2009

Stopping "How" from Driving "What": Advice on Avoiding Measurement Item Mis-specification

D. Alan Ladd
Washington State University, darin.ladd@email.wsu.edu

Andrew Hardin
University of Nevada - Las Vegas, andrew.hardin@unlv.edu

Mark A. Fuller
University of Massachusetts - Amherst, dean@som.umass.edu

Follow this and additional works at: http://aisel.aisnet.org/icis2009

Recommended Citation
http://aisel.aisnet.org/icis2009/191

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
STopping “How” from Driving “What”: Advice on Avoiding Measurement Item Mis-Specification

Completed Research

D. Alan Ladd  
Washington State University  
Todd 442, P.O. Box 644743  
Pullman, WA 99164-4743  
darin.ladd@email.wsu.edu

Andrew Hardin  
University of Nevada, Las Vegas  
BEH 328 - Box 451234  
Las Vegas, Nevada 89154-1234  
andrew.hardin@unlv.edu

Mark A. Fuller  
University of Massachusetts, Amherst  
121 Presidents Drive  
Amherst, MA 01003  
dean@som.umass.edu

Abstract

The recent debate on measurement item mis-specification has renewed interest in the construct explication process; however, the focus of this debate has been the interaction between latent variables and their measures (“how”), instead of between latent variables and the constructs they represent (“what”). This paper highlights the problems created when framing the formative-reflective discussion in terms of “how” instead of “what.” We argue it is necessary to precede measurement item construction (or appropriation) with careful construct definition and operationalization by focusing first on what is being measured, rather than on how it will be measured. Two solutions for better understanding measurement item construction are proposed: 1) carefully defining assumptions and boundary conditions of a construct in its theoretical context, and 2) carefully defining assumptions and boundary conditions of a latent variable in its measurement context. Recommendations regarding correct interpretation of the measurement process by authors, reviewers, and editors are offered.

Key Words: Measurement, Formative, Reflective, Operational
STOPPING “HOW” FROM DRIVING “WHAT”: ADVICE ON AVOIDING MEASUREMENT ITEM MIS-SPECIFICATION

Introduction

Is it possible that “how” and “where” affects “what” we measure? Methodologists argue that the answer is “yes” (Cook & Campbell, 1979; Cronbach & Meehl, 1955; Shadish, Cook, & Campbell, 2002). Researchers attempting to measure unobservable phenomena must consider not only “what” is measured but also “how” and “where” measurement occurs. Recently, the “how” debate centered on formative and reflective measures (Borsboom, Mellenbergh, & Heerden, 2003; Fornell & Bookstein, 1982; Heise, 1972; Law & Wong, 1999). Unfortunately, the debate appears unresolved (Bagozzi, 2007; Bollen, 2007; Howell et al., 2007a, 2007b), perhaps in part due to the “how” (measurement) overshadowing the “what” and “where” (conceptual) debates.

As described by Sethi and King (1991), vigorous (and continual) methodological debate is important to establish validity of constructs used to create a cumulative IS tradition (Keen, 1980). Methodology concerns more than just measurement—it concerns all researchers’ actions—from idea conception to results reporting. In this study we focus on the methodological process as it pertains to the conception and measurement of constructs. Although content validity is familiar territory in academic debate, it would appear that the correct sequence of events leading from conception to measurement seems to have been overlooked somewhat in the current formative vs. reflective debate.

This paper discusses problems created when framing measurement debates in terms of “how” instead of “what.” Next, the conceptual, operational, and observational planes are described, and the importance of addressing each level of abstraction successively is argued. Two solutions for better item construction are proposed: 1) carefully defining assumptions/boundary conditions of a construct in its theoretical context, and 2) carefully defining assumptions/boundary conditions of a latent variable in its measurement context. Finally, implications for the IS discipline are discussed of framing measurement as theory driven, rather than statistical-convenience driven.

Problem: Is the “How” Driving the “What?”

Recently, spirited debate regarding “formative” and “reflective” measurement items spilled over from psychology into IS literature (Hardin et al., 2008a, 2008b; Marakas et al., 2008). These debates clarified many issues concerning the observational-operational interaction (Diamantopoulos & Siguaw, 2006), but did not address issues arising from the conceptual-operational interaction (illustrated in MacKenzie et al., 2005):

• A certain amount of dogma appears to exist within research circles whose world view is entirely “formative” or “reflective.” As will be shown, rigid adherence to an internally-consistent model may be appropriate; however, when viewed in terms of the conceptual and operational planes, constructs are neither formative nor reflective. Only when viewed in the light of observational level assumptions can they be viewed as such.

• Measures are often inappropriately adapted from previous studies. As will be shown, item appropriateness requires consideration of the original conceptual plane and context within which the items were developed.

• Statistical tools have emerged that allow formative indicator weights to be more easily estimated—especially in Information Systems research where the Partial Least Squares (PLS) technique is gaining popularity (Gefen & Straub, 2005; Fornell & Bookstein, 1982). Researchers are experimenting with new ways of using the tools—some of which are not preceded by proper theoretical justification, and may not be consistent with the designers’ original intent. For example, a researcher who chooses to model a construct “formatively” or “reflectively” based on a post-hoc comparison of item loadings, rather than on a priori theoretical justification.

• Reviewers “legislating from the bench,” asking researchers to reformulate measures and models as “formative” or “reflective” based on their (often incorrect) preconception of a construct as inherently formative or reflective.

• Researchers unquestioningly accepting reviewer suggestions to respecify measures as “formative” or “reflective” in order to achieve publication—avoiding the theoretical debate and issues in favor of convenience.

1 The terms “observational-operational” and “conceptual-operational” are explained in subsequent sections.
Readers, unaware of the theoretical debates, use journals as “research exemplars” to guide future research. While advancing scientific knowledge based upon prior literature is certainly desirable, problems can arise when measurement decisions are based on blind adherence to others’ work, rather than theoretical justification. Clearly, as a field, we desire research that is theoretically grounded and properly executed. Is there a way, then, to change the way we view this debate, and in doing so change how we treat measurement item construction? We argue that the answer to this question is an emphatic “yes,” and that it begins by refocusing our attention from the “how”—or observational plane—to the “what”—or conceptual plane. Further, clarification is required on how the move from the conceptual plane to the observational plane requires passing through the “where”—or operational plane. This discussion naturally depends on what is meant by the different levels at which a construct or theory exists: the conceptual plane, the operational plane, and the observational plane. It is to these issues we now turn.

**Description: the Conceptual, Operational, and Observational Planes**

**The Challenge of Properly Defining and Operationalizing Constructs**

Assuming that concerns of theory appropriation are first addressed (Truex et al., 2006), modern scale construction begins with construct definition, then proceeds to scale design, pilot testing, administration of test, and test validation (Spector, 1992; Churchill, 1979). This paper concentrates on the first two steps: construct definition and scale design, expanding the logic connecting these steps.

Even though the conception of a construct may evolve over time, it is important to define constructs conceptually, because as Spector (1992, p. 13) suggests: a) “many constructs are theoretical abstractions,” b) “construct definition is an essential first step in item measurement that cannot be overlooked,” and c) “a construct cannot stand alone, but only takes on meaning as part of a broader theoretical network that describes relations among many constructs.”

Nunnally (1967, p. 3) adds: a) “. . . measurement always concerns some particular feature of objects,” b) “. . . measurement requires a process of abstraction,” and c) “emphasizing that measurement always concerns a particular attribute . . . forces us to carefully consider the nature of an attribute before attempting to measure it.”

According to methodologians, it is not possible to measure a psychological construct directly (Lord & Novick, 1968; Nunnally, 1967), either because it does not exist outside of its social construction (Burrell & Morgan, 1979; Rosenthal & Rosnow, 2008), or because it has no natural metric (Shadish et al., 2002; Brown, 2006). Instead, researchers measure *something*, represented by a square in Figure 1 that they hope approximates the psychological construct, represented by an ellipse in Figure 1 (Nunnally, 1967). Unfortunately, imprecision in language and symbology betrays researchers because they cannot precisely portray the entirety of any given conceptual model. For this reason, constructs themselves are traditionally defined (Cook & Campbell, 1979; Shadish et al., 2002) at a level of abstraction removed from latent variables used in measurement models (Brown, 2006; Kline, 2005).

Perhaps a better description of what takes place during the construct definition and scale development process is that researchers define a construct (ellipse) in a way that makes *operational* sense to them given the intended context of study. This *operational* definition may be slightly different across studies, and hence one might argue that the exact same dimensions of a construct are not consistently captured. To the extent that operational definitions differ across studies, a researcher must address to what extent construct validity holds for any given set of studies. Perhaps the research community would be better served by adding another symbol to its repertoire, like a square with rounded edges, to remind us that *latent variables are not themselves constructs, but rather, operationalized constructs.*

Figure 1 clarifies that constructs, operational definitions, and measurement indicators are connected by inductive inference, and indicates the appropriate order in which scale construction should occur (from top to bottom).

---

2 Baxter (2009) provides a recent discussion of these issues, along with three illustrative examples.

3 Constructs are, therefore, neither inherently “formative” nor “reflective” (Wilcox et al., 2008).

4 Or, one could consider latent variables *partially operationalized* constructs (Rosenthal and Rosnow, 2008).

5 Inductive inference usually refers to an abstraction from the observational to conceptual level (Churchill, 1990). As logicians conceive of it, though, inductive inference applies equally to abstractions made from the conceptual to observational level. It is in this respect that the term “inductive inference” is used. Logicians prefer “abductive inference,” or “inference to best explanation” to when inferring from observations to theory (Sober, 1991, p. 23).
A construct is: “a concept, model, or schematic idea” (Shadish et al., 2002, p. 506), e.g., “self efficacy,” or “decision quality.” Regardless of whether one’s methodological paradigm is subjectivist, objectivist, or in-between, this definition is useful because it focuses not on whether concepts exist outside of the mind, or whether they are a construction of it; but rather on whether or not a concept can be named or articulated. Social science constructs, at the conceptual plane, are unexplicated, inexact (Lord & Novick, 1968), and therefore, unmeasurable. As soon as one begins to attempt to further define a construct, one leaves the conceptual plane and begins moving towards the operational plane, where construct definitions are context-dependent, and specific to a given nomological net.

Operational Plane

An operational definition is: “an empirically based definition, that is, the meaning of a variable in terms of the operations that are used to measure it or the experimental method involved in its determination” (Rosenthal & Rosnow, 2008, p. 753). In fact, three interrelated inferences must occur to operationalize a construct: a) definition of the construct, b) consideration of its immediate context, and c) explanation of the nomological network in which the construct acts (Schwab, 1980). Assumptions must be made about how the construct acts within its nomological network, and boundary conditions must be set concerning what the construct includes/excludes in a given context. These assumptions and boundary conditions—by restricting a construct’s range and interpretation—increase the probability that the construct can be measured, albeit at the expense of some of its conceptual richness.

---

6 As Rosenthal and Rosnow (2008) conceive of it, an operational definition includes both of the bottom two planes in Figure 1, we distinguish between an operational definition and its measures because of the intuitive distinction between the two as popularly conceptualized, i.e., a measurement model (see Brown, 2006, p. 51). It is possible that in some instances latent variables can be defined with such precision that their measures are obvious (e.g., height or weight); however, as we treat the subject in this paper this would appear to be a highly improbable occurrence.

7 Cronbach and Meehl (1955, p. 291) define nomological networks as “the interlocking system of laws which constitute a theory.” This definition implies a tight conceptual linkage between constructs that operate in a methodological fashion, but also implies that, at the operational level, they share the same context.

8 Some authors (Bagozzi, 2007; Bollen & Lennox, 1991; Edwards & Bagozzi, 2000; Jarvis et al., 2003; MacKenzie et al., 2005) argue theories are more instructive when phrased: “under what conditions does A cause B.”
Regardless of a researcher’s methodological paradigm\(^9\), what becomes apparent from this discussion is that important decisions about a construct are made apart from the conception of measurement. For example, a subjectivist, who perceives the construct “decision quality” as: “the shared conception of a group’s agreement on its satisfaction with a decision outcome” will proceed with a quite different operationalization of decision quality than will an objectivist-positivist who defines decision quality as: “the extent to which the focal decision is objectively correct.” Thus, context necessarily informs and constrains measurement, which occurs at the observational plane.

**Observational Plane**

The observational plane discussed by MacKenzie et al. (2005), refers to the plane of observables (or indicators when used in the context of a measurement model). At this level, observations can be recorded, with the resulting data available for analysis (whether tables filled with numbers collected during an experiment, or a log of recorded sentence fragments in a participant-observation study). Because, in most cases, an inductive inference is required to connect manifest variables with latent variables\(^10\), a researcher must again make more restrictive decisions regarding the importance of confounds that may act to resist the measurement of latent variables as defined in the operational plane. This is where most methodological texts pick up; describing the model connecting the operational plane to the observational plane as a “measurement model” (Brown, 2006; Kline, 2005). Notably, it is at this plane that the formative vs. reflective debate becomes relevant insofar as it concerns the assumptions a researcher makes about how to model construct dimensionality and error at the operational (latent variable) level; as well as how to (or whether to) maximize generalizability of the latent variable outside its original intended measurement context.

Having outlined the challenges posed by construct operationalization, and proposing definitions of the conceptual plane, the operational plane, and the observational plane, it is now possible to discuss the two processes implied by the three plane model: 1) abstracting from the conceptual plane to the operational plane, and 2) abstracting from the operational plane to the observational plane. We will now address these two steps in order, while offering recommendations on proper explication of constructs aimed at reducing the occurrence of measurement item “mis-specification.” We begin with the process of abstraction from the conceptual plane to the operational plane.

**“Mis-specification” Solution #1: Carefully Define Assumptions and Boundary Conditions of a Construct in its Theoretical Context**

When moving from the “what” (conceptual plane) to the “where” (operational plane), assumptions such as a construct’s definition, context, and explanation of how a construct interacts with its nomological network must be made. As each of these conceptual-operational considerations is addressed, a construct becomes more specific and specialized to a given time/place where it will be measured (Diamantopoulos & Winklhofer, 2001). Carefully defining assumptions and boundary conditions at this first level of abstraction in such a way that they are internally consistent greatly reduces the possibility of measurement item mis-specification when abstracting from the operational to observational plane by ensuring measures are related to a theory in its operationalized form.\(^11\)

Understanding assumptions and/or boundary conditions will accomplish three complementary goals: first, it will ensure the researcher is prepared to measure the construct s/he is interested in with as little error as possible; second, accounting for a construct’s definition, context, and explanation of how a construct interacts with its nomological network must be made. As each of these conceptual-operational considerations is addressed, a construct becomes more specific and specialized to a given time/place where it will be measured (Diamantopoulos & Winklhofer, 2001). Carefully defining assumptions and boundary conditions at this first level of abstraction in such a way that they are internally consistent greatly reduces the possibility of measurement item mis-specification when abstracting from the operational to observational plane by ensuring measures are related to a theory in its operationalized form.\(^11\)

\(^9\) It is at this level that subjectivists and objectivists part company in some respects, because the subjectivists might argue that any definition of a construct must account for its socially-constructed nature, while an objectivist might argue from one of two perspectives: a) positivists might argue that the construct is a real entity apart from its conception in the mind of a human, or b) functionalists might argue that the construct is the statistical convenience used to aggregate like-features of existing sets of real entities. Interestingly, subjectivists and functionalists may reach many of the same conclusions in practice—though they may disagree in philosophy.

\(^10\) One notable exception is pure “path modeling,” where observations are compared to other observations; however, one must note that even in path modeling an inference is made about the treatment of error variance (Kline, 2005).

\(^11\) In other words, if a reviewer suggests to an author that the measurement items used in a paper are “mis-specified” as formative when they should be reflective, or vice-versa, we argue that at least part of the embedded meaning of this phrase is that the author is using the wrong measures based on the theory as it is operationalized. Perhaps it is more useful to replace the phrase “measurement mis-specification” with a description of where the reviewer feels the author’s disconnect in logic occurred when moving from the conceptual plane to the observational plane.
it will assist reviewers in following the logic of the researcher with respect to the measurement items chosen; third, it will establish the researcher’s contribution within the larger IS domain. Journal space is at a premium, and therefore, a discussion such as the one proposed must be made as brief as possible; however, we wish to make clear our recommendation that this step always be taken, even if it is brief.

For example, if a researcher proposed to measure computer self-efficacy (CSE), he/she would want to specify whether it was general CSE (Compeau & Higgins, 1995) or task-specific CSE (Marakas et al., 1998). This might include a discussion of how the CSE used in this study is similar to, and different from, previously measured CSE constructs. Further, the researcher would then wish to iterate which tasks CSE was meant to be generalized to (e.g., spreadsheet use in a business setting, versus an informal setting). If adopting a CSE measure from another study, a researcher must carefully reposition the latent variable as it was initially measured into its new nomological network, and ensure that it is consistent with that new context. It is important to note that even a slight change in definition, such as changing the task domain of task-specific CSE from spreadsheet operations to computer-aided decision support, may require either: a) significant additional justification, or b) measurement item revalidation. As described in this section, failure to perform these two steps might result in measurement of a latent variable other than the one that was intended to be measured.

Clearly, these recommendations are not new; however, one need only look to the TAM literature for examples of measures being adapted to new setting with little more than a reporting of reliability. The purpose of our discussion is to reemphasize the importance of defining “what” we are measuring in determining “how” we will measure it, and the definitions can and do change across contexts. If the mere act of defining a construct changes “what” is measured, then this step must be carefully considered, and must precede subsequent steps, in order to ensure that construct measurement is consistent with our theories. The next step, then, is to decide how to best choose measurement items that appropriately model an operationalized latent variable in its context, and hence make the appropriate assumption about how to model the error that remains. It is to this we now turn.

**“Mis-specification” Solution #2: Carefully Define Assumptions and Boundary Conditions of a Latent Variable in its Measurement Context**

Earlier we established that a construct’s operational definition drives the measurement model, which then drives tool selection. Moving from the operational plane to the observational plane, three additional _operational-observational_ interactions must be considered: latent variable dimensionality, generalizability, and error assumptions.

**Latent Variable Dimensionality**

As MacKenzie et al. (2005) explain, one of the most critical errors that can be made is to misconceive latent variable dimensionality (e.g., a multidimensional latent variable is treated as unidimensional or vice versa). While we agree that this is a critical error, we also believe it is critical that consideration be given to the focal nature of all latent variables in a nomological network. This gives rise to two inter-related compromises: a) incomplete re-validation of non-focal latent variables, and b) replacing a multidimensional latent variable with an omnibus measure.

First, if a latent variable is not central to a theory, for example a control variable intended to account for excess variability, it might be possible to use a proxy. For example, the use of enrollment in school lunch programs is often used as a surrogate for socioeconomic status. Because this implies incomplete validation of a measure in its context, construct dimensionality may be unknown, and applicability to the context limited. This practical consideration is only warranted if the researcher believes that any confounds created through misappropriation will not adversely affect the primary effects of interest. In this case, it is incumbent upon the researcher to acknowledge this limitation of the study, or else risk propagating a sub-optimal measure to future researchers.

Second, regardless of a latent variable’s centrality, if it is possible to conceive of a latent variable in different ways at this stage, as we argue it is, one must consider the practicality of measuring the latent variable. This is necessary because researchers rarely have unlimited resources with which to capture all possible: a) operationalizations of focal latent variables, b) dimensions of focal latent variables, or c) components of a nomological network. For these reasons, it is likely that some latent variables may be more fully captured than others. For example, if a latent variable is seen as multidimensional, with three components, yet a researcher can only spare three questions in a survey, the researcher must carefully consider whether three omnibus measures are more appropriate than one question for each of the three dimensions. Once again, it is incumbent upon the researcher to carefully consider how to allocate scarce resources, and equally important to acknowledge resource allocation as a limitation of the study.
**Generalizability**

The problem of misappropriation of theories and measures out of their original context is well-documented (as discussed by DiMaggio, 1995). For this reason, authors should carefully consider the needs of future researchers with respect to the reuse of their measures and theories. This is a two-pronged problem: a) carefully iterating the original context and nomological network of the latent variable, and b) carefully considering the measures chosen. Carefully iterating the original context and nomological network of the latent variable—through assumptions and boundary conditions—helps future researchers understand whether or not they are faithfully appropriating latent variables. This, in turn, increases generalizability in the sense that explicit boundaries help a future researcher separate theory from its context. Unfortunately, even if latent variables are faithfully appropriated into new contexts, any modification of the measures chosen decrease generalizability; i.e., changing measures alters the meaning of latent variables, even if the labels of those latent variables stay the same (Diamantopoulos, 2009). Item modification may be considered, but only if this modification helps adapt a latent variable into a new context, and only if the possible change in meaning is acknowledged and/or investigated.

**Error Modeling**

In order to ensure internal consistency as well as generalizability, a researcher must decide which of the three methods of dealing with variance in indicators, and their associated error (or disturbance) terms, that s/he will use. In each case, assumptions are made that may reduce the generalizability of a latent variable in different ways; therefore, the researcher must carefully consider which error types are less damaging to a latent variable’s internal and generalizability in any given study. Figure 2 depicts a simplified representation of formative measures (1a), reflective measures (1b), and a combined model with both formative and reflective measures (1c).

---

**Formative measurement error (Figure 2a) and its implications**

In a formative measurement model (Bollen & Lennox, 1991; Petter et al., 2007; Heise, 1972), error is defined as any deviation from full variance accounted for in the latent variable by the formative indicators, indicated by the “disturbance term” (represented by the ellipse labeled \( D_x \) in Figure 2a). In this case, the disturbance term represents the existence of two possible conditions (implied by Diamantopoulos et al., 2008): 1) a latent variable that overlaps the focal latent variable that was not measured, and 2) the exclusion of a formative indicator that should have been included but was not. It does not include measurement error, because formative measurement assumes that \( I_1 \) and \( I_2 \) perfectly comprise \( X \) (a broad assumption practically, if not conceptually).\(^{12}\)

The implications of these assumptions are fourfold. First, if an overlapping latent variable is not measured, the researcher has failed to capture the appropriate nomological network for the focal latent variable. This is a significant problem according to many methodologists because formative indicator weights are dependent on the relationship of the formative measure and its nomological network (Bagozzi, 2007). Second, if the researcher fails to include a complete census of measures, it is possible that the latent variable will act differently within the nomological network, because omitting indicators changes the definition of the latent variable. Unfortunately, in

---

\(^{12}\) In some formative models, also termed “composite” models (Bagozzi, 2007; Bollen, 2007), this disturbance is not modeled because a variable is composed of—or defined by—its measures.
this second situation there is no way to estimate the size of the problem; therefore, in the absence of additional corroborating studies, the generalizability of the latent variable will be somewhat suspect. Third, because error is not modeled at the item level, it is entirely possible that the latent variable measured is different (more precisely, less) than it appears because outside influences are causing the measurement items to act differently than they should—again a possible failure of the researcher to properly specify the nomological network surrounding a measurement model. Fourth, if a formative latent variable is defined in terms of its indicators, and indicators are missing per the disturbance term, the formative latent variable may not be generalizable to other settings.

Reflective measurement error (Figure 2b) and its implications

In a reflective measurement model (Bollen & Lennox, 1991; Lord & Novick, 1968), the error terms (E in Figure 2b) represent measurement error, and are represented by an ellipse (latent variable). In this case, assuming a unidimensional latent variable (a broad assumption practically, if not conceptually), any deviation from a perfect factor loading is due to “unique variance” encompassing: 1) random error variance and 2) systematic, indicator-specific variance (latent factors, external to the model, that only affect one specific indicator) (see Brown pp. 13).

The implications of these assumptions are again fourfold. First, failure to model error at the latent variable level presumes unidimensionality. Because it is possible to test for unidimensionality, a researcher should confirm its existence by: a) using an exploratory factor analysis on a separate data set, or b) if a measure is appropriated, a confirmatory factor analysis on the new data set with a theoretical justification for why the contexts studied were similar (Brown, 2006). Second, existence of a unidimensional latent variable does not imply that the latent variable is atomistic (indivisible). From a functionalist perspective, it is possible to argue that the shared variance between the indicators is just that—shared variance. For this reason, researchers must ensure the assumptions underlying the unidimensional latent variable are made explicit—reserving the right of future researchers to decompose the latent variable if necessary. Third, to the extent that unique variance is due to systematic causes, the measures captured something that is not relevant to the focal study. Additional work is therefore required to enforce the boundary conditions of the theory as they apply in the measurement model. Fourth, because random error variance and systematic variance cannot be separated with current tools, it is not possible to determine the extent to which the theory and sample are represented by the measures. In either of these cases, generalizability is suspect.

Regardless of how one chooses to model error, any failure to account for all possible sources of error systematically reduces the explanatory power of a model. Therefore, choice of one type of measurement model over another implies acceptance of the error type not modeled. Unfortunately, a model that accounted for all sources of error might prove unwieldy, and therefore be of limited use. Until statistical theory and techniques are developed that can better model the entirety of a researcher’s intent, the other option is to continue the use of these flawed models within the contexts in which they were originally designed—in other words, by carefully iterating and explicating the assumptions/boundary conditions of our models. If, as described above, assumptions/boundary conditions are properly addressed, the choice of which error terms to model, and which error terms to disregard, might become more obvious. In either of these cases, assumptions and boundary conditions are better investigated by starting with a combined model, then paring off the pieces that are less problematic to a given study.

Starting with a combined model (Figure 2c)

As shown by Petter et al. (2007), many latent variables are measured both formatively and reflectively, implying that practical considerations drive measurement item specification. As discussed above, considerations of a latent variable’s definition, context, nomological network, dimensionality, centrality to the focal theory, and generalizability of the latent variable and its measures all precede decisions about what type of indicators to use. Therefore, when moving from the operational plane to the observational plane, it might be useful to begin by

---

13 Loadings, and hence the error term are calculated to increase the loading of the focal latent variable on downstream latent variables (see, for example Franke et al., 2008; and Hardin et al., 2008a).

14 If the authors had a preference for usage of the term “mis-specification,” this would be it: “mis-specification” of the context and/or nomological network surrounding a latent variable.

15 Thus, it could be argued that use of an “error term” at the latent variable level is inconsistent with the principles of formative measurement, because the mere existence of an error indicates an incompletely defined construct.
conceiving of the focal latent variable as a combined model (shown in Figure 2c), instead of formative or reflective. Then, if theoretically justified or practically required, it might be possible to pare the model into a subset of its possible “formative” or “reflective” components. This process forces one to explicitly justify assumptions and boundary conditions surrounding the measurement context, and to justify modeling some error sources instead of others. In fact, in some contexts it might be appropriate to actually model latent variables with a combined model.

In this section we discussed how assumptions and boundary conditions set at the observational level are in part driven by a latent variable’s nomological network, and also discussed how inattention to the nomological network might lead to an internally inconsistent theory. We will now present an example using a common IS construct—decision quality—to further illustrate the concepts outlined thus far.

**Example: Decision Quality**

**Construct Definition**

Consider the construct “decision quality.” Because it is not an existing entity, but rather an attribute, it fits into a discussion of social sciences constructs. A decision is an action; however, it is also an outcome. Therefore, in its simplest form, decision quality refers to the quality of a decision outcome. In this sense, one must define what decision is of interest in the context of a given study; in other words, it is not helpful to try to study all decisions—for example, one must decide between the general class of decisions exemplified by explicit task-oriented decisions, versus the class of decisions exemplified by the autonomous decisions required to steady a glass of water in one’s hand. Next, one must define what is meant by “quality” with respect to a decision; in other words, only one instantiation of the attribute is of interest. Unfortunately, this still is not precise enough, because decision quality can be conceived of in many ways. For example, decision quality can be modeled in terms of absolute outcomes, e.g., success or failure, profit, performance, individual members’ satisfaction with decision outcomes, etc. Likewise, decision quality can be modeled in terms of process quality (Carroll & Johnson, 1990), e.g. quality of each step in the decision making process. What is clear from this mental exercise is that a substantial number of assumptions must be made in order to define a social sciences construct. Also clear from this mental exercise is that a definition is not complete until its context is included.

**Construct’s Immediate Context**

In Figure 3, decision quality is shown in a hypothetical context. In this context, decision quality is seen as the result of a fit between decision requirements and the ability of a decision support system (DSS) to support those decision requirements. This context allows the researcher to draw a boundary condition around the theory specifying that the decisions of interest are only those made as a result of certain specified decision requirements supported by a DSS. This boundary condition furthers the understanding of where the focal decisions occur, but also begins to limit the applicability of this construct to other contexts. For example, this theory may not be applicable to decisions made without computer assistance. It is not yet clear whether groups or individuals are making decisions; therefore let us assume that the context is decisions made by a group. Let us also assume that multiple decisions are being modeled—in other words, performance outcomes are not the result of a single decision, but of the aggregation of the results of multiple decisions. Because the decisions made are in support of a task, we can now draw a boundary condition around the decisions of interest: only those in support of the focal task performance.

![Figure 3: Conceptual Depiction of “Decision Quality” in a Given Context (Example)](image-url)
Construct’s Nomological Network

To consider decision quality as it exists in the theory shown in Figure 3, one must think about it in additional detail. For example, what other theories must be integrated or deconflicted with this theory. One obvious answer is Task-technology Fit (TTF) (Goodhue & Thompson, 1995)—the model on which this example is based; however, exclusion of the usage construct from the model, and inclusion of decision quality, implies that additional assumptions were added to the TTF model that must be explained. For example, we could make an explicit assumption that usage of the computer system is mandatory (vs. optional usage in TTF), as implied in the model. This assumption also rules out Universal Technology Acceptance Model (UTAUT) (Venkatesh et al., 2003) concerns that must otherwise be addressed within the TTF framework if computer system use is optional.

Finally, let us assume that the time frame required for the decision and task performance are short enough that it is not possible for the decisions or the performance to affect the decision requirements, decision support system capabilities, or the fit between the two; in other words, there is no feedback loop modeled in this system. This assumption helps draw the distinction between this theory and theories like adaptive structuration theory (AST) and structuration of technology (SOT) that assume feedback processes (DeSanctis & Poole, 1994; Orlikowski & Robey, 1991). Again, these assumptions and boundary conditions, shown in Figure 4, limit the applicability of this model to short-term decisions made under the conditions of mandatory system usage.

Upon considering the above, we are left with a latent variable as it exists within a context, and within a nomological network. What should be apparent is that, in order to define what we wished to measure, it was necessary to consider where it would be measured, but not how it would be measured. In fact, as predicted, the formative-reflective discussion was completely excluded from this discussion. It is to this next task—how we measure decision quality—that we now turn.

Latent Variable Dimensionality and Error Modeling

Once conceptual and operational concerns are addressed, the researcher can begin to consider measurement. The process of identifying and distilling measurement items is actually a three-step process: writing measurement items (both formative and reflective), performing a mental exercise considering latent variable dimensionality and error modeling, and item revision. Only after this exercise, we argue, is it appropriate to proceed to item testing. To illustrate this process, we begin by showing an initial attempt at item-generation as shown in Figure 5 (a combined model that presumes step 1 was previously accomplished). This hypothetical combined model uses measures that could be used together, or separated into formative (Figure 5a) or reflective (Figure 5b) components.

On the formative side of the model (5a), decision quality is shown as a latent variable measured with three formative items. In an attempt to reduce the disturbance term $D_Q$, the researcher must consider whether or not such measures are supported theoretically by the operationalized latent variable, form a census of the latent variable, or form more than a census of the latent variable, then modify or add to the items accordingly. For example, based on the context of this study, it is possible that $I_2$ is too broad—unsupported by the operationalized latent variable, and possibly introducing additional variance not intended—and must therefore be decomposed into resource types specific to the focal decision (such as human resources and computing resources). Further, if decision correctness, time, and
resources can be used to model decision quality, one must ask if there is any benefit (other than model identification) in asking about team or expert perceptions of decision quality (Figure 5b). The answer to this might be constrained by time and space allotted, although considering the centrality of this latent variable to the theory a researcher might decide to collect this information, as well.

On the reflective side of the model (Figure 5b), “decision quality” is shown as a unidimensional latent variable measured with three reflective items. Before attempting any measurement, the researcher may wish to ask if there is sufficient conceptual reason for using the items selected based on possible sources of error. For example, what happens if experts’ ratings in I₄ disagree with team members’ ratings in I₃? Perhaps experts might use a heuristic that emphasizes time spent to reach a decision, while team members base their perceptions on decision correctness. If so, then the latent variable is actually multidimensional and may be better served with two expert judgments and two team perceptions—one each for time spent and decision correctness. Finally, what about “satisfaction with the team’s decision”—is this measuring the same thing as “decision quality?” Again, a third dimension might be offered based on the argument that satisfaction with a decision is not the same thing as a quality decision. Because decision quality is central to the proposed model, the researcher in this example might want to investigate these assumptions further before proceeding to a pre-test of the measures, but already it appears as though item I₆ might not capture a unidimensional latent variable.

This example illustrates problems researchers might encounter when attempting to move from an operational definition to the observational model. Significant error or disturbance terms are clues available only after the fact; therefore, it is important for the researcher to consider what these errors/disturbances may mean in the context of the model and attempt to minimize them (or at least make an attempt to understand their sources).

**Generalizability**

Using item I₆ (Figure 5), it is possible to make one final point regarding model specification. Considering the nomological network of the theory depicted in Figure 4, the researcher may wish to ask if it justifiable to include team satisfaction as a measure, considering the model’s immediate nomological network and assumptions. For example, the immediate downstream variable of interest is “performance outcomes,” and the assumption made earlier was that this theory only applied to “short-term” decisions; therefore, it is possible that inclusion of team satisfaction is irrelevant in this context. Also, if one includes satisfaction as a measure, this assumption moves the entire theory towards inclusion of longer-term outcomes, and therefore, towards the inclusion of adjacent theories (such as UTAUT, if systems use is voluntary, and SOT). Explanation of these assumptions and boundary conditions will help establish the contexts under which the items can be properly appropriated, and those to which they should not be appropriated.

As shown in this brief example, assumptions and boundary conditions affect every decision made from the definition of a construct to the measurement items chosen. Concerted effort is required to think about how decisions
made on one plane of a construct’s existence affect other planes in order to achieve during the development of an internally consistent theory. In addition, both the “where” of a construct’s context, and the “how” of its measurement items affect “what” is measured; therefore, a researcher wishing to measure a construct must consider how increasingly restrictive assumptions and boundary conditions affect a construct’s generalizability.

Conclusion

If measurement item mis-specification is indeed as common as described, and as damaging as argued, then immediate attention is required to correct the course of IS scholarship by protecting its constructs from misinterpretation. Therefore, in an effort to reduce measurement item mis-specification, this paper addressed a paradigm for approaching construct definition and measurement item construction that avoids many of the pitfalls leading to mis-specification. Specifically, the following observations and associated solutions were proposed:

First, the current formative/reflective debate is symptomatic of a larger problem with construct definition and operationalization, namely the misconception that constructs and measures are context-independent. In addressing this problem, a distinction was made between the conceptual plane, the operational plane, and the observational plane. Therefore, this paper recommends that theory should drive measurement, not the opposite. While this suggestion may not appear particularly provocative to some readers, it is important given the current confusion surrounding the use of formative versus reflective measurement that the discussion of proper measurement procedures be revisited. (The debates on rigor versus relevance and theory building serve as prior examples of this form of discourse). Related to this recommendation, assumptions and boundary conditions affect latent variables and their measurement items; therefore, caution is suggested when appropriating measures from previous studies.

Second, by explicating the logic connecting construct definition to measurement item construction, this paper recommended two solutions, outlined as a process in Table 1. First, researchers were advised to carefully define the assumptions and boundary conditions of a construct in its theoretical context. The process of considering a construct’s definition, immediate context, and nomological network prepares a researcher to measure the focal theory with as little error as possible. Second, researchers were cautioned to carefully define the assumptions and boundary conditions of latent variables in their measurement context. The process of considering a latent variable’s dimensionality, generalizability, and error, prepares a researcher to a) choose measurement items that appropriately model an operationalized latent variable in its context, and b) make the appropriate assumption about how to model the error that remains. These two solutions were illustrated using the concept of decision quality as an example.

Third, the implications of this research may extend to reviewers and editors, as well. For reviewers, it might be helpful to move beyond just the identification of mis-specified constructs, and to aid researchers in understanding where mis-specification errors occur. Next, reviewers are reminded to pay close attention when measures are appropriated from other studies to ensure they are used consistently with their original intent, including their associated assumptions and boundary conditions. For editors, the long-held policies of premiere IS journals that consider measurement development papers to be lacking substantial contribution may be detrimental to the discipline. Instead, theory-driven pieces that explicate constructs—including measurement development papers—should be encouraged. This change in policy will encourage additional rigor during measurement development—and more thoughtful appropriation of previously developed measures—by properly rewarding such efforts with publication in top tier journals.16

Finally, considering the discussion of the process of construct explication and measurement item development, this paper argued that it might be time to re-think the term “construct mis-specification”. If there is mis-specification in research, it would appear that it begins at the conceptual plane, and continues through the conceptual-operational interaction—long before the observational process begins.

---

16 Another approach may be to establish a premier IS journal focused on methods, similar to the role of Psychological Methods in the Psychology discipline.
<table>
<thead>
<tr>
<th>Explication Step</th>
<th>What</th>
<th>Why?</th>
</tr>
</thead>
</table>
| 1. Construct Definition | • Internally-consistent  
  • Acknowledge limitations | Constructs are unmeasurable. The act of definition begins to establish a focal latent variable (FLV). |
| 2. Immediate Context | • In which situations/populations does FLV exist/apply?  
  • In which situations/populations does FLV not exist/apply?  
  • What tradeoffs were applied to use the FLV in this study’s context? | Context is necessary to define a measurable LV, but context restricts how a latent variable behaves, and reduces generalizability. |
| 3. Nomological Network | • What other theories/LVs exist in this context? Under what conditions?  
  • How do the other theories/LVs affect this construct?  
  • Are there unmeasured LVs that also affect the construct?  
  • What theories/LVs do not apply to this construct in this study?  
  • How central is the LV to the theory?  
  • Does the LV exist, or is it a functional convenience? | Defining a nomological network helps describe how a LV will behave, but also further constrains its applicability to diverse contexts. If LV is a functional convenience, it may not be applicable to different contexts. |
| 4. Dimensionality | • Is latent variable central to theory?  
  o If YES:  
    ▪ What are the different dimensions?  
    ▪ Can I measure all different dimensions? | Dimensionality drives measures. Some LVs might be defined as functional conveniences (to account for shared variance or combine into an omnibus measure for convenience). |
| 5. Generalizability | • What about the measurement context is unique?  
  • How applicable/generalizable is FLV to other contexts?  
  • Are the measures chosen context-specific?  
  IF APPROPRIATING MEASURES  
  • Are measures in their original context/nomological network?  
  • If not, how does this change the ability to measure the FLV?  
  • Do adapted items still capture the original FLV, or something different?  
  • Acknowledge limitations | LV definition as well as causal relationships may change in different contexts. Informing future researchers of boundary conditions of the theory/LV/measures will better help them appropriate measurement items. |
| 6. Error Modeling | • Start with combined (conceptual) model  
  • Define formative items for FLV  
    o Error sources unaddressed?  
    o How error contaminates FLV?  
    o Acknowledge limitations | Formative and reflective measurement items capture LVs differently, making different assumptions about how to treat measurement error. Considering both types force researchers to address these error types, and to address how the items chosen minimize or accept the error types most applicable to a given LV in a given context/nomological network. |
  • Define reflective items for FLV  
    o Error sources unaddressed?  
    o How error contaminates FLV?  
    o Acknowledge limitations  
  • Does combined model better address error sources?  
  IF APPROPRIATING MEASURES  
  • Follow steps 1-6 in their entirety, then:  
    • Are appropriated measures best possible items?  
    • If not, then produce/adapt items to reduce error contamination of FLV |
References


