2017

Towards a Synthesis of the Guidelines for the Development of Measurement

Sander Paul Zwanenburg
University of Otago, info@sanderpaul.com

Towards a Synthesis of the Guidelines for the Development of Measurement

Sander Paul Zwanenburg
Department of Information Science
University of Otago
Dunedin, New Zealand
Email: sander.zwanenburg@otago.ac.nz

Abstract

The literature on the development of measurement help researchers with steps before and after data is collected. For example, some guidelines aid the modelling of constructs and the generation of indicators while others help with analytical procedures to use indicator data in estimation. While valid measurement requires a coordination between these two sets of tasks, their corresponding streams of literature have been largely separate. With this paper, we aim to contribute toward more integral guidance, which should help researchers attain validity of measurement. We propose a conceptual framework that helps researchers identify the logical steps of measurement and their interdependencies. We show how the framework can be used to identify the sources of error specific to a measurement. We further provide initial guidelines on how error-prone measurement designs can be remedied. Finally, we suggest how future research could take next steps in establishing more integral guidelines.

Keywords Measurement, construct, indicator, model, operationalization.

1 Introduction

The literature on the development of measurement has helped to address two sets of issues. Some guidelines aid with issues before data collection such as modelling a construct conceptually (Barki 2008), generating items and their response scales (Tourangeau et al. 2000), designing a questionnaire (Dillman 2000), and so on. Other guidelines help with the steps after data collection, to produce and evaluate measurement estimates and its derivatives, such as the fit of research models. While these two streams of literature have each accumulated their own insights, we believe that synthesizing these insights would improve their usefulness and help researchers develop valid measurement.

Naturally, the conceptual, operational, and inferential aspects of measurement are intricately connected (Petter et al. 2012; Zwanenburg 2015a). The use of indicator data to draw inferences about a construct ought to reflect the relation between that data and the meaning of that construct. This relation depends on how the construct was modelled conceptually with a set of indicators, and how these indicators yielded data.

Yet, while it is well established that measurement ought to be driven by theory (DeVellis 2003; Hair et al. 2010; Petter et al. 2012), guidelines have provided limited help in combining a substantive and methodological understanding and using it integrally to attain validity. Guidelines on generating items, for example, do not consider the implications of the design of items for estimation (Dillman 2000). Conversely, the guidelines on formal measurement models carry standardized assumptions about the relation between the construct and the indicators yet do not provide recommendations on how to satisfy those assumptions through measurement design (Aguirre-Urreta et al. 2012). While some guidelines do explicitly consider both aspects, they do so separately, like by suggesting to generate items and specifying a measurement model sequentially, rather than holistically (MacKenzie et al. 2011). Finally, methodological research on sources of error like common method bias has produced recommendations to avoid specific errors, but has not yet integrated these insights into broader measurement guidelines (Burton-Jones 2009; Podsakoff et al. 2012; Sharma et al. 2009).

A danger of this lack of synthesis is that decisions in measurement design neglect due consideration of their implications. This could explain a common mismatch between the design of measurement and estimation procedures, or a reliance on unrealistic assumptions such as complete independence of errors in factor analytic models with homogeneous indicators (Aguirre-Urreta et al. 2012; Jarvis et al. 2012; Petter et al. 2012; Zwanenburg 2015b). These issues can threaten the validity of research conclusions (Burton-Jones 2009; Viswanathan 2005).

Our long-term objective is to deepen our understanding of the connection between the tasks before and after data collection, and use this to provide integral guidelines for the development of measurement. With this paper, we take initial steps. We aim to develop a conceptual framework to aid the design of valid measurement. To do so, we modify a common visualization of measurement models to disentangle the relationship between a construct and its indicators. The resulting diagram visualizes the structure of measurement and the locations where error can undermine its validity. Using this structure, we analyse the literature on sources of error. We then provide recommendations on designing indicators based on an understanding of the most potent sources of error and the assumptions of formal measurement models. These steps should help ensure a good fit between the conceptual, operational, and inferential design aspects. Finally, we suggest how research could further help the development of integral guidelines on the development of valid measurement.

2 Analysing the Conduct of Measurement

The objective of measurement is to obtain estimates of a construct that fit the meaning of that construct. This fit – better known as the validity of measurement[1] – is the logical basis for drawing inferences, such as research conclusions (Burton-Jones 2009). In most cases, many potential sources of error can undermine validity and complicate the task of measurement (Viswanathan 2005).

Using indicators helps break it down. An indicator can represent a proxy of a construct, a manifestation of it, a part of a construct, a dimension of it, or any other element that stands in some relation to the

[1] In this paper, validity of measurement refers to the match between what is measured and what is to be measured (Markus and Borsboom 2013) and differs from what is termed 'construct validity,' a property of test score interpretations (see e.g. Borsboom et al. 2009). It also differs from validity as 'the lack of *systematic* error' (e.g. Carmines and Zeller, 1979; Adcock and Collier 2001), as complementary to 'reliability' as the lack of *random* error. In our definition, reliability is a form of validity. Other forms, like content validity, cross validity, face validity, refer to specific tests that can indicate problems with validity based on the domain of the construct, the sample, or the inspection of measurement.

construct. Ideally, these elements are easier to capture than the construct itself while the resulting data neatly informs the construct estimates or its derivatives.

Panels A and B of Figure 1 depict the two most popular measurement models: the reflective model (Panel A) and the formative one (Panel B). In those two models, the relationship between the construct and the indicators are represented with arrows: in a reflective model, changes in a construct are assumed to *lead to* changes in (all) the indicators, since they reflect the construct (MacKenzie et al. 2011). In a formative model, changes in a construct are assumed to *originate from* changes in the indicators, as they form the construct (MacKenzie et al. 2011). Sources of error are modelled as directly affecting either the indicators (in a reflective model) or the construct (in a formative model[2]).

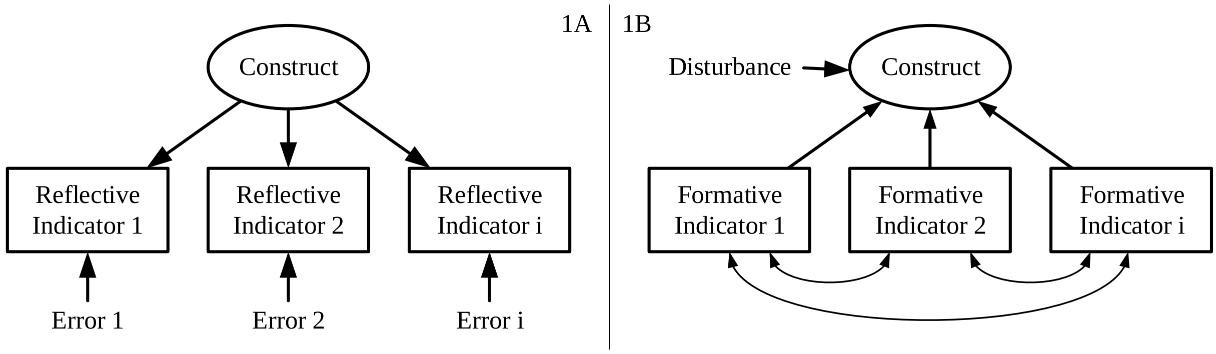

Figure 1: The reflective measurement model (Panel A) and the formative measurement model (Panel B)

The use of these models does not necessarily make measurement more straightforward, as revealed from an extensive debate on their appropriate use. Common issues are model misspecification, ignorance of model assumptions during indicator design, unaccounted sources of error, and confusion about when to use which model. We posit that the models oversimplify the relationship between construct and indicator. Indeed, as Bagozzi (2010) noted on the relation between a construct and a reflective indicator:

> *"It seems to me that the relationship in question is not causal, per se, but rather one of hypothetical measurement. That is, the relationship is between an abstract, unobserved concept and a concrete, observed measurement hypothesized to measure the concept; the relationship is part logical, part empirical, and part theoretical (conceptual), with the inferred factor loading representing, in and of itself, only part of empirical meaning of the relationship"* (p 210).

Disentangling the relation between a construct and its indicators can analyse and clarify the conduct of measurement. In Figure 2, we depict the overall task of measurement in Panel A and a decomposition of it in Panel B. These diagrams differ from the models in Figure 1 by distinguishing between the targets of a measurement (i.e. what is to be captured) in hollow shapes and the actual data (i.e. what is actually captured) in solid shapes. In a sense, the basic structure in Panel 2B is an *unfolded* version of those in Figure 1.

[2] This disturbance term is typically unidentifiable and ignored in estimation.

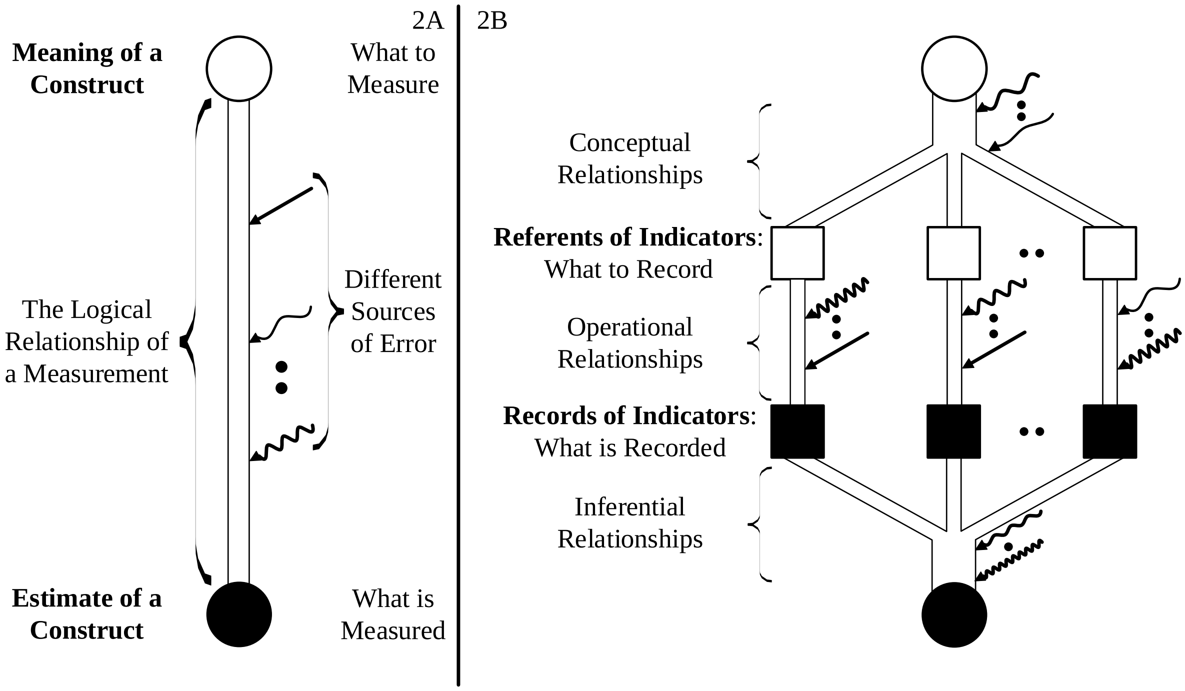

Figure 2: The overall task of measurement (Panel 2A) and a decomposition of it (Panel B)

This ‘unfolded’ model of measurement allows for an analysis of the overall relationship between measurement estimates and what they ought to represent. It is broken down into the following:

- **Conceptual relationships**: the relationship between the meaning of a construct and what an indicator is intended to record, i.e. its referent.
- **Operational relationships**: the relationship between the referent of an indicator and what is actually recorded, i.e. its records.
- **Inferential relationships**: the relationship between the records of an indicator and the estimates of a construct.

Any error in measurement is due to a deficiency in any of these relationships. Figure 2 shows how different sources of error, indicated by differently shaped arrows can undermine these relationships. It illustrates how they can interfere with the measurement process in different ways and at different locations. This map of potential error allows for a systematic evaluation of sources of error. With it, we may recognize, for example, that at the conceptual level a construct may not always manifests itself in a manifestation referred to by one of its indicators (Cook et al. 1979). Operationally, a respondent may misinterpret a question or give a dishonest answer (Podsakoff et al. 2012). At the inferential level, a factor model may be misspecified or unidentifiable (Aguirre-Urreta et al. 2012).

This decomposition can be used for developing or validating any kind of measurement, with any number of indicators, whether they are reflective, formative, or otherwise. It can be used independently of the overall method of measurement, whether operations rely on perception or on detection, and estimation techniques. In other words, measurements vary in the number and the nature of their constituent conceptual, operational, and inferential relationships.

Further analysing the nature of these relationships into more elementary logical links can deepen our understanding of the validity of a (proposed) measurement. Constructs may be multi-dimensional, where the conceptual relationship between a construct and its indicators is recursive, going through sub-constructs. If a relationship relies on causality, examining this causal mechanism may help identify sources of error. For example, a respondent in a questionnaire survey will see an item, read it, interpret it, evaluate it, and report on it, with each step involving its own potential interference. Table 1 illustrates a range of sources of errors associated with logical links of a typical case of measurements, involving a reflective model based on self-reported questionnaire data. In the Appendix, a similar table exposes common sources of error associated with a more generic logical links, applying to a wider variety of measurements.

Logical Link	**Associated Sources of Error**
Construct to Manifestation	A well-defined construct may be inappropriately modelled in terms of its manifestations (Aguirre-Urreta et al. 2012; Jarvis et al. 2012; Petter et al. 2012). A proclaimed manifestation may in fact not be consistent with the definition of the construct. It may be caused by other constructs, or the construct may only give rise to it under conditions that are not met, or it may not stand in a relationships to it as modelled (Cook et al. 1979; Rigdon 2014).
Manifestation to Question	A question may inappropriately capture a manifestation of a construct. For example, a question may refer to something else or may be unclear (e.g. Dillman 2000; Netemeyer et al. 2003; Tourangeau et al. 2000). Its response scale may be confusing, inconsistent with the question, or unable to capture accurate answers (e.g. Verhagen et al. 2015).
Question to Evaluation	Participants may inappropriately evaluate a question. For example, they may lack the motivation, energy, vocabulary, and other cognitive abilities to do so (Churchill 1979; Nunnally et al. 1994; Viswanathan 2005). A question may be too difficult or illegible. The instructions may be unclear and the time pressure and incentives may be inappropriate. The time of the day, the location, and the order of the questions may have an unintentional influence on their evaluations (Dillman 2000; Drury et al. 1997; Harrison et al. 1996; Podsakoff et al. 2012; Schwarz et al. 1992). A participant may have distracting thoughts and feelings while evaluating a question, due to idiosyncratic associations with certain words, or perceptions of fatigue, hunger, pain, noise, a phone ringing, or even due to simultaneous actions (e.g. Edwards 2008).
Evaluation to Response	An evaluation may not be reported. Questions may be too sensitive to answer honestly (e.g. Dillman 2000; Netemeyer et al. 2003; Tourangeau et al. 2000). The participant's anonymity or the lack thereof may affect the honesty of the response. A participant may have certain response tendencies (e.g. Podsakoff et al. 2003), or lack the motivation or incentives to provide accurate answers (e.g. Aronson et al. 1998; Podsakoff et al. 2003; Podsakoff et al. 2012; Richman et al. 1999).
Response to Record	An accurate response may be inappropriately recorded due to its illegibility, a data entry mistake, a technical failure, and so on.
Record to Factor Score	A well-recorded response may not be used appropriately for calculating factor scores as its assumptions may be violated. The specification of the factor model may deviate from the conceptual model, i.e. it may deviate from the specified relations between a construct and the referents of its indicators (e.g. Petter et al. 2007; Rigdon 2013).
Factor Score to Estimate	A factor score may be an inappropriate estimate when the assumptions underlying the factor analysis are violated. These typically include local independence, linearity of relationships, and homogeneity of relationships across entities (e.g. Becker et al. 2013; Havlicek et al. 1977; Jarque et al. 1987; Petter et al. 2007). Further, factors scores are indeterminate: their validity depends on the arbitrary method chosen to calculate them (Mulaik 2010; Rigdon 2012).

Table 1: Sources of error associated with common logical links in a typical case of measurement

3 Evaluating the impact of sources of error

A deeper understanding of the most potent sources of error allows us to generate a better quantitative map of how records of each indicator might relate to the construct, like the examples in Figure 3. Specifically, it help in evaluating the joint distribution of the indicator and the construct. Ideally, this joint distribution is barely affected, such that values on the indicator clearly inform values on the construct, like in the first diagram with Indicator A. More realistically, sources of error can have various effects on this joint distribution, like in the diagrams with Indicator B, C, and D. Considering the potential sources of error, researchers may evaluate how they may impact an indicator.

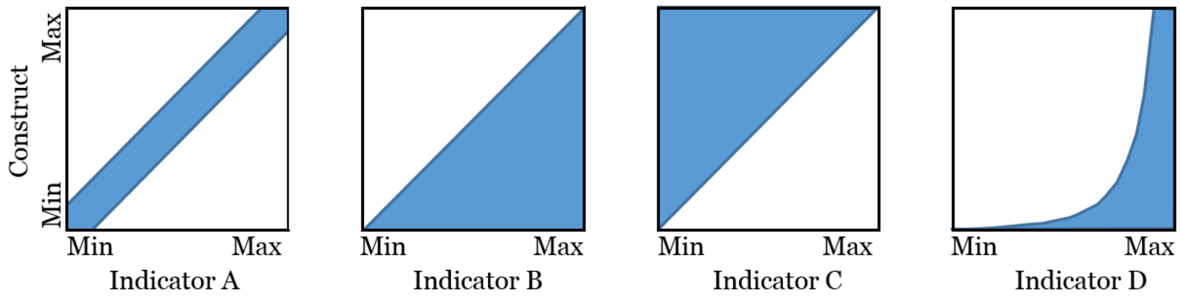

Figure 3: Illustrations of joint distributions of indicators and constructs. Indicator A carries information about the construct across its scale; Indicator B and C carry more information at the lower and higher end of their scales respectively; Indicator D only carries information about the construct at low and medium levels of its scale.

Consider one specific source of error in a measurement where an indicator refers to the frequency of an event that is an effect of the construct, and the effect could also be triggered by other causes. Here, the joint distribution might be similar to that of Indicator B in Figure 3, where high levels of the indicator tell us little about the construct. In other words, the source of error, i.e. the existence of alternative causes, reduces the information corresponding to high values on the indicator. If the indicator would feature in a standard factor analysis, this error would violate the assumption that measurement error is normally and independently distributed as the level of error depends on the level of the construct.

Another, similar example is the bias introduced through socially desirability. When the response scale of an indicator refers to behaviours that vary in their social desirability, the values corresponding to the least socially desirable should be more indicative of the actual behaviours than those corresponding to the others. Similarly, when a certain skill is being tested with a quiz, and a question is very easy, only wrong answers are indicative of the level of skill. This may follow a joint distribution similar to Indicator D in Figure 3.

While evaluating the magnitude of sources of error and the joint interference of multiple sources on the use of an indicator may be difficult, an analysis of the logical links of measurement should at least ease the identification of the most potent sources of error and foster a basic understanding of their effects. Table 1 has shown one such analysis; Table 2 in Appendix 1 shows more examples of logical links of measurement, and how sources of error interfere with them.

The evaluation of indicators should not only provide an understanding of which sources of error interfere in what fashion. It should also clarify how common these sources of error are across the indicators, as some interference may be highly local (specific to one indicator only) and other interference may be more global. Ultimately, this understanding of how and where the most potent sources of error would interfere with the indicators of a measurement, like illustrated in Panel 1A, should prove highly instructive.

4 Dealing with Sources of Error

Ultimately, the impact of an interference on the validity of measurement depends on an interplay of its locality, its magnitude, and the transformation of its impact. The locality refers to where this interference affects the measurement, as illustrated in Figure 2B. This might be at the level of specific indicators—yielding 'local' interference—or related to the entire set, yielding 'global' interference. The magnitude refers to the local severity of this interference. The transformation of its impact refers to how the interference ultimately impacts validity. For example, it might be controlled for when it is explicitly modelled, or it might be aggravated, for example when multiplying indicator values. Considerations of these aspects of interference can help determine whether and where remedies are needed most.

In the example given earlier, where a non-exclusive effect of a construct was to be recorded, one response is to locally adjust the measurement by evaluating whether there may be exclusive effects of a construct. If possible, it could replace the referent of the indicator. The adjusted indicator might have a joint distribution with the construct more akin to that of Indicator A in Figure 3. When this is not feasible, one may attempt to leverage the performance of the indicator at low values, relying on it less with higher values. Perhaps another indicator could complement it, one that carries more information at higher values, like Indicator C in Figure 3. Ideally, the expected distribution of error in the indicators should inform how the records of multiple indicators are used for estimation.

In estimation procedures with reflective indicators, some indicator-specific error can be effectively controlled for, often to the extent this error is normally and independently distributed. However, in many measurements indicators resemble one another or their error is not distributed normally or independently. These indicators may need to be remedied. The best remedies for flawed indicators depend on whether these flaws are local (specific to one indicator) or global (affecting a larger set), as illustrated in Figure 4.

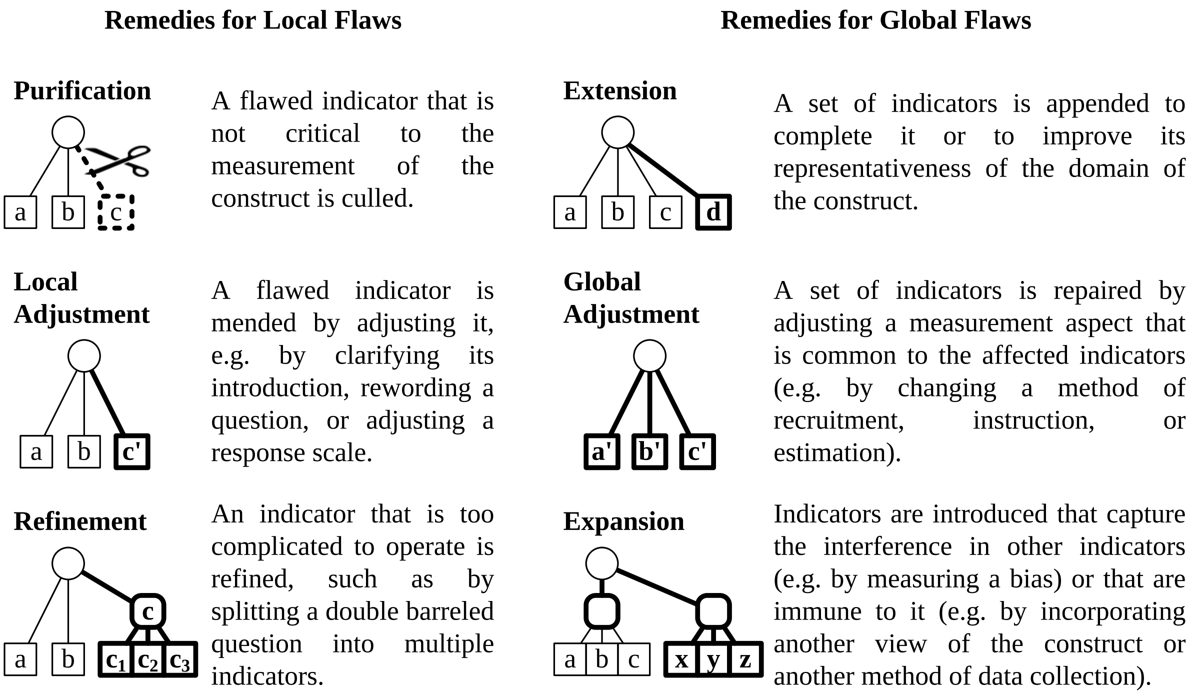

Figure 4: Types of Conceptual Remedies to Local and Global Flaws

A purification, a local adjustment, or a refinement can remedy a local flaw (for more details, see e.g. Clark et al. 1995; Dillman 2000; Haynes et al. 1995; MacKenzie et al. 2011; Tourangeau et al. 2000). An extension, a global adjustment, or an expansion may remedy a flaw that is unspecific to one indicator (Burton-Jones 2009; Chin et al. 2012; Podsakoff et al. 2003; Williams et al. 2010). Extensions and expansions may involve the introduction of indicators that do not repair one flaw, but alleviate the impact of all flaws that are common to the extant set of indicators but not the new one (Podsakoff et al. 2012). Examples of diversifying the indicators include appending self-report with peer-report; a single survey with the momentary assessment method; and a measurement of a construct as a sum of its parts with an indicator that refers to the multiplication of its dimensions. Diversifying indicators works well when new indicators are not clearly inferior to extant ones (Burton-Jones 2009).

While improving the quality of an indicator is generally more preferable than complementing it, it is often harder. Interference tends to stem from a complex and hidden interplay of contextual, idiosyncratic, and circumstantial factors. For example, a researcher may not know fully the conditions under which a construct gives rise to its manifestations. It may be impossible to ensure that participants interpret all questions as intended and report answers accurately. Sometimes, shielding the measurement process against one threat exposes it to another. For example, indirect questioning may help prevent socially desirable responses but it may capture content outside of the construct's domain (Fisher 1993). One may wish to rely on methods different from question-and-answer by recording response times, eye movements, skin conductance, or other physical phenomena (Bradley et al. 1992; Carlson 2013; Segerstrom et al. 2007). As shown in *Table 2*, such alternative means may suffer from their own sources of error that are hard to prevent. Inevitably, potential threats to validity can be found along the entire logical relationship between a construct and its estimate. In Spector's (2006, p230) words, “each operationalization of a variable or method-trait combination carries with it a unique set of potential biases.”

When interference is inevitable, we can alleviate its impact on validity by isolating it in one of multiple indicators (Burton-Jones 2009; Houts et al. 1986; Nunnally et al. 1994). As peculiarities of individual indicators, sources of error often have less impact on the validity of measurement. Depending on the method of estimation, their impact may be diluted, cancelled out, or controlled for (Chin et al. 2012; Harman 1976; Kim et al. 2010).

In sum, the degree with which sources of error undermine the validity of measurement depends on where they interfere, how severely they interfere, and how much of their impact can be reduced. Often, their presence is inevitable and their impact elusive. Yet even while deficient, a thorough evaluation of these sources of error should inform the design of measurement as it can lead to better validity.

5 Discussion

The task of measurement is often split into those before and those after data collection. Guidelines that should help researchers measure their construct of interest have helped them write questions, become more aware of sources of errors, and use estimation techniques. In the latter category, information systems journals have been a prominent venue for advancing ideas. Yet despite progress in each of these areas, the integration between these issues presented in the guidelines has been pedestrian. If researchers are to leverage the value within these guidelines, they are left with the arduous task of mentally reconciling and integrating the insights as they develop measurement.

With this paper, we hope to contribute to a body of literature that provides integral guidelines, reflecting a synthesis of the diverse yet intricately interrelated issues in measurement. Our modified visualisation of measurement models analyses the logical relationship that connects a construct’s estimate with its meaning. This analysis, along with an overview of commonly encountered sources of error, should prove useful in identifying the sources of error that threaten the validity of a measurement. We further hope to help researchers find appropriate remedies for cases of measurement where this threat might undermine the conclusions based on the measurement.

Future contributions to this envisioned body of literature could focus more deeply on the connection between the design of indicators and specific estimation procedures, including common factor models and Item-Response Models. This should help inform researchers the fit between indicator design and estimation procedures. Our hope that such contributions will open up new possibilities for researchers to design measurement in a way that suits the context of that measurement, allowing for valid measurement and reducing the risk of drawing incorrect research conclusions.

6 References

Adcock, R., and Collier, D. 2001. "Measurement validity: A shared standard for qualitative and quantitative research," *American Political Science Review* (95:3), pp 529-546.

Aguirre-Urreta, M. I., and Marakas, G. M. 2012. "Revisiting Bias Due to Construct Misspecification: Different Results from Considering Coefficients in Standardized Form," *MIS Quarterly* (36:1), pp 123-138.

Aronson, E., Wilson, T. D., and Brewer, M. B. 1998. "Experimentation in social psychology" (pp. 99-142). In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology*. New York: McGraw-Hill.

Bagozzi, R. P. 2010. "Structural equation models are modelling tools with many ambiguities: Comments acknowledging the need for caution and humility in their use," *Journal of Consumer Psychology* (20:2), pp 208-214.

Barki, H. 2008. "Thar's gold in them thar constructs," *ACM SIGMIS Database* (39:3), pp 9-20.

Becker, J.-M., Rai, A., Ringle, C. M., and Völckner, F. 2013. "Discovering unobserved heterogeneity in structural equation models to avert validity threats," *MIS Quarterly* (37:3), pp 665-694.

Borsboom, D., Cramer, A. O., Kievit, R. A., Scholten, A. Z., and Franic, S. 2009. "The end of construct validity," In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications*, pp 135-170. Charlotte, NC: IAP Information Age Publishing.

Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. 1992. "Remembering pictures: pleasure and arousal in memory," *Journal of experimental psychology: Learning, Memory, and Cognition* (18:2), p 379.

Burton-Jones, A. 2009. "Minimizing method bias through programmatic research," *MIS Quarterly* (33:3), pp 445-471.

Carlson, N. 2013. *Physiology of Behavior*, (11 ed.) Pearson Education, Inc.: New Jersey.
Carmines, E. G., and Zeller, R. A. 1979. *Reliability and validity assessment*, Sage publications.
Chin, W. W., Thatcher, J. B., and Wright, R. T. 2012. "Assessing common method bias: problems with the ULMC technique," *MIS Quarterly* (36:3), pp 1003-1019.
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.
Cook, T. D., and Campbell, D. T. 1979. *Quasi-experimentation: Design and analysis for field settings*, (Rand McNally.
DeVellis, R. F. 2003. *Scale development: Theory and applications*, Sage Publications: Thousand Oaks, California.
Dillman, D. A. 2000. *Mail and internet surveys: The tailored design method*, (Wiley New York.
Dimoka, A. 2012. "How to conduct a functional magnetic resonance (fMRI) study in social science research," *MIS Quarterly* (36:3), pp 811-840.
Drury, D. H., and Farhoomand, A. 1997. "Improving management information systems research: question order effects in surveys," *Information Systems Journal* (7:3), pp 241-251.
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. 2008. "To prosper, organizational psychology should... overcome methodological barriers to progress," *Journal of Organizational Behavior* (29:4), pp 469-491.
Fazio, R. H., and Olson, M. A. 2003. "Implicit measures in social cognition research: Their meaning and use," *Annual review of psychology* (54:1), pp 297-327.
Fisher, R. J. 1993. "Social desirability bias and the validity of indirect questioning," *Journal of Consumer Research*), pp 303-315.
Goertz, G. 2006. *Social science concepts: A user's guide*, Princeton University Press.
Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. 1998. "Measuring individual differences in implicit cognition: the implicit association test," *Journal of Personality and Social Psychology* (74:6), p 1464.
Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. 2010. "Multivariate data analysis: A global perspective," New Jersey, Pearson Prentice Hall.
Harman, H. H. 1976. *Modern factor analysis*, University of Chicago Press: Chicago.
Harrison, D. A., McLaughlin, M. E., and Coalter, T. M. 1996. "Context, cognition, and common method variance: Psychometric and verbal protocol evidence," *Organizational behavior and human decision processes* (68:3), pp 246-261.
Havlicek, L. L., and Peterson, N. L. 1977. "Effect of the violation of assumptions upon significance levels of the Pearson r," *Psychological Bulletin* (84:2), p 373.
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.
Houts, A. C., Cook, T. D., and Shadish, W. R. 1986. "The person-situation debate: A critical multiplist perspective," *Journal of personality* (54:1), pp 52-105.
Jarque, C. M., and Bera, A. K. 1987. "A test for normality of observations and regression residuals," *International Statistical Review/Revue Internationale de Statistique* (55:2), pp 163-172.
Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2012. "The Negative Consequences of Measurement Model Misspecification: A Response to Aguirre-Urreta and Marakas," *MIS Quarterly* (36:1), pp 139-146.
Ketchen, D. J., Ireland, R. D., and Baker, L. T. 2013. "The Use of Archival Proxies in Strategic Management Studies Castles Made of Sand?," *Organizational research methods* (16:1), pp 32-42.
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.
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.
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.
Mulaik, S. A. 2010. *Foundations of Factor Analysis*, (2nd edition ed.) Taylor and Francis Group: Boca Raton, FL.

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.) McGraw-Hill: New York, NY.

Petter, S., Rai, A., and Straub, D. 2012. "The critical importance of construct measurement specification: a response to Aguirre-Urreta and Marakas," *MIS Quarterly* (36:1), pp 147-155.

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.

Richman, W. L., Kiesler, S., Weisband, S., and Drasgow, F. 1999. "A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews," *Journal of Applied Psychology* (84:5), p 754.

Rigdon, E. E. 2012. "Rethinking partial least squares path modeling: in praise of simple methods," *Long Range Planning* (45:5), pp 341-358.

Rigdon, E. E. 2013. "Lee, Cadogan, and Chamberlain: an excellent point... But what about that iceberg?," *AMS review* (3:1), pp 24-29.

Rigdon, E. E. 2014. "Comment on “Improper use of endogenous formative variables”," *Journal of Business Research* (67:1), pp 2800-2802.

Schwarz, N., and Sudman, S. 1992. *Context Effects in Social and Psychological Research*, Springer-Verlag: New York.

Segerstrom, S. C., and Nes, L. S. 2007. "Heart rate variability reflects self-regulatory strength, effort, and fatigue," *Psychological Science* (18:3), pp 275-281.

Sharma, R., Yetton, P., and Crawford, J. 2009. "Estimating the Effect of Common Method Variance: The Method—Method Pair Technique with an Illustration from TAM Research," *MIS Quarterly* (33:3), pp 473-490.

Spector, P. E. 2006. "Method variance in organizational research truth or urban legend?," *Organizational research methods* (9:2), pp 221-232.

Tourangeau, R., Rips, L. J., and Rasinski, K. 2000. *The psychology of survey response*, (Cambridge University Press.

Verhagen, T., van Den Hooff, B., and Meents, S. 2015. "Toward a Better Use of the Semantic Differential in IS Research: An Integrative Framework of Suggested Action," *Journal of the Association for Information Systems* (16:2), pp 108-143.

Viswanathan, M. 2005. *Measurement error and research design*, (Sage.

Williams, L. J., Hartman, N., and Cavazotte, F. 2010. "Method variance and marker variables: a review and comprehensive CFA marker technique," *Organizational research methods* (13:3), pp 477-514.

Zwanenburg, S. P. 2015a. "How to Tie a Construct to Indicators: Guidelines for Valid Measurement," International Conference on Information Systems, Fort Worth.

Zwanenburg, S. P. 2015b. "Valid Construct Measurement Using Multiple Models," European Conference on Information Systems, Münster, Germany.

Appendix 1

Link	Associated Sources of Error
Construct to Referent	A construct may be inappropriately modelled in terms of its manifestations, effects, constituent parts, dimensions, or other referents of indicators that stand in some relation to the construct (Goertz 2006; Petter et al. 2012). For example, a specification of its parts may be incomplete, superfluous, or it may contain overlapping parts (MacKenzie et al. 2011). A dimensional model may inappropriately specify how dimensions combine to make up the construct (Law et al. 1998).

Referent to Referent	The referent of an indicator may be modelled in terms of its own referents, yielding a hierarchical model (e.g. Edwards 2001; Law et al. 1998; Polites et al. 2012). These links can suffer from the same sources of errors as the links between the construct to be measured and the referents of its immediate indicators.
Referent to Detection	A referent of an indicator may be implemented through physical detectors, such as those that aim to detect heart rate, skin conductance, eye movements, gamma waves, and so forth. Errors may stem from the processes of designing, installing, calibrating, and operating the instrument, depending on the specific apparatus. Further, using them to infer referents of more abstract constructs can be problematic (Dimoka 2012; Fazio et al. 2003). For example, we still know little about how to best infer people's stress, affect, and reward from detections of skin conductance, heart rate variability, and activation of the nucleus accumbens (basal forebrain) respectively (e.g. Carlson 2013).
Referent to Record	A referent of an indicator may be linked directly to a record when relying on past measurements, or 'secondary data', stored in databases, documents, and logs. The sources of error corresponding to this link include all inconsistencies between the original measurement operations and what the result of these operations – i.e. the record – is taken to mean (Ketchen et al. 2013; Markus et al. 2013).
Referent to Stimulus	A referent of an indicator may be inappropriately implemented into a question, picture, sound, or any other linguistic or non-linguistic stimulus. For example, a sound may be inaudible, a question illegible, or the membership of a stimulus to an intended category may be ambiguous (e.g. Dillman 2000; Greenwald et al. 1998).
Stimulus to Response	A good stimulus may not produce the appropriate response, when, for example, the instructions are unclear or evoke an inappropriate degree of time pressure, social pressure, or other forms of stress. The lag with which stimuli are presented may obscure the interpretation of response times (e.g. Greenwald et al. 1998).
Record to Estimate	Records may be inappropriately combined to produce estimates. The mathematical procedures of such combination may involve assumptions that do not hold (Rigdon 2012; Viswanathan 2005).
Record to Model Estimate	Records can be used to test the fit of models or the support for inter-construct hypotheses without separately obtaining estimates for constructs. Sources of error are violations to the assumptions underlying these estimation techniques, related to linearity, normality, or independence of distributions (e.g. Becker et al. 2013; Havlicek et al. 1977; Jarque et al. 1987; Mulaik 2010; Petter et al. 2007; Rigdon 2012).

Table 2: Sources of Error Associated with Logical Links of Measurement

Acknowledgements

I thank Israr Qureshi and Ali Farhoomand for their support in the development of the ideas presented in this manuscript and three anonymous reviewers their help in revising the manuscript.

Copyright