurement. The purpose of this study was to propose a method to gain convergent and discriminant evidence for formative measurement. A mediator perspective was adopted to propose a series of steps to test the validity of formative measurement. The data collected supports our method and showed that the method could keep those indicators which should belong to a formative measurement model and teasing out those which should not be part of the measurement. Our method can guide further social and behavioral research on how to gather convergent and discriminant validity evidence for formative measurement, and contribute a potential solution to one of the issues surrounding the application of formative measurement raised by recent literature (Edwards, 2011).

It is admitted that conclusions drawn from our method are dependent upon the data from a single example with one data set. In the results above that we showed that PE2, PE3, PE4 and PE5 did not significantly influence PE. Therefore, those four indicators may not belong to PE. However, the decision whether PE2, PE3, PE4 and PE5 are to be retained based on statistical results (convergent and discriminant validity) and other validity evidences (e.g., content validity) would

be necessary. Any scale refinement should be based on both empirical and theoretical information and not rely solely on empirical data. For formative measurement, indicator weights are dependent on specified structural models (Bollen &Davis, 2009), and the relative contribution of indicator weights is model dependent (Bollen et al., 2001; Hauser & Warren, 1997). Therefore, the choice should be based on "theoretical relevance" (Cenfetelli & Bassellier, 2009). If PE2, PE3, PE4 and PE5 represent unique and important domain of PE, they should be kept despite the fact that they do not significantly influence PE in this context with an eye in refining how they are assessed.

Because the procedures of measurement development and validation are quite complex, researchers may find that the focal latent variable cannot mediate the relationship between certain causal indicators and outcome variables. Consider the context with reflective measurement only. Even if researchers have followed strict procedures to develop indicators, it is still possible for several reflective indicators to have insufficient discriminant validity (e.g., cross-loadings are high) (MacKenzie et al., 2011). Based on previous discussions, cross-loadings for reflective indicators are similar to direct effects which cannot be mediated by the latent variable from a formative measurement model (Figure 4 and 5). When the latent variable measured with causal indicators cannot mediate the relationship between certain causal indicators and outcome variables, these corresponding indicators are problematic (Diamantopoulos, 2011; MacKenzie et al., 2011). Our method can detect these indicators and warn researchers that their measurement models are not be supported.

Limitation and Directions for Future Research

A few limitations should be recalled when applying the proposed method. First, the application of statistical testing is based on relevant literature (e.g., Bollen, 1989; Bollen & Lennox, 1991). As MacKenzie et al. (2011) argue, "indicator validity is captured by the significance and strength of the path from the indicator to composite latent construct" (p. 315). Bollen (2011) also argued that "a coefficient of a causal indicator with the wrong sign or that is not statistically significant would appear to be invalid and a candidate for exclusion" (p. 365). A significance test was relied on in the first stage of examining convergent and discriminant validity (Table 1). After the first stage, it is the difference of path coefficients between the second and the third stage that is important in supporting validity claims (Table 1). It is fully acknowledged that the exclusive focus on statistical significance ignores the problem that in large samples, effects that are

trivial in magnitude can be statistically significant. However, in smaller samples where power is too low to be effective, even appreciably large effects may not be statistically significant in smaller samples. Therefore, when researchers apply our method and are in the first stage of our method, they may also want to check the statistical power to ensure that there is adequate power to detect medium to large effects.

Second, because the residual from formative measurement can only be identified when there are at least two paths emitting from the formative measurement model, at least two other latent variables measured by reflective indicators are needed. This limitation is due to the underlying attribute of formative measurement. One potential way to solve that issue is to add a reflective indicator to that measurement model so that only one other latent variable is needed. In this context, the formative measurement model still emits two paths: one to its reflective indicator and one to another outcome latent variable. Note that our method is fully consistent with recent debate of the disturbance term for formative measurement (Cadogan & Lee, 2013). Specifically, Cadogan and Lee (2013) suggested that using formative latent variables (formative measurement with the disturbance term) should be suspended until researchers developed corresponding measurement theories; meanwhile, other alternatives could be used, such as formative composite variables (formative measurement without the disturbance term). Therefore, after gathering convergent and discriminant validity evidence for formative measurement, researchers should apply formative composite variables in their model testing. As discussed above, our model is just to validate formative measurement, not to test theories developed containing formative measurement.

Third, for our method, the number of indicators used in reflective measurement should be at least four. As discussed above, for reflective measurement, the minimum number of indicators should be at least three. However, if there are only three indicators in a reflective measurement model (like TP in the previous data), the number of indicators from that measurement model will become two when we move one indicator to the formative measurement model and test if the latent variable measured with causal indicators can mediate the effect from that indicator. With only two indicators a latent variable will be unidentifiable.

Fourth, the analysis employed indicators from previously published studies. There was no control over model fit, strength of relationship between variables, and so on. Even though this may reflect reality, future studies can employ Monte

Carlo techniques to further validate the proposed under a variety of conditions (e.g. degree of model misspecification, strength of loadings).

References

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*(2), 229-237. doi:10.1037/1082-989X.12.2.229

Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly, 35*(2), 261-292.

Barki, H., Titah, R., & Boffo, C. (2007). Information system use-related activity: An expanded behavioral conceptualization of individual-level information system use. *Information Systems Research, 18*(2), 173-192. doi:10.1287/isre.1070.0122

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182. doi:10.1037/0022-3514.51.6.1173

Bollen, K. (1984). Multiple indicators: internal consistency or no necessary relationship? *Quality and Quantity, 18*(4), 377-385. doi:10.1007/BF00227593

Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K. (2007). Interpretational confounding is due to misspecificaiton, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*(2), 219-228. doi:10.1037/1082-989X.12.2.219

Bollen, K. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly, 35*(2), 359-372.

Bollen, K., & Davis, W. R. (2009). Causal indicator models: identification, estimation, and testing. *Structural Equation Modeling, 16*(3), 498-522. doi:10.1080/10705510903008253

Bollen, K., Glanville, J., & Stecklov, G. (2001). Socioeconomic status and class in studies of fertility and health in developing countries. *Annual Review of Sociology, 27*, 153-185. doi:10.1146/annurev.soc.27.1.153

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Methods, 110*(2), 305-314. doi: 10.1037/0033-2909.110.2.305

Brettel, M., Engelen, A., Müller, T., & Schilke, O. (2011). Distribution channel choice of new entrepreneurial ventures. *Entrepreneurship: Theory and Practice, 35*(4), 683–708. doi:10.1111/j.1540-6520.2010.00387.x

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.

Cadogan, J. W., & Lee, N. (2013). Improper use of endogenous formative variables. *Journal of Business Research, 66*(2), 233-241. doi:10.1016/j.jbusres.2012.08.006

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly, 33*(4), 689-707.

Chandon, P., Wansink, B., & Laurent, G. (2000). A benefit congruency framework of sales promotion effectiveness. *Journal of Marketing, 64*(4), 65-84. doi:10.1509/jmkg.64.4.65.18071

Chin, W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (p. 295-336). Mahwah, NJ: Lawrence Erlbaum.

Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*(3), 397-412. doi:10.1177/0013164407310130

Curtis, R. F., & Jackson, E. F. (1962). Multiple indicators in survey research. *American Journal of Sociology, 68*(2), 195-204.

Diamantopoulos, A. (2006). The error term in formative measurement models: interpretation and modeling implications. *Journal of Modelling in Management, 1*(1), 7-17. doi:10.1108/17465660610667775

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly, 35*(2), 335-358.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1201-1302. doi:10.1016/j.jbusres.2008.01.009

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicator. *Journal of Marketing Research, 38*(2), 269-277. doi:10.1509/jmkr.38.2.269.18845

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods, 14*(2), 370-388. doi:10.1177/1094428110378369

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155-174. doi:10.1037/1082-989X.5.2.155

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research, 61*(12), 1229-1237. doi:10.1016/j.jbusres.2008.01.011

Gefen, D., & Straub, D. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems, 16*, 91-109.

Hardin, A. M., Chang, J. C., Fuller, M. A., & Torkzadeh, G. (2011). Formative measurement and academic research: In search of measurement theory. *Educational and Psychological Measurement 71*(2), 281-305. doi:10.1177/0013164410370208

Hauser, R. M., & Warren, J. R. (1997). Socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology, 27(*1*)*, 177-298. doi:10.1111/1467-9531.271028

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007) *Psychological Methods 12*(2), 238-245. doi:10.1037/1082-989X.12.2.238

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007b). Reconsidering formative measurement. *Psychological Methods 12*(2), 205-218. doi:10.1037/1082-989X.12.2.205

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199-218. doi:10.1086/376806

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement: American Council on Education / Praeger Series on Higher Education* (4th ed., pp. 17-64). Westport, CT: Praeger Publishers.

MacCallum, R. C. & Browne, M. W. (1993). The use of causal indicators in covariance structure models: some practical issues. *Psychological Bulletin 114*(3), 533-541. doi:10.1037/0033-2909.114.3.533

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology, 90*(4), 710-730. doi:10.1037/0021-9010.90.4.710

MacKenzie, S. B., Podsakoff, P.M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. *MIS Quarterly, 35*(2), 293-334.

Messick, S. (1995). Validity of psychological assessment, *American Psychologist*, *50*(9), 741-749. doi:10.1037/0003-066X.50.9.741

Muthén, L.K. & Muthén, B.O. (1998-2012). *Mplus User's Guide*. (7th Ed.). Los Angeles, CA: Muthén and Muthén.

Pavlou, P. A., & Gefen, D. (2005). Psychological contract violation in online marketplaces: antecedents, consequences, and moderating role. *Information Systems Research, 16*(4), 372-299. doi:10.1287/isre.1050.0065

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4)*, 623-656.

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 3*(4), 369-379. doi:10.1080/10705519609540052

Straub, D., Boudreau, M., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems 13*, 380-427.

Wang, X., Jessup, L. M., & Clay, P. F. (2015). Measurement model in entrepreneurship and small business research: a ten year review. *International Entrepreneurship and Management Journal, 11*(1), 183-212. doi:10.1007/s11365-013-0285-0

Woodworth, R. S. (1928). Dynamic psychology. In C. Murchison (Ed.), *Psychologies of 1925* (pp. 111-126). Worcester, MA: Clark University Press.