Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2015 Research-in-Progress Papers | ECIS 2015 Proceedings

Spring 5-29-2015

Valid Construct Measurement Using Multiple Models

Sander Paul Zwanenburg
The University of Hong Kong, info@sanderpaul.com

VALID CONSTRUCT MEASUREMENT USING MULTIPLE MODELS

Research in Progress

Zwanenburg, Sander Paul, The University of Hong Kong, Hong Kong, info@sanderpaul.com

Abstract

Invalid measurement of constructs in survey research often remains undetected and can lead to false conclusions. An important determinant of a construct's measurement validity is how it is modeled. A construct can often be modeled in different ways, such as the sum of its parts or the cause of its effects. Since each of these models is associated with a unique set of errors, the common practice of specifying only a single model undermines validity. Current guidelines on measurement have not focused on how better validity can be achieved by comparing and combining multiple models. In this paper we provide a framework for the development and use of multiple models. This, we hope, would lead researchers view their construct of interest from different perspectives and thus measure it more validly.

Keywords: Measurement, models, method, indicator, construct, item

1 Introduction

Invalid measurement has long been a hidden problem in the field of Information Systems (Boudreau et al. 2001; MacKenzie et al. 2011; Straub 1989). Validation is often hampered by unclear conceptual definitions (MacKenzie 2003) and unused tools and techniques (MacKenzie et al. 2011).

Meta-analytical and methodological studies suggest that the hidden problem of invalidity impedes theory development. After synthesizing a subset of eleven meta-analyses on estimating treatment effects, Wilson and Lipsey (2001, p419) conclude: "our indication that different operationalizations of what is presumed to be the same outcome construct within the same treatment domain can lead to quite different results is disconcerting." More broadly, the risk of using and continuing to use invalid operationalizations is indeed disturbing because it may lead us to draw the wrong theoretical and practical conclusions (Burton-Jones 2009; Viswanathan 2005).

Researchers have devoted much attention to the underlying reasons for measurement error. They have identified many common sources of error and proposed various ways to control for them, both before and after data collection (e.g. Burton-Jones 2009; Podsakoff et al. 2003; Podsakoff et al. 2012; Tourangeau et al. 2000; Viswanathan 2005). The multitude of potential errors can be mind-boggling; many published studies do not even discuss errors or their potential effects on results (Schmidt and Hunter 1999). Schmidt and Hunter (1999, p183) conclude that, "failure to control for biases induced by measurement error has retarded the development of cumulative research knowledge."

Designing a measurement model that controls for error is difficult because errors often correspond to multiple of its aspects, such as how questions are asked, when, where, to whom, and how the meaning of these questions relate to the construct (Burton-Jones 2009). In Spector's (2006, p230) words, "each operationalization of a variable [] carries with it a unique set of potential biases."

Modeling a construct from multiple perspectives can result in more valid measurement (Campbell and Fiske 1959). This relies on triangulation; the same logic that underlies the common practice of measuring using multiple reflective items. Like applying a reflective item, applying an entire model is a way of taking a look at an entity's position on the construct and capturing a reflection. Any reflection is to some degree distorted, contaminating our estimation. We can remove some of this contamination by using multiple reflections that are distorted differentially. Thus, analogous to reflective items, models can be compared and combined in a larger model.

The principle of triangulating across models has rarely been adopted. Exceptions include various information retrieval constructs such as relevance (Harter 1992). While methods have been proposed to leverage the principle across methods, they have focused on *evaluating* validity, not on *improving* it (Campbell and Fiske 1959; MacKenzie et al. 2011).

In this paper, we provide a measurement framework that aims to help researchers operationalize their construct in multiple models. We typify ways to tie a construct to its indicators, and how taking such steps iteratively can generate complementary models. Subsequently, we discuss when the multi-model approach is most beneficial, what its limitations are, and where future research is needed.

2 A Structure of Measurement

The key to measuring a construct is the items that tie it to observations. As shown if Figure 1, items specify what is to be observed in order to measure the construct. While a combination of items indicates a construct conceptually, each item separately relates operationally to one observation for each individual, firm, or any other entity to be measured. Records of these observations combine mathematically to produce estimates, from which to draw measurement inferences.

Measurement is valid to the extent an estimate, i.e. what *is* measured, matches the meaning of the construct, i.e. what is *to be* measured (Markus and Borsboom 2013; Rigdon 2013). As can be seen from Figure 1, this match has three components: one conceptual, one operational, and one mathematical. More error in any of those components means lower measurement validity.

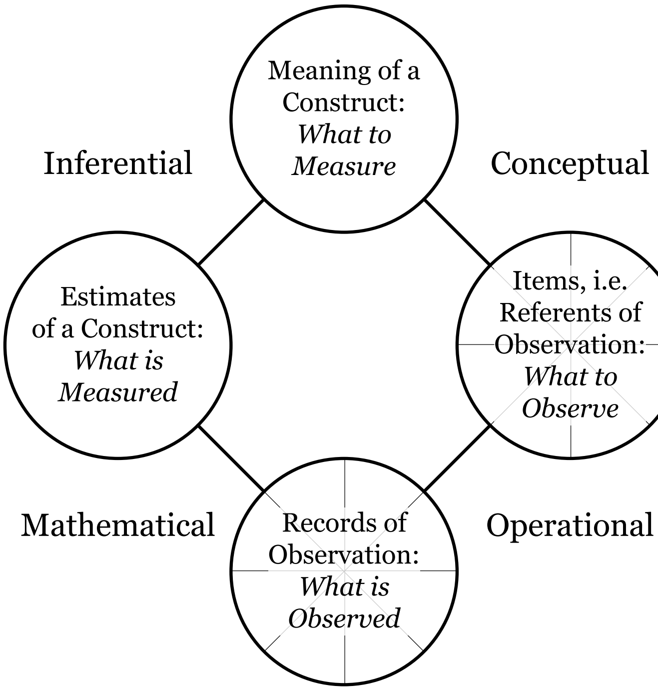

Figure 1. A Structure of Measurement

Error is conceptual when a construct is not equivalent to the combination of items. This error occurs, for example, when items correspond to only *some* parts of a construct, or when they correspond to effects of a construct that are also effects of something else. Operational errors occur when what is observed deviates from what was to be observed. For example, a respondent misinterprets a question or lies about it. Mathematical errors are due to the researcher combining the records of observations in a way that departs from what the conceptual relationship stipulates. Without these three types of error, a construct is measured with perfect validity: what is measured equals what was to be measured.

2.1 The Conceptual Relationship

The relationship between what to measure and what to observe can take many forms. The simplest relationship is when an item is defined as equivalent to the construct, being its only indicator. Gender and age are examples of constructs typically assessed with one item. In those cases, the match between the construct and the item is perfect; no conceptual error exists.

Many constructs, of course, are measured with multiple items (Bergkvist and Rossiter 2007). The degree to which the combination of multiple items is equivalent to the meaning of a construct – we will call this *conceptual validity* for short – depends on the logic and knowledge that underlies this relationship. In an application of measuring frequency of technology use, Carlson and Grossbart (1988) included three items: use on weekdays, use on Saturdays, and use on Sundays. Logic and knowledge – in this case about time – allowed the researchers decompose the construct into multiple parts. Logic and knowledge also allow us to evaluate that a weighted sum of an individual's positions on these items is equivalent to that on the construct, meaning the conceptual component of validity is perfect.

Measuring technology use, or any construct for that matter, can be compared to measuring the mass of a pie. Carlson and Grossbart (1988) cut the pie at one angle – the angle of the structure of the week – to get three pieces that are perhaps more operable than the whole. This angle is one of an infinite number of angles at which the pie, or technology use frequency, can be cut. Other angles could divide it into its different features, locations, situations, durations, and so on. Here, the underlying logic is that of summing parts to make up the whole. This conceptual analysis is valid as long as the items together represent the entire conceptual domain of the construct (Churchill 1979; Clark and Watson 1995; Haynes et al. 1995; MacKenzie et al. 2011).

We could also cut based on causal logic, breaking up a construct into its causes or its effects. In practice, this logic is probabilistic, making perfect conceptual validity unattainable. But these angles may help define more operable pieces, i.e. pieces that can be measured without much operational error. To stretch the analogy, asking your guests how full they are after eating your pie and if they want another one is a crude way of measuring its mass, but it is worth considering in case the pie has been eaten already. These are conditions where physical measurement is complicated if not impossible; these conditions often apply in behavioral information systems research.

Whether we use mereology or causality to analyze the construct, the resulting pieces stand in the *same* qualitative relationship to the construct. They are all constituent parts, all causes, or all effects. An alternative is to cut pieces with qualitatively different relations to the construct. For example, we could measure the mass of the pie by multiplying its density with its volume. Similarly, we could measure technology use by multiplying use in a past period with a growth rate. While at first glance such dimensions may seem to complicate the measurement problem, contextual conditions may exist in which they ease it.

So far, all these examples of conceptual analysis rely on the knowledge of the *substance* of the construct. But knowledge on *method* may also drive a conceptual analysis. For example, a writer of a question that should capture the meaning of the frequency of technology use may grow concerned with potential differences in interpretation: would respondents interpret 'smartphone use' only as *active* use or as both active and passive, including listening to music? Some guidelines advise to generate items that differ in syntax, rather than semantics (Churchill 1979; Netemeyer et al. 2003). Generating items that refer to '*using the smartphone*' and '*the smartphone being used*' may lessen the problem of ambiguous stimuli.

This approach is an example of breaking up a construct's method of measurement into pieces. It is cut at an angle of one measurement aspect, in this case *how* a question is asked. Other analyzable aspects are *to whom* a question is asked, i.e. by using multiple informants (Kumar et al. 1993), and *when* it is asked, i.e. by using ecological momentary assessment (also called the experience sampling method; Csikszentmihalyi and Larson 1987; Hektner et al. 2007). Which aspect to cut depends on the magnitude of error associated with that aspect. By analogy, we could weigh the entire pie using multiple scales in case we do not trust any particular scale.

These examples involve qualitatively similar items. Yet similar to a substantive analysis, a methodical analysis can involve cuts that yield qualitatively different pieces. We could cut up the measurement of the mass of the pie into the measurement of its weight and that of the weight-mass conversion factor. Similarly, when we are interested in *deviant* technology use, we could cut it up into a convenient piece and a corrective piece, such as self-reported deviant use and the tendency to give social desirable answers (Nederhof 1985; Podsakoff et al. 2012).

While the discussion so far has provided examples of a single step of conceptual analysis, it has ignored that linking a construct to a combination of items often consist of taking multiple consequent steps. Many constructs relate to their items through intermediate indicators (Edwards 2001; Law et al. 1998; Polites et al. 2012).

In the pie example, we can first equate the mass of the pie to the multiplication of its weight with a weight-mass conversion factor, and then equate the weight of the pie to the sum of the weights of its slices, etc. As summarized in Table 1, in each step the construct is conceptually analyzed based on method or substance so as to equate it to a combination of indicators that are more operable. This continues until all indicators are items, i.e. when every indicator refers to one observation for each entity to be measured. That last step completes the conceptual design of a measurement model.

Relation to Construct across Indicators	How an Indicator Relates to its Construct	
	Substantively	Methodically
Same	A substantive aspect of the construct distinguishes the indicators. The logic relies on mereology or causality related to the substance of a construct; indicators refer to its constituent parts, its causes, or its effects (Bollen and Bauldry 2011; MacKenzie et al. 2011).	A measurement aspect distinguishes the indicators. The logic relies on an understanding of how this aspect corresponds to sources of error that are difficult to isolate. Indicators could refer to *how* a question is asked (Netemeyer et al. 2003), *to whom* (Kumar et al. 1993), *when* (Csikszentmihalyi and Larson 1987), and even *where*.
Different	Indicators refer to concepts that relate differently to the construct. The underlying logic relies on the 'spatial' structure of a construct. Examples include multiplying a construct's dimensions (Law et al. 1998; Vroom 1964), and, conversely, dividing a multi-dimensional concept of which the construct is a dimension by its other dimension(s).	An indicator can help measure the construct by measuring the bias of another indicator. This relies on the assumption that a particular method error can be isolated. Examples include confirmatory factor models where all items suspected of suffering from one type of bias are loaded on a factor representing that bias (Harman 1976; Nederhof 1985; Podsakoff et al. 2003; Podsakoff et al. 2012).

Table 1. A typology of the relationship between a construct and a combination of indicators

2.2 The Mathematical Relationship

The mathematical relationship mirrors the conceptual relationship: each step in the link between the construct with its items corresponds to one unique step in combining records of observation. When the conceptual step is deterministic, it prescribes exactly how the mathematical step is made. When it is not, specifying a model can be more complicated.

Various mathematical models exist, with varying assumptions about the type of scales of variables, the distributions of the variables, the form of their relationship, and the modeling of error (Mellenbergh 1994; Nunnally and Bernstein 1994). Many applications of measurement models draw on classical test theory, in which an observation is equated to a linear combination of the true score, i.e. the position on the referent of that observation, and the error score. In a popular extension, confirmatory factor analysis, multiple indicators are modeled in this way, typically under the assumption that all error scores are mutually independent. Other extensions allow for modeling different types of indicators (Jöreskog and Goldberger 1975), or multiple layers of indicators (Edwards 2001).

Much attention has been paid to how the *direction* of the conceptual link should determine its mathematical model (Blalock 1964; Bollen and Lennox 1991; Bollen and Bauldry 2011; Diamantopoulos et al. 2008; Diamantopoulos and Winklhofer 2001; Edwards 2001; Edwards and Bagozzi 2000; Kim et al. 2010; MacKenzie et al. 2011; Petter et al. 2007). This discussion assumes that (1) multiple entities are measured, (2) all indicators relate to the construct in qualitatively the same way, and (3) the link is best characterized as unidimensional. If indicators can be thought to be dependent of the construct, this direction is reflective, flowing from the construct to the indicators. If the construct is thought to be dependent of its indicators, this direction is formative, flowing from the indicators to the construct (MacKenzie et al. 2011). In terms of the 'same' row of Table 1, cause and part indicators are formative, while effects and method indicators are reflective.

Reflective indicators are commonly advocated or assumed to be modeled through factor analysis, although some have argued for simply summing scores on reflective indicators (Spector 1992). How formative indicators are best modeled has been topic of an on-going debate (Bollen and Bauldry 2011; Diamantopoulos and Temme 2013; Edwards 2011; Howell et al. 2013; Lee et al. 2013; MacKenzie et al. 2011; Markus and Borsboom 2013; Rigdon 2013).

In our view, most generally, the measurer is best advised aim to adopt a mathematical model that fits the measurement problem at hand. That is the model that best fits the understanding of the conceptual link between a construct and a combination of indicators, within the context of the entire conceptual relationship specifically and that of the inquiry more generally.

This context of inquiry limits the mathematical options when only a single entity is measured and little prior knowledge exists on the relationships between the variables. Further, when multiple links in the conceptual relationship are non-deterministic (Edwards 2001), or the measurement fits into a larger structural model, one may consider the advantages of modeling multiple links integrally, such as with structural equation modeling (Chin 1998; Jöreskog and Sörbom 1993).

Importantly, while a conceptual link is the basis for the choice of the mathematical model, an understanding of these models should help researchers in designing the conceptual relationship that can be easily mirrored mathematically.

3 A Process of Measurement

The goal of measuring a construct is to infer the position of an entity on the construct, i.e. to describe the entity according to the meaning of the construct. The meaning of a construct is thus the starting point of measurement (MacKenzie et al. 2011).

Knowing what to measure can be confusing (MacKenzie 2003). A construct's name, i.e. its label or term, can also denote other constructs as it can carry multiple meanings. Further, one meaning can be described in different ways; a construct can be defined in different languages and syntaxes. Definitions of a construct can also highlight different aspects of its meaning. As long as they are consistent, multiple definitions of the same construct help specify its meaning, i.e. they demarcate it (Barki 2008; Goertz 2006; MacKenzie et al. 2011). These definitions must be consistent; inconsistent definitions do not define the same construct. The meaning of a construct can never be fully specified; any description of a construct is to some degree ambiguous (Kaplan 1964; Van de Ven 2007).

It may be helpful to visualize a construct as a node in a hyper-dimensional web of nodes connected by strings. The web is an internally consistent map of reality, with nodes representing concepts and strings. The strings position the nodes, giving it meaning. When we define a construct, we take a string and attach its node to other nodes. We may define these nodes too in a similar way or assume their location is bound by our prior knowledge or common understanding of them.

Attaching strings will limit the potential space the node can occupy, analogous to demarcating the meaning of the construct in the space of reality by providing definitions. Ideally, this is done in a way that specifies the position best. Since this reality is hyper-dimensional, there will always be degrees of freedom. How we best attach strings to limit them depends on the larger goal of inquiry. If this goal is to assess the relationship between two constructs, i.e. to investigate the distances between two nodes, the definitions of the constructs should remove as much as variance in the possible distances between their nodes. This will help make stronger inferences, revealing more of the shape of reality (Kaplan 1964).

Measuring a construct, then, is approximating the location of its node, or estimating a point on the surface of reality. In the analogy, this is by way of assessing the distance with other nodes. We will infer a position more reliably when assessing these distances with nodes in multiple directions. Similarly, triangulation in the measurement of a construct will be more successful when based on more diverse concepts.

This visual analog highlights not only the use of our *structure* of understanding in measurement but also that of its *process*. Experimenting with measuring distances to different concepts should help determine how to best fix a position. Specifically, commonly advised practices for developing measurement instruments include pretests and pilot tests, in which responses to stimuli are compared with what was expected (Churchill 1979; Dillman 2000; MacKenzie et al. 2011). These tests can aid the level of a stimulus, an item, a combination of indicators, and the level of constructs, in case of assessing nomological validity (MacKenzie et al. 2011). Such testing may involve thought experiments, expert judgments, participant interviews, surveys, and actual experiments. All these approaches can help validate and calibrate measurement.

The focus in this paper is on the design of an initial condition with multiple models as these can benefit the most from subsequent testing.

3.1 Conceptual Design

The conceptual design of a measurement model consists of tying a construct to a combination of items. We can specify a combination of items through one or more consecutive steps of conceptual analysis. As shown in Table 1, at each step, we can generate indicators that are heterogeneous or homogenous, and based on substance or on method. This latitude allows for a plethora of conceptual measurement models.

As a first step, we can consider a construct's potential ties: what it causes, what it is caused by, how it changes over time, what it is composed of, what it is part of, what its dimensions are, what it is a dimension of, who perceives it, how it is recorded, and so on (Goertz 2006; MacKenzie 2003; MacKenzie et al. 2011). While some of these ties may be subject to hypotheses, those that are better understood may provide a portal for measurement.

This evaluation often relies on consulting prior literature and theories, reviewing extant measures, asking experts, using focus groups, and conducting explorative surveys (Churchill 1979; MacKenzie et al. 2011; Netemeyer et al. 2003). Coupled with logic, such inquiry allows us to draw a map of the construct, positioning it in a web of concepts.

How promising these concepts are as potential portals to specifying a measurement depends not only on how well their relationship with the construct is understood; they also should make the construct more operable. Thus, evaluating potential indicators requires iterating this mapping exercise, keeping in mind the goal of minimizing conceptual and operational error. Each step should introduce as little conceptual error as possible, while also move toward potential items. Items are best specified at such a level that observations can be reasonably assumed to correspond to their referent, i.e. operational error is minimal. Knowledge on the meaning of the construct and on operations of measurement will help maintain direction, focusing on feasible ways to tie a construct to observations. See Figure 2 for how a part of such an iterative map could look like. Each of the rectangles corresponds to one model.

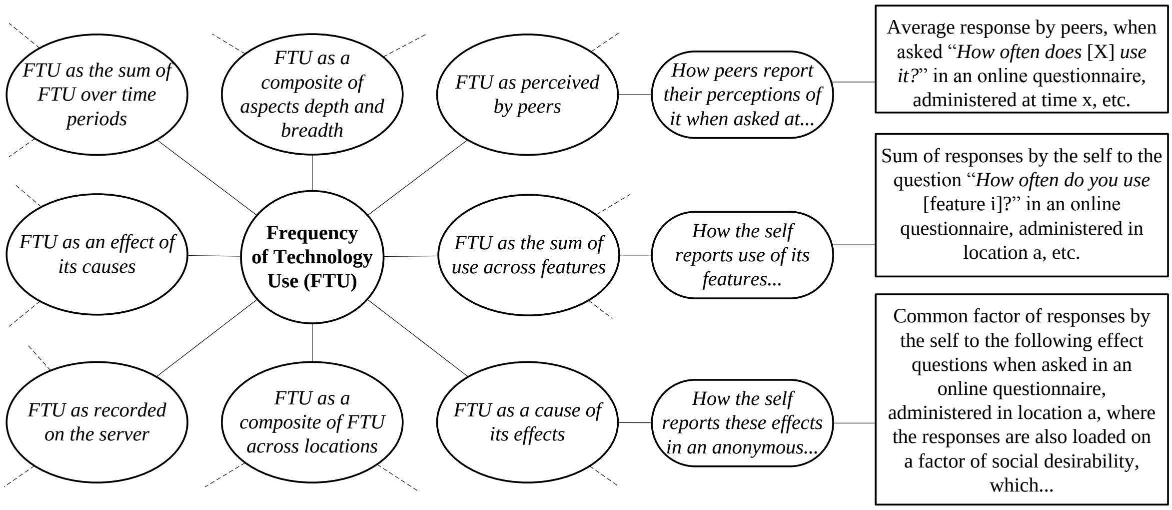

FIGURE 2: MAPPING POTENTIAL MODELS OF FREQUENCY OF TECHNOLOGY USE

Each model's error consists of the conceptual errors introduced in each of the links that tie the model to the construct, and the operational error of its items. This analysis of error should help assess the relative quantity and quality of error across models. Models that have clearly the most error may be discarded, while models with heterogeneous errors quality are best retained; models that are erroneous for different reasons carry the promise of triangulation.

Figure 3 illustrates this comparison with an imaginary example in which we know the quality and quantity of errors across five models. Two of them are clearly inferior, while three others are heterogeneous in their error, making them complementary.

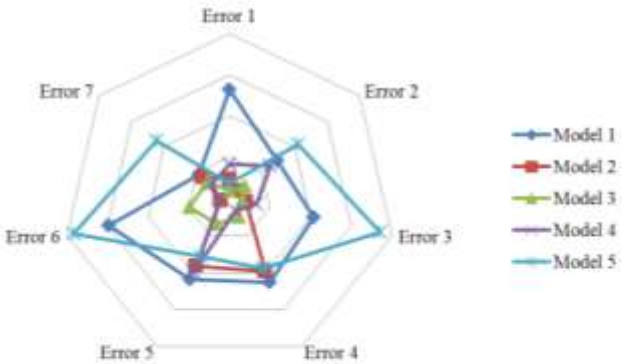

FIGURE 3: COMPARING MODELS FOR TRIANGULATION

In a recursive way, an entire model can be seen as a reflective indicator. Both reflective items and entire measurement models aim to capture the entire meaning of a construct. Similar to evaluating reflective items, we could put multiple models to the test. Their relative performance may be evaluated using the same criteria as those developed for evaluating reflective items (Fornell and Larcker 1981; MacKenzie et al. 2011). In place of item scores, we can use composite scores, factor scores, or any other point estimate of the construct produced by a mathematical model. An evaluation of the models' relative performance may suggest refining, dropping, or retaining them. When we retain multiple models we may use them in a factor analysis or a composite so as to produce valid, multi-model measurement.

4 Discussion

In this paper, we argue why and how we can use multiple models for better measurement validity. But when is this approach worth the effort? More models mean more effortful design, and perhaps costlier operations.

In general, more valid measurement can prevent mistakes in drawing research conclusions (Viswanathan 2005). Such mistakes are expensive when they lead a stream of research down a blind alley. Multiple models are more likely to prevent wrong inferences when research conclusions are more sensitive to measurement validity, such as when studying small effects or using small samples. That is, better measurement validity can compensate for lack of statistical power. Further, using multiple models should pay off more when even the best measurement model of a construct suffers from much error.

Thanks to technology, operating costs of measurement methods have dropped. For instance, the internet and mobile devices have made it easier to measure through the momentary assessment method. This approach complements more traditional approaches, being less sensitive to errors specific to location and time (Hektner et al. 2007; Podsakoff et al. 2012). Innovations in digital payments and location-based services will reduce the administrative burden of such measurement methods, and increase its usefulness.

One limitation of multi-model measurement is that it may complicate standardization of measurement. While standardization generally helps in comparing multiple studies, the premise of meaningful comparisons is that constructs are measured with sufficient validity. The adoption of multiple models can help ensure this premise. We would thus argue that the adoption of multiple models aides comparison.

Future research could develop mathematical models that better fit the simultaneous use of multiple measurement models of the same construct. Using factor analysis with indicators that are themselves factor scores, for example, involves a loss of information, which may affect the ability to identify and evaluate structural models. An alternative of specifying a multi-level measurement model may be difficult since it introduces cross-model relationships that may be difficult to evaluate.

Error can threaten the validity of measurement in a plethora of ways, especially in survey research. Conventional procedures and techniques to control for these errors are limited because they assume a single model of a construct. Many measurement errors are specific to its particular underlying view of the construct. In this paper, we advance why, how, and when combining multiple models can lead to better measurement validity. Multiple models may be generated by applying a structural and process framework presented in this paper. We hope that by contextualizing and typifying the ways in which a construct can be tied to observations this paper will lead researchers come up with complementary models allowing them to measure construct with better validity.

Acknowledgements

I am thankful to Ali Farhoomand, Israr Qureshi, Nikolaos Soultanidis and four anonymous reviewers for their constructive feedback on an earlier version of this manuscript.

References

Barki, H. 2008. "Thar's Gold in Them Thar Constructs," *ACM SIGMIS Database* (39:3), pp 9-20.

Bergkvist, L., and Rossiter, J.R. 2007. "The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs," *Journal of marketing research* (44:2), pp 175-184.

Blalock, H.M. 1964. *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

Bollen, K., and Lennox, R. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin* (110:2), p 305.

Bollen, K.A., and Bauldry, S. 2011. "Three Cs in Measurement Models: Causal Indicators, Composite Indicators, and Covariates," *Psychological methods* (16:3), p 265.

Boudreau, M.-C., Gefen, D., and Straub, D.W. 2001. "Validation in Information Systems Research: A State-of-the-Art Assessment," *MIS Quarterly* (25:1), pp 1-16.

Burton-Jones, A. 2009. "Minimizing Method Bias through Programmatic Research," *MIS Quarterly* (33:3), pp 445-471.

Campbell, D.T., and Fiske, D.W. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin* (56:2), p 81.

Carlson, L., and Grossbart, S. 1988. "Parental Style and Consumer Socialization of Children," *Journal of Consumer Research* (15:1), pp 77-94.

Chin, W.W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," *Modern methods for business research* (295:2), pp 295-336.

Churchill, G.A. 1979. "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of marketing research* (16:1), pp 64-73.

Clark, L.A., and Watson, D. 1995. "Constructing Validity: Basic Issues in Objective Scale Development," *Psychological Assessment* (7:3), p 309.

Csikszentmihalyi, M., and Larson, R. 1987. "Validity and Reliability of the Experience-Sampling Method," *The Journal of nervous and mental disease* (175:9), pp 526-536.

Diamantopoulos, A., Riefler, P., and Roth, K.P. 2008. "Advancing Formative Measurement Models," *Journal of Business Research* (61:12), pp 1203-1218.

Diamantopoulos, A., and Temme, D. 2013. "Mimic Models, Formative Indicators and the Joys of Research," *AMS review* (3:3), pp 160-170.

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

Dillman, D.A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. Wiley New York.

Edwards, J.R. 2001. "Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework," *Organizational research methods* (4:2), pp 144-192.

Edwards, J.R. 2011. "The Fallacy of Formative Measurement," *Organizational research methods* (14:2), pp 370-388.

Edwards, J.R., and Bagozzi, R.P. 2000. "On the Nature and Direction of Relationships between Constructs and Measures," *Psychological methods* (5:2), p 155.

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

Goertz, G. 2006. *Social Science Concepts: A User's Guide*. Princeton University Press.

Harman, H.H. 1976. *Modern Factor Analysis*. Chicago: University of Chicago Press.

Harter, S.P. 1992. "Psychological Relevance and Information Science," *Journal of the American Society for information Science* (43:9), pp 602-615.

Haynes, S.N., Richard, D., and Kubany, E.S. 1995. "Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods," *Psychological Assessment* (7:3), p 238.

Hektner, J.M., Schmidt, J.A., and Csikszentmihalyi, M. 2007. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Sage.

Howell, R.D., Breivik, E., and Wilcox, J.B. 2013. "Formative Measurement: A Critical Perspective," *ACM SIGMIS Database* (44:4), pp 44-55.

Jöreskog, K.G., and Goldberger, A.S. 1975. "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association* (70:351a), pp 631-639.

Jöreskog, K.G., and Sörbom, D. 1993. *Lisrel 8: Structural Equation Modeling with the Simplis Command Language*. Scientific Software International.

Kaplan, A. 1964. *The Conduct of Inquiry: Methodology for Behavioral Science*. Chandler.

Kim, G., Shin, B., and Grover, V. 2010. "Investigating Two Contradictory Views of Formative Measurement in Information Systems Research," *MIS quarterly* (34:2), pp 345-365.

Kumar, N., Stern, L.W., and Anderson, J.C. 1993. "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal* (36:6), pp 1633-1651.

Law, K.S., Wong, C.-S., and Mobley, W.M. 1998. "Toward a Taxonomy of Multidimensional Constructs," *Academy of Management Review* (23:4), pp 741-755.

Lee, N., Cadogan, J.W., and Chamberlain, L. 2013. "The Mimic Model and Formative Variables: Problems and Solutions," *AMS review* (3:1), pp 3-17.

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

MacKenzie, S.B., Podsakoff, P.M., and Podsakoff, N.P. 2011. "Construct Measurement and Validation Procedures in Mis and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly* (35:2), pp 293-334.

Markus, K.A., and Borsboom, D. 2013. *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning*. Routledge.

Mellenbergh, G.J. 1994. "Generalized Linear Item Response Theory," *Psychological Bulletin* (115:2), p 300.

Nederhof, A.J. 1985. "Methods of Coping with Social Desirability Bias: A Review," *European Journal of Social Psychology* (15:3), pp 263-280.

Netemeyer, R.G., Bearden, W.O., and Sharma, S. 2003. *Scaling Procedures: Issues and Applications*. Sage.

Nunnally, J.C., and Bernstein, I.H. 1994. *Psychometric Theory*, (3 ed.). New York, NY: McGraw-Hill.

Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp 623-656.

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), pp 879-903.

Podsakoff, P.M., MacKenzie, S.B., and Podsakoff, N.P. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It," *Annual review of psychology* (63), pp 539-569.

Polites, G.L., Roberts, N., and Thatcher, J. 2012. "Conceptualizing Models Using Multidimensional Constructs: A Review and Guidelines for Their Use," *European Journal of Information Systems* (21:1), pp 22-48.

Rigdon, E.E. 2013. "Lee, Cadogan, and Chamberlain: An Excellent Point... But What About That Iceberg?," *AMS review* (3:1), pp 24-29.

Schmidt, F.L., and Hunter, J.E. 1999. "Theory Testing and Measurement Error," *Intelligence* (27:3), pp 183-198.

Spector, P.E. 1992. *Summated Rating Scale Construction: An Introduction*. Sage.

Spector, P.E. 2006. "Method Variance in Organizational Research Truth or Urban Legend?," *Organizational research methods* (9:2), pp 221-232.

Straub, D.W. 1989. "Validating Instruments in Mis Research," *MIS quarterly*), pp 147-169.

Tourangeau, R., Rips, L.J., and Rasinski, K. 2000. *The Psychology of Survey Response*. Cambridge University Press.

Van de Ven, A.H. 2007. *Engaged Scholarship: A Guide for Organizational and Social Research: A Guide for Organizational and Social Research*. Oxford University Press.

Viswanathan, M. 2005. *Measurement Error and Research Design*. Sage.
Vroom, V.H. 1964. *Work and Motivation*. New York: Wiley.
Wilson, D.B., and Lipsey, M.W. 2001. "The Role of Method in Treatment Effectiveness Research: Evidence from Meta-Analysis," *Psychological methods* (6:4), p 413.