**Thinking about Measures and Measurement in Positivist Research:**

**A Proposal for Refocusing on Fundamentals**

Andrew Burton-Jones
The University of Queensland Business School
and Centre for the Business and Economics of Health
St. Lucia, Qld, Australia
abj@business.uq.edu.au

Allen S. Lee
School of Business
Virginia Commonwealth University
Richmond VA, USA
allenslee@alum.mit.edu

**ABSTRACT**

We challenge two taken-for-granted assumptions about measurement in positivist research. The first assumption is that measures and measurements are relevant for quantitative, but not qualitative, research. We explain why they apply to both types of research. The second assumption we challenge is that existing measurement practices are unproblematic, even if researchers sometimes vary in how well they enact them. We explain why current norms (both espoused and enacted) are deficient in some important ways because they fail to emphasize the fundamental issues of measures and measurements. Drawing on symbolic logic, we provide a framework to help positivist researchers to assess efforts in measuring and measurement regardless of their quantitative or qualitative orientation. The framework provides more parsimonious and broadly applicable guidance than available to date and suggests the need to refocus on measurement fundamentals.

Keywords: Research methods, quantitative, qualitative, measures, measurement.

**Accepted at *Information Systems Research***

**Submitted for publication**

**January 2017**

**This is the last version of the paper submitted prior to its acceptance and publication.**

**It is provided in open access in accordance with Australian Research Council policy.**

**Funding:** This research has been made possible by the support of the University of Queensland Business School and the Australian Research Council [FT130100942].

## INTRODUCTION

We challenge two assumptions regarding measurement that we believe are taken for granted in much positivist research. First, we believe many researchers assume that measurement is relevant for quantitative, but not qualitative, positivist research. Although some qualitative guidelines discuss measurement (Eisenhardt, 1989; Pare, 2004), others do rarely (Yin, 1994) or not at all (Shanks, 2002). Measurement guidelines typically exclude qualitative research (Bagozzi, 2011; MacKenzie et al., 2011) or state that measurement is inherently quantitative (Michell, 1997). General descriptions of research (such as the page for quantitative research on Wikipedia that contains 17 references to measurement, and the page for qualitative research that contains none) reinforce this belief. Not surprisingly, positivist qualitative research studies often exclude reference to measurement at all (Sarker and Lee, 2003) (see our Supplementary Appendix for further details).

The second assumption we challenge is that existing measurement practices are unproblematic. One indicator of this assumption has been the long-held acceptance of basic concepts for validating measures, such as convergent validity, discriminant validity, nomological validity, and scale reliability. Another indicator has been the long-held acceptance and gradual expansion of existing guidelines, e.g., from Churchill (1979) to Straub (1989) to Straub et al. (2004) to Mackenzie et al. (2011), essentially agreeing upon and expanding upon an accepted script (Grover and Lyytinen, 2015).

We argue against the first assumption. We believe quantitative and qualitative researchers have been missing an opportunity to learn from each other, not realizing they both engage in the same practice. This insight is also vital for mixed-methods researchers who engage in measurement with both methods but have little guidance for doing so.

We also argue against the second assumption because existing practices, despite their strengths, fail to address the full landscape of measurement issues and fail to prioritize their fundamental aspects. We believe that confirmatory quantitative researchers, in particular, are caught in a bind. Due to the gradual accumulation and refinement of existing practices, these researchers face an almost unworkable number of tests to comply with. MacKenzie et al. (2011) summarized best practice with a 10-step process in which several steps had over 10 sub-steps or considerations. Bagozzi (2011 p. 288) writes: "There are so many ways …and so many issues to consider … that we might be apt to throw up our hands in frustration." We will show that many of the existing practices fail to address the fundamental issues involved. Rather than continuing to expand and refine existing practices, a shift in emphasis is required—a new framework is needed that helps researchers to understand the common fundamentals they need to address. We will show there are only a small number of fundamental issues yet the practices that address them are seldom used.

We direct our arguments towards *positivist* research. Because there are many varied definitions of positivism, we make our ontology and epistemology explicit. Ontologically, we commit to the existence of an objective reality. Epistemologically, we commit to the justification of knowledge via logic. Hofweber (2004) writes: Logic [deals] only with the most generally valid forms of reasoning … It applies no matter what one is thinking or reasoning about." Thus, it applies to both quantitative and qualitative research.

We are not aware of any detailed proposals for measurement in quantitative and qualitative IS research. Burton-Jones's (2009) proposal for minimizing method bias provided an initial step, but we go beyond his focus on measurement accuracy to include the quality of measures too. We will argue that measurement practices in IS research, both espoused and enacted, need improvement. We hope our paper will ignite dialog on this issue and help researchers to rethink their practices and undertake the slow, ambitious, and programmatic work required to address the fundamental issues.

## DEFINITIONS AND THEIR LOGICAL FOUNDATIONS

We build our definitions using the following terms: statement variables, statement constants, individual variables, and individual constants. Klenk (1994 p. 110, pp. 196-199) explains the general case of what "variable" and "constant" mean. We incorporate and extend Klenk's explanation. Because of the unfamiliarity of these terms to some IS researchers, we provide several examples (see Table 1).

**Table 1: Examples of Variables and Constants**

| Example | Individual variable | Individual constants | Example | Statement variable | Statement constant |
|---|---|---|---|---|---|
| **1** | *x* | 3 | **2** | *x+y = y+x* | 3+4 = 4+3 |
| | | 4 | | | 7+1 = 1+7 |

| | | | | | |
|---|---|---|---|---|---|
| **3** | Manager | Roy | **5** | A manager will resist new IT | Roy will resist the new Oracle enterprise software |
| | | Don | | | |
| **4** | New IT | New Oracle enterprise software | | | Don will resist the new Microsoft email system |
| | | New Microsoft email system | | | |

**Example 1.** Consider the algebraic symbol $x$. It is appropriately called a *variable* because the value that it takes can vary. For instance, "$x$" can take values, known as *constants*, such as "3" or "4."

**Example 2.** Consider the formula $x+y=y+x$. It is a *statement variable* with two *individual variables*, x and y. It can take values, such as "3+4=4+3" or "7+1=1+7," each of which is a *statement constant*.

**Examples 3 and 4.** Consider the term "Manager." It is a word rather than a number, but it too can be called a *variable*. It could take values, such as "Roy" or "Don," as a *constant*. Likewise, the *individual variable* "new IT" might take a value such as "new Oracle enterprise software" as an *individual constant*.

**Example 5.** Another *statement variable* is "a manager will resist new IT." It contains two *individual variables*, "manager" and "new IT." The *statement variable* "a manager will resist new IT" can take a value, or *statement constant*, such as "Roy will resist the new Oracle enterprise software." The *statement constant* contains *individual constants*, such as "Roy" and "New Oracle enterprise software."

We introduce these terms for three reasons. First, compared to any other set of terms we are familiar with, they have the extremely valuable property of not being tied to any single research orientation, whether qualitative, quantitative, exploratory, or confirmatory. Second, they provide very straightforward definitions for concepts associated with measurement. This is important because such concepts are notoriously difficult to define (Borsboom, 2009; Lissitz, 2009). Our definitions follow:

- **To measure** is to assign, to a variable, a value or constant. This could be to assign an individual constant to an individual variable or to assign a statement constant to a statement variable.
- A **measurement** is the individual constant or statement constant.
- A **measure** is the individual variable or statement variable.
- An **empirical referent** is the real-world phenomenon that both a variable (as a measure) and a

constant (as a measurement) refer to.

The third reason we introduce these terms is that they allow us to propose a more holistic framework for thinking about measurement than currently available. The most common definition of measurement stems from Stevens (1946): *the assignment of symbols (such as numbers or words) to objects or events according to rules*. Despite its wide applicability, most researchers interpret it narrowly, focusing on the assignment of *numbers* to *attributes* of an object (Nunnally and Bernstein, 1994 p. 4). This narrower view reflects only the quantitative assignment of an individual constant to an individual variable. It omits both the possibility of assigning qualitative symbols and the possibility of assigning statement constants to statement variables. Our definition allows us to offer a more comprehensive measurement framework that allows us to account for both qualitative and quantitative research and the measurement of relationships not just constructs.

We briefly note here that directly assigning a constant to a variable, as in Table 1, is but one means by which a variable may acquire a value. When outlining future research at the end of the paper, we name another means, quantification theory.

## SITUATING MEASURES AND MEASUREMENTS IN FOUR CATEGORIES

We now show how the concepts defined above can be used to describe measures and measurements in IS research. We start with Table 2, which situates measurement in its larger context.

**Table 2: Situating Measurement in its Larger Context***

| | **What is being measured?** | |
|---|---|---|
| | an empirical referent is being measured, such as:<br>a) a person's behavioral intention to use a new IT, or<br>b) the richness of a piece of communication, or<br>c) resistance to a new information system in an organization | a relationship between empirical referents (where the relationship itself is another empirical referent being measured), such as the relationship between:<br>a) a person's behavioral intention to use a new IT and the person's perception of its usefulness, or<br>b) the leanness of communication medium and the richness of a piece of communication, or<br>c) resistance to a new information system and the political balance of power in an organization |
| **What is the measure?** | 1. an **individual variable**<br>Examples:<br>1a) The construct referring to a person's behavioral intention to use a new IT is operationalized as | 2. a **statement variable**<br>Examples:<br>2a) The relationship that a person's behavioral intention to use a new IT increases when his or her attitude towards the IT increases and when his or her perceived |

| | | |
|---|---|---|
| | the measure (individual variable), **BI**.<br>1b) The construct referring to the richness of a piece of communication is operationalized as the measure (individual variable) **communication richness level.**<br><br>1c) The construct referring to an organization member's resistance to a new information system is operationalized as the measure (individual variable) **acts of resistance**. | usefulness of the IT increases is operationalized in the form of a general equation as the measure (statement variable): $\mathbf{BI=ß_0+ß_1A+ß_2U}$.<br><br>2b) The relationship between communication using lean communication media and communication richness is operationalized in the textual statement (statement variable): **communication using a non-face-to-face communication medium has a low communication richness level**.<br><br>2c) The relationship between resistance to a new information system and the organization's political balance of power is operationalized in the textual statement (statement variable): "a new information system, which specifies a distribution of intraorganizational power embodying a loss to certain organizational members, will lead to **acts of resistance** to the new system" |
| **What is the measurement?** | 3. an **individual constant**<br>Examples:<br>3a) A measurement taken of the individual variable "BI" is an individual constant, such as **the number 3.2** (resulting from the score given by a person on a questionnaire).<br><br>3b) A measurement taken of the individual variable "communication richness level" associated with an instance of communication is **low communication richness level** (resulting from comments made by a person participating in a field study).<br><br>3c) A measurement taken of the individual variable "acts of resistance" is **the non use of the new information system and the maintaining of valuable data in manual ledger books in Golden Triangle Corporation**.[1] | 4. a **statement constant**<br>Examples:<br>4a) A measurement taken of the statement variable "$BI=ß_0+ß_1A+ß_2U$" is a statement constant, such as **the fitted equation "BI=2.3+1.4A+2.2U"** (resulting from applying the general equation to a given population).<br><br>4b) A measurement taken of the statement variable "communication using a non-face-to-face communication medium has a low communication richness level" is a statement constant, such as "**communication using this email system has a low communication richness level**" (resulting from conversations with research participants in a field study of communication in their organization).<br><br>4c) A measurement taken of the statement variable "a new information system, which specifies a distribution of intraorganizational power embodying a loss to certain organizational members, will lead to **acts of resistance** to the new system" is a statement constant, such as "the new information system at Golden Triangle Corporation, which specifies a distribution of intraorganizational power embodying a loss to certain of its members, led to **the non-use of the new information system and the maintaining of valuable data in manual ledger books."** |

* Key: BI: Behavioral intention, A: Attitude, U: Usefulness, ß: population parameter.

[1] cf. Markus (1983).

Using the Technology Acceptance Model (TAM) (Davis et al., 1989), we identify U (a person's perceived usefulness of an IT), A (the same person's attitude toward the IT), and BI (the same person's behavioral intention to use this IT) as **individual variables**. Note that a researcher must make a choice about how to operationalize a theoretical construct in the form of an individual variable. A researcher may use one question, or set of questions, while another researcher may use a different question, or set of questions (Boudreau et al., 2001 p. 12). A value that BI takes for a given person, such as 3.2, is an **individual constant**. This might reflect, for example, the score that a person gives to one or more questions about BI on a survey. We regard 3.2 as a **measurement** of BI, which we regard as a **measure**. The relation "$BI=\beta_0+\beta_1A+\beta_2U$" is a **statement variable**. When applied to a population through a random sample, it can be instantiated as the **statement constant**, "BI=2.3+1.4A+2.2U." In this way, we regard the general equation "$BI=\beta_0+\beta_1A+\beta_2U$" as a **measure** and we regard the fitted equation "BI=2.3+1.4A+2.2U" as a **measurement** of it.

We stress three points about Table 2. The first is the importance of our extending the traditional approach of measuring individuals to statements. Quantitative researchers typically use "measure" and "measurement" when referring to *constructs* and use other terms such as "estimate" to refer to assessments of *relationships*, but because relationships have empirical referents, they deserve and require measuring and measurement no less than any other empirical referent. Our framework provides a way to describe the measurement of both constructs and relationships, providing a more complete basis for empirical work.

The second point is that the concepts in Table 2 are flexible enough to account for various approaches to measurement (or "measurement models," Straub et al. 2004). Quantitative and qualitative researchers often propose constructs with relationships among dimensions, codes, or items (Edwards, 2001; Glaser and Strauss, 1967). For instance, a survey researcher might measure trust with three items, the first referring to beliefs about competence, the second to beliefs about benevolence, and the third to beliefs about integrity. Another researcher might consider these beliefs to be sub-concepts (dimensions), and measure each one by a single item, or might create multiple items for each one (Gefen et al., 2003; MacKenzie et al., 2005 pp. 713-716). Many terms can be used to describe these different options, such as formative, reflective, latent, superordinate, aggregate, or profile (Edwards, 2001; Goertz, 2006; Law et al., 1998). Regardless, all of these options can be described using the concepts we provided. Each element can be viewed as an individual variable, relationships among elements can be stated using statement variables, and each variable can be instantiated with a constant. The variables that a researcher focuses on in a study will simply reflect that researcher's chosen unit of analysis. Our concepts can apply regardless.

The third point is straightforward: individual constants, individual variables, statement variables, and statements constants are not limited to quantitative research. For example, the table shows how they can apply to phenomena examined in qualitative research such as "communication richness" (Lee, 1994) and "balance of power" and "resistance" (Markus, 1983).

## PRIMARY MEASUREMENT AND MEASURE ISSUES

The four logical concepts we have discussed so far allow us to propose a way to assess "rigor" in measuring and measurement: *by researchers' justification for their individual constants, individual variables, statement variables, and statements constants.* To facilitate such justifications, we now discuss the primary issues that researchers must address for each concept. Table 3 provides a summary.

**Table 3: Primary Measurement and Measure Issues**

| | What is being measured? | |
|---|---|---|
| | an empirical referent | a relationship among empirical referents |
| What is the measure? | I. an **individual variable**<br>Primary "measure" issue:<br>• Does the individual variable reflect the theoretical concept? | II. a **statement variable**<br>Primary "measure" issue<br>• Does the statement variable reflect the theoretical relationship? |
| What is the measurement? | III. an **individual constant**<br>Primary "measurement" issue:<br>• Does the individual constant reflect the empirical referent being theorized as it is actually instantiated in the real world? | IV. a **statement constant**<br>Primary "measurement" issue:<br>• Does the statement constant reflect the empirical relationship being theorized as it is actually instantiated in the real world? |

### Primary issue with individual variables

The primary issue with an individual variable is how well it reflects (i.e., how well as a "measure" it shares the meaning of) the theoretical concept of interest.[2] This can be assessed by judging the extent to which the content expressed in the definition of the theoretical concept—its conceptualization—is shared by the content expressed in the concept's measures—its operationalization. Cenfetelli (2004) gives an example, noting that measures associated with *low* use of a site may not be valid measure of *inhibitors*, as these concepts mean different things, e.g., few cues for trusting a site need not equate to a reason to distrust it.

To clarify this primary issue, note that in confirmatory work, researchers start with a theoretical concept and then move to its operationalization whereas in exploratory work, researchers start with operationalizations (or just data) and then move to theoretical concepts. Either way, it is desirable for one's operationalizations (e.g., questionnaire items or low-level codes) and one's theoretical concepts to map to each other. Three failures can occur: (1) overrepresentation: the operationalization reflects the theoretical concept plus other unintended phenomena, (2) misrepresentation: the operationalization reflects something other than the theoretical concept; and (3) underrepresentation: the operationalization reflects part of the theoretical concept but not all of it. Each problem indicates a lack of "content validity" (Messick, 1989).

A precondition to achieving *shared* meaning between an operationalization and theoretical concepts is to ensure that the meaning of each one is clear. The meaning of a concept can be specified by outlining the properties associated with the class of things in the world that the concept refers to (Goertz, 2006; Sartori, 1970). For instance, the concept perceived ease of use is defined as *the extent to which a person believes that using a technology will be free of effort* (Davis, 1989). This definition identifies one property: "effort-free use." However, it is typically measured using items that cover properties ranging from ease of use, to flexibility, to clarity of interaction (Davis 1989), which has led some to argue that it exhibits operational overrepresentation (Evermann and Tate, 2011; Segars and Grover, 1993).

Researchers have long sought shared meaning between operationalizations and theoretical concepts:

- In *quantitative research*, researchers have created procedures (such as card sorting) to create measures with high content validity (Moore and Benbasat, 1991) and have critiqued the content

[2] "Shared meaning" can also be defined more broadly to refer to meanings shared among people (e.g., in interpretive work, between researchers and participants), but we do not consider such broader definitions in this paper.

validity of existing measures and proposed revised ones in their place (Segars and Grover, 1993). In quantitative research using archival proxies, researchers have stressed the need to check whether archival proxies match theoretical concepts by checking with those who create or use the proxies or who understand their interpretation in practice (Ketchen et al., 2012 p. 38; Wennberg, 2005 p. 15).

- Likewise, in *qualitative research*, when researchers code interview transcripts, they typically create multiple operational codes to reflect types of comments in the transcripts and then relate these codes to more abstract concepts in their coding hierarchy (Flick, 2009). Overrepresentation occurs when the operational codes reflect more than just the theoretical concept they are linked to. Misrepresentation occurs when the operational codes do not reflect the theoretical concepts they are linked to. Underrepresentation occurs when the operational codes reflect part (but not all) of the meaning of the theoretical concept. To ensure such high-quality coding, qualitative researchers often report the steps they took to ensure their coding was complete (i.e., that all relevant segments of data were examined) and that codes were applied accurately and consistently (Spiggle, 1994).

Striving for shared meaning between concepts and measures does not mean striving for an isomorphism (perfect correspondence)—the goal of so-called operationalists, long since dropped (Grace, 2001). Two phrases—'surplus meaning' and 'hermeneutic leap'—clarify the desired goal. Quantitative researchers use the phrase 'surplus meaning' to acknowledge that their *constructs* contain surplus meaning: any set of indicators will fail to reflect a construct completely (Jarvis et al., 2003 p. 202; MacCorquodale and Meehl, 1948 p. 104). Qualitative researchers use 'surplus meaning' to acknowledge that the *empirical world* always contains more meaning, because more interpretations are always possible (Van Leeuwen, 1981 p. 1). Because of surplus meaning, it is unrealistic to seek an isomorphism between measures and theoretical concepts. Yet, researchers who share the ontological and epistemological assumptions at the start of this paper still seek theoretical conceptions that reflect reality. The interpretive phrase 'hermeneutic leap' therefore helps. It suggests that researchers must be willing to leap beyond what is given: "they must make the hermeneutic leap from what the text says to what they think it means" (Stock, 1983 p. 522). This applies equally to quantitative researchers who must be willing to create items to reflect a construct, even though they know the items will fail to reflect it exactly, or use archival measures to measure revealed preferences even if they know it is a rough proxy. The leap should just not be too far. In short, researchers who subscribe to the assumptions we stated earlier must seek and try to justify “a close fit” (Easterby-Smith et al., 2008 pp. 424-425) between their individual variables (measures) and theoretical concepts.

**Primary issue with individual constants**

The primary issue with an individual constant is how accurately it reflects the particular instantiation of the empirical referent being theorized. That is, is the measurement of the instantiation accurate? Zetterberg (1965 p. 122) gave a droll example, noting that 98% of respondents in his survey said their table manners were as good at home as when at other peoples’ homes, presumably indicating substantial social desirability bias. To this day, inaccuracies in data collection remain common (Meade and Craig, 2012).

When considering this issue, it is useful to consider how inaccuracies can arise in primary and secondary data. Primary data is data that would not have been produced except for a research study. In such data, inaccuracies often occur because research methods cause them—the problem of “method bias” (Burton-Jones, 2009). For example, a researcher may interview employees to ask if they will adopt a new IT. Users may say ‘yes,’ even if they plan to resist it, because it is a more socially acceptable answer. Secondary data is data that is produced independently of a research study but that a researcher subsequently appropriates for his/her purposes. In such data, method bias can still occur because the researcher may have little knowledge of the data’s production and the biases that it contains. After all, organizational records of an event might not reflect what transpired, but what stakeholders wished others to believe (Van Maanen and Pentland, 1994). In both primary and secondary data, even if the original records are accurate, method biases can still occur if researchers make errors in interpreting or transforming the data.

Researchers have long stated the importance of being sensitive to possible inaccuracies in primary and secondary data (Webb et al., 2000). Because qualitative research involves more intensive, rather than extensive, observations than does quantitative research, qualitative researchers are often more able than quantitative researchers to get close to the context in which data is produced and, thereby, increase the likelihood of obtaining accurate data (Becker, 1996). In contrast, quantitative researchers using secondary data, especially archival proxies, are inherently limited in their ability to check for accuracy.

A general strategy for researchers to promote accuracy is to (a) understand the process by which empirical phenomena are produced and recorded and (b) identify and account for instances in this process where inaccuracies occur (Burton-Jones, 2009).

- In *qualitative* work, this involves taking steps to understand informants' language and the context in which documents in the field are produced, and behaviors performed (Becker, 1996; Heider, 1988).
- In *quantitative* work, this involves taking steps to understand how respondents interpret researchers' language and how secondary data such as records of behavior are produced and, where applicable, archived (Karabenick et al., 2007; Webb et al., 2000). For research using archival proxies, in particular, this can require the researcher to investigate the archive itself, e.g., whether it is audited and whether its data compares well with other sources (Houston, 2004; Payne et al., 2003)
- Finally, in *quantitative and qualitative work*, this involves researchers understanding (e.g., by self-appraisal or by checking with other researchers/respondents) how their biases can improperly, or their pre-understandings can constructively, influence their interpretations (Reichardt and Rallis, 1994).

## Primary issue with statement variables

The primary issue with a statement variable is how well it reflects (i.e., how well it shares the meaning of) the theorized relationship. For example, consider the IS Success Model (DeLone and McLean, 1992). It theorizes that "*There is a significant, positive relationship between Information Quality and User Satisfaction*" (Petter and McLean, 2009 p. 161). Sun et al. (2014 p. 191) explain how this relationship might be operationalized linearly or non-linearly—the relationship is not specified clearly enough to know.

Our introduction of the concept of "statement variable" exposes the severity of what we call *the problem of the linear measure*. Consider the relationship: "a person's behavioral intention to use a technology is related to the person's perception of the usefulness of the technology." A quantitative researcher might operationalize this relationship as: $BI=ß_0+ ß_1U$, but this is just one possibility. In the past, many researchers have simply assumed linearity. Once we realize that this is not simply a statistical matter but a matter of *constructing a new measure*, the need to question this assumption becomes clear.

Even if we assume that a researcher uses a regression model to operationalize a relationship, he or she can still consider the form (linear, curvilinear, or power), direction (positive or negative), coefficients (constant and slope), and limits (e.g., continuous or limited to a range of values) (Shoemaker et al., 2004 p. 59). In TAM, for instance, the relationship between its variable was originally specified as follows:

> BI is viewed as being jointly determined by the person's attitude toward using the system (A) and perceived usefulness (U), with relative weights estimated by regression: BI= A + U [Equation 4]. The A-BI relationship represented in TAM implies that, all else being equal, people form intentions to perform behaviors toward which they have positive affect. … The U-BI relationship in equation (4) is based on the idea that, within organizational settings, people form intentions toward behaviors they believe will increase their job performance…. (Davis et al., 1989 pp. 985-986).

In this quote, we can see that the authors specified the directions of the relationships (i.e., positive) and the form but not the coefficients or limits. We do not imply that every empirical paper should be able to specify every element; rather, it should be a goal for a research program over time.

The rigor with which a relationship is operationalized can be assessed in the same way as noted for individual variables: by assessing the extent to which the content expressed in the conceptualization of the relationship is shared by the content expressed in its operationalization. For instance, a confirmatory researcher may propose several theoretical mechanisms for a relationship but may only operationalize one (or none) of them (Mathieu et al., 2008), or he or she may specify several boundaries of a theoretical relationship but may operationalize them only loosely through a rough set of control variables (Spector and Brannick, 2011). In both cases, the theoretical relationship and the operationalization could fail to share meaning. Shoemaker et al. (2004 pp. 54-56) offer a simple way to assess shared meaning. For each relationship in a theory, they create a multi-column table, with a column defining the constructs in the relationship, a column defining the relationship theorized between the constructs, and a column describing the operationalized relationship. A researcher can then examine the correspondence between the columns.

Researchers can also take steps during theory building and testing to improve the link between the conceptualized and operationalized statements of a relationship:

- In *confirmatory* work, researchers can conceptualize a theory in such a way at the outset that its implied operational form is clear (Bagozzi, 1984).

- In *exploratory* work, one can pursue similar aims working backwards. For example, using TETRAD analysis, researchers can analyze all possible relationships among a set of variables and discover the most plausible subset (those that survive refutation) (Im and Wang, 2007).
- In *middle ground* approaches, researchers can increase the precision of theoretical and operational statements by iterating between concepts and data over time. For instance, in the grounded theory method, Urquhart et al. (2010, p. 367) stresses the need to focus on: "the precise nature of the association between constructs" and urges researchers to continue to code and analyse data until the relationships are closely tied to the data and the low-level codes associated with them.

**Primary issue with statement constants**

The primary issue with a statement constant is how accurately it reflects (i.e., how accurate a "measurement" it is of) the theorized relationship in the actual, particular form in which it is instantiated in the real world. For example, suppose that a researcher operationalizes a relationship with the equation $y=ß_0+ß_1x_1+ß_2 x_2$. A statement constant could be $y=1.1+2.3x_1+0.8x_2$, resulting from taking a sample of data from a population. Even aside from any errors introduced by the sampling process, the accuracy of this measurement could be good or bad, depending on what the true (but unknown) relationship is.

Ensuring accurate statement constants is a challenge in all empirical research:

- *Qualitative researchers*, for instance, can be challenged by mutually conflicting interpretations (Heider, 1988)
- *Quantitative researchers*, for instance, can be challenged to determine whether it is more accurate to include a poor proxy or omit it and thereby introduce bias (Trenkler and Stahlecker, 1996).

For both quantitative and qualitative researchers, we propose three approaches to verifying the accuracy of statement constants: verifying predictions, processes, or truth.

The first approach is to test predictions. We follow the logic that Lee and Hubona (2009 pp. 246-255) suggest for using prediction intervals. Consider the statement variable $BI=ß_0+ß_1A+ß_2U$ and the constant $BI=2.3+1.4A+2.2U$ that results from applying the equation to a given population. Researchers can test this prediction by, first, selecting a person who was part of the population studied, but not part of the sample from which the equation was estimated, measuring this person's values for A, U, and BI, and then comparing the predicted value of BI for this person with the observed value. If the observed value falls outside a prediction interval, then there is reason to conclude that the statement constant inaccurately reflects the true relationships in that population. Furthermore, the larger the proportion of a group of such out-of-sample persons for which this is true, the greater the confidence with which one may draw this conclusion. The same logic can be used in qualitative work. Consider the statement variable "communication using a non-face-to-face communication medium has a low communication richness level" and the constant "communication using this email system has a low communication richness level" (resulting from interviews with workers in a particular organization). Researchers can check if their interpretations change by sampling another worker in the same organization whom they did not interview before, learning how he/she communicates using the system, and determining if the statement constant applies equally for this person. If it does not (for instance, if this person reports a high communication richness level), then one may conclude the statement constant is inaccurate, and the larger the proportion of a group of such previously uninterviewed persons for which this is true, the greater the confidence with which one may draw this conclusion. In sum, the aim is to actively test predictions and treat failed predictions seriously rather than simply assume accuracy, or wave away failed predictions as mere outliers.

A second approach to verifying accuracy is to try to verify the underlying process that an empirical relationship is assumed to reflect. For instance, TAM researchers typically assume that the relationships they estimate reflect real-world processes. Kim and Malhotra (2005 p. 743) write: "TAM presumes that [BI] is formed as a result of conscious decision-making processes." Thus, one way to test the accuracy of the relationships is to verify the mediating mechanism(s) (Mathieu et al., 2008)—to check if users *do* engage in conscious decision-making processes before forming their usage intentions, or instead, follow a different process, e.g., informed by habit or emotion. This can be checked by using a method that sheds light on the process. Experimental researchers have long used process tracing to check their assumptions about the processes that individuals engage in (Todd and Benbasat, 1987). We are not familiar with any process tracing studies of TAM, but such an approach could be used. Researchers with a preference for lab settings could use functional magnetic resonance imaging (fMRI) to examine if/how individuals engage in conscious decision-making processes when forming their intentions to use an IT (Dimoka et al., 2007) and researchers with a preference for field settings could use experience sampling methods to learn the cognitive and emotional process that lead people to use IT in specific ways (Ortiz de Guinea and Webster, 2013).

A third approach that researchers can use to test the accuracy of statement constants is to verify the truth of relationships reflected in the data. Perhaps the most common example of this in IS research is the use of robustness checks to ensure that estimates of relationship are as unbiased as possible (Tambe and Hitt, 2014). However, this approach can become even more focused when combined with the second approach—when researchers have evidence regarding the underlying process. For example, assume that a lab researcher follows the second approach above by collecting verbal protocols of users' decision-making processes (Ericsson and Simon, 1993). The accuracy of these protocols depends on subjects telling their thoughts truthfully. The same applies in field studies. Eisenhardt and Graebner (2007 pp. 29-30) write: "It is … crucial to [determine] the … logical link between the constructs …. [This] can be drawn from … an informant explaining the logic…." In such cases, the accuracy of the data depends on whether informants tell the truth. These issues are simply more manifestations of the general problem of method bias and can be addressed by understanding where potential biases can arise in data collection and using multiple methods to minimize their impact on the researcher's conclusions (Eisenhardt and Graebner, 2007 p. 28).

**Summary of primary issues**

Figure 1 summarizes our arguments. It distinguishes between theory, reality, and empirical research, where the last of these bridge the first two, much like Campbell's (1974 p. 449) "trinism…of data, theory, and real world." Figure 1 does not distinguish between operationalizations and data because in *confirmatory* work, operationalizations and data are always distinguishable. For example, researchers might operationalize a construct by creating questions for a structured interview script. Responses to these questions are considered data. However, in *exploratory* work, a researcher may take field notes and use these notes as evidence. Because the notes are filtered by the researcher's perceptions and sense-making, the line between operationalization and data is blurry. The researcher may even use an *in vivo* code to operationalize a theoretical concept (Spiggle, 1994), where such a code reflects a research subject's exact words (Strauss, 1987). At this limit, the distinction between operationalization and data disappears.

In short, Figure 1 is broad enough to apply to any positivist research, whether confirmatory or exploratory, and whether qualitative, quantitative, or mixed methods.

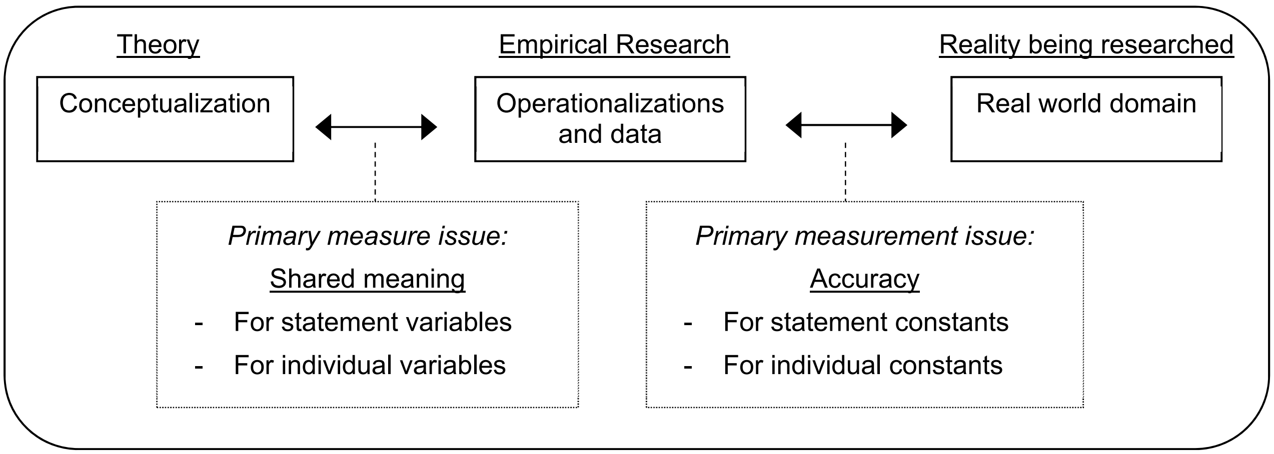


**Figure 1: Summary of the Primary Issues with Measures and Measurement**

As the figure shows, the primary "measure" issue in empirical research is the extent to which there is *shared meaning* between researchers' conceptualization and their operationalizations, and the primary "measurement" issue is the *accuracy* of data obtained about the world.[3] The criterion of shared meaning applies to individual and statement variables, while the criterion of accuracy applies to individual and statement constants. Shared meaning and accuracy are equally important. When accuracy is high, researchers can say that they have measured *something* successfully, but only when accuracy *and* shared meaning are high can researchers say that they have measured what they theorized, or theorized what they measured. Some pre-existing frameworks for research tend to emphasize accuracy over meaning, or vice versa. For instance, Bhaskar (1979 p. 59) wrote: "…precision in meaning now assumes the place of

[3] As we just observed, the distinction between operationalization and data can be blurry in exploratory work. In fact, there could be times in an exploratory study when the distinctions between theory, operationalization, data, and real world, are all somewhat blurry. Nevertheless, the principles proposed here still apply: measures and measurements will improve to the extent researchers are guided by the dual criteria of shared meaning and accuracy.

accuracy in measurement as the a posteriori arbiter of theory." Our framework emphasizes both aspects.

The issues we have identified are sufficiently general that they can apply to all empirical research, *irrespective* of the type of data (quantitative/qualitative/mixed), one's research approach (confirmatory, exploratory, or mixed), or the type of constructs that one is studying (uni- or multi-dimensional, formative or reflective). They can be considered foundational criteria, therefore, for assessing measures and measurements. The value of general criteria is that they can help researchers who work in any given tradition as well as those working across traditions (Mingers, 2001; Venkatesh et al., 2013). To date, general criteria for measurement have not been forthcoming, perhaps because of the vast differences in techniques and terminologies across communities. Our analysis helps to address this problem.

In prior sections, we noted practices that researchers could use to address the primary issues. Table 4 provides a summary, which can serve as a guide for research. We list the practices in Table 4 at a high level of abstraction because each practice can be enacted in different ways. We recognize that researchers know many of these practices, yet the next section shows that they are not necessarily well-understood or enacted. Researchers should be encouraged and rewarded for their creativity in applying them to the context of a given study.

**Table 4: Summary of Recommended Practices for Addressing Primary Issues**

| Logical concept | Practices for addressing primary issues |
|---|---|
| Shared meaning for individual variables and statement variables | • Clearly specify the meaning of each theoretical concept and its operationalization, ensuring that the properties of each one are clear (Goertz, 2006). In the case of individual variables, such properties will characterize a particular phenomenon (e.g., properties that reflect 'ease of use'). In the case of statement variables, such properties will characterize relationships among phenomena (e.g., properties that reflect how 'ease of use' affects 'intention to use an IT').<br>• Validate the degree to which the meaning of the theoretical concept is shared by the meaning of the operationalization. For individual variables, some validation procedures such as card-sorting (Moore and Benbasat 1991) and reviews by expert academic and practitioner panels (Ketchen et al. 2012) are well-known, but more sophisticated approaches exist (e.g., Schriesheim et al., 1999). For statement variables, analogous practices can be used, such as the multi-column table procedure in Shoemaker et al. (2004, pp. 54-56). Validation should be conducted in the knowledge that surplus meaning cannot be eliminated entirely. |
| Accuracy for individual constants and | • Understand the process by which empirical phenomena are produced and recorded (in primary and secondary data) and identify and account for instances in this process where inaccuracies may occur (e.g.., by minimizing method bias) (Burton-Jones, 2009; Webb et al., 2000). This |

| | |
|---|---|
| statement constants | includes learning how secondary data archives are produced and managed (where applicable) and understanding how one's own biases and pre-understandings affect one's measurements.<br>• Verify the accuracy of measurements of relationship by verifying predictions (e.g., using prediction intervals), verifying processes (e.g., identifying mediators and using process tracing techniques), and verifying truth (e.g., by testing for robustness and minimizing method bias). |

## RAMIFICATIONS FOR QUANTITATIVE AND QUALITATIVE RESEARCH

Several ramifications flow from our analysis that have implications for empirical research in IS.

### Understanding how errors cascade in empirical work

The first ramification is the need to understand how problems in accuracy and shared meaning can cascade through research. First, they can accumulate as researchers move horizontally through Figure 1.

- In *confirmatory* research, the problems accumulate from the left to the right, i.e., if researchers fail to ensure shared meaning between their theoretical constructs and operationalizations, this restricts their ability to measure the constructs they theorized.
- In *exploratory* research, problems accumulate from the right to the left, i.e., no matter how sophisticatedly researchers explore their data, they cannot have faith that their conclusions reflect reality unless steps are first taken to ensure the accuracy of their data.

Second, the problems can also accumulate as research moves from the study of individual constants and variables to statement constants and variables, because statements involve relationships among concepts. Straub (1989) stressed this point for confirmatory quantitative research. We emphasize its generality for all research.

Overall, the cascading effect of problems with measures and measurements suggests the need to move back and forth between the three realms of Figure 1 over time—checking one's conceptualizations, one's operationalizations, and one's understanding of the empirical domain. Likewise, it suggests the need to move back and forth between one's understanding of constructs and the relationships among them.

### The need for additional measuring

A second ramification is that researchers need to engage in *additional* work to improve their measures and measurements, beyond current practice. This is because many issues we have discussed are rarely addressed in IS research. This problem applies to all four logical concepts we have introduced:

- ***Individual variables***. We noted earlier that having shared meaning between a theorized construct and operationalized measure is very similar to the notion of "content validity." However, Boudreau et al. (2001) found that only 23% of articles in their sample verified the content validity of measures.
- ***Individual constants***. We discussed earlier the need to minimize method bias to obtain accurate data about individual constants. However, in a review of articles in top IS journals during 1995-2005, King et al. (2007) found that only 37% of studies mentioned method bias, let alone addressed it. In our Supplementary Appendix, we provide an updated review (2006-2015), which shows that practices have not improved greatly.
- ***Statement variables***. We noted earlier how researchers have many choices when operationalizing relationships, such as whether the relationships should be linear or nonlinear, but they often fail to explicate or justify these choices. We also observed that most researchers think of content validity as something that applies only to constructs, not relationships too. We also noted the need for grounded theory researchers to clearly specify relationships and tie them closely to their data, but according to Urquhart et al. (2010), this is what many grounded theory researchers perform least effectively.
- ***Statement constants***: IS researchers rarely carry out the three procedures we noted earlier for ensuring the accuracy of statement constants, with the exception of the increasing use of robustness checks in econometric style research (Tambe and Hitt, 2014). IS researchers almost never test prediction intervals (Lee and Hubona, 2009), and they rarely use alternative methods to verify processes or reduce method bias. Confirmatory, quantitative studies often rely on just one method (either purely archival data or purely cross-sectional, questionnaire data) (Burton-Jones, 2009).

**The need to place more focus on fundamental issues**

A third major ramification is the need to refocus on the *fundamental* issues of measures and measurement. IS researchers carry out many tasks to justify measures and measurements, but not all provide strong evidence for the quality of measures or the accuracy of measurements. We illustrate this issue by analyzing examples of practices recommended in quantitative and qualitative work. The common lesson is that current guidelines fail to fully address the fundamental issues of measures and measurement.

**An Example from Quantitative Research**

Confirmatory, quantitative researchers have many guidelines for measures and measurement. In this section, we draw on Straub et al. (2004), who provide a particularly exhaustive account of the tasks that confirmatory quantitative researchers can carry out to validate measures and measurements. In Table 5, we examine each of these elements of validity, focusing on the extent to which they are relevant to issues regarding the quality of the measures in a study or the quality of measurements.

We briefly review the four categories of elements in Table 5. The first element, content validity, refers to the extent to which a researcher's conceptualization of a construct is reflected in his/her operationalization of it. Because this is a matter of shared meaning, it is of vital importance for the quality of measures. It is not a measurement issue because it bears no relation to the accuracy of the data. Even from the perspective of measures, however, Straub et al. (2004) observed that many researchers had begun to consider content validity as *not* being a critical issue to address. Like several other methodologists, we would suggest that it *is* critical to address (Lissitz and Samuelson, 2007; Rossiter, 2002). Like them, we suggest that a shift in emphasis is required because content validity is a vital aspect of high quality measures.

**Table 5: Analyzing Common Validation Tasks in Confirmatory, Quantitative Research**

| **Validity elements (from Straub et al. 2004, pp. 385-386)*** | **Relevance for the measure issues that we outlined (i.e., the extent to which conceptualizations and operationalizations share meaning)** | **Relevance for the measurement issues that we outlined (i.e., the accuracy of data)** |
|---|---|---|
| 1. Content validity | High | None |
| 2. Construct validity | | |
| • Discriminant validity | Low | Moderate |
| • Convergent validity | Low | Moderate |
| • Nomological validity | Low | Moderate |
| • Common method bias | Low | High |
| 3. Reliability | | |
| • Internal consistency | None | Moderate |
| • Inter-rater reliability | None | Moderate |
| • Unidimensional | None | Moderate |

| | | |
|---|---|---|
| 4. Manipulation validity | None | High |

* Table 5 excludes some tests rarely used in IS research (factorial validity, concurrent validity, split-half reliability, test-retest reliability, and equivalent forms), but the pattern in the table would not change by including them.

The second category is construct validity. As Table 5 shows, we assessed the first three elements of construct validity (all except common method bias) in the same way because they all use a similar, correlational, logic. That is, they depend on researchers examining correlations reflected in the loadings of measures on constructs (discriminant and convergent validity) or among constructs in a model (nomological validity). The problem with all these criteria is that they involve assessments of correlations rather than assessments of the semantics of measures or the truth of data. Borsboom et al. (2004 p. 1066) remark: "The crucial mistake is the view that validity is about correlation." A good way of illustrating this problem is to bring in the last element of construct validity—common method bias. If a researcher's data is affected by common method bias, it is quite possible that the other elements of construct validity will appear satisfactory *because* of such bias (e.g., measures will load well and predictions will be confirmed) (Feldman and Lynch, 1988; Sharma et al., 2009). The supportive results might provide comfort that *something* was measured accurately (hence the label "moderate" in the right column of Table 5). But, in and of itself, the supportive results cannot provide comfort that a researcher measured the intended constructs and relationships.

The third category is reliability. Reliability refers to measurement consistency or precision. It is well known that reliability alone does not imply validity. For example, a person may consistently misunderstand questions in a questionnaire. For this reason, Table 5 lists reliability as having no relevance for the quality of measures. The relationship between reliability and the accuracy of measurements is complex. Highly consistent responses can be inaccurate measures of the construct of interest, but accurate measures of something else (e.g., what the respondent mistakenly thought you were asking about). Therefore, consistent and precise responses do not necessarily imply accurate measurements. For this reason, we listed the relevance of reliability as "moderate" in Table 5.

The fourth category is manipulation validity. The most common way in which researchers assess it is via a manipulation check (Straub et al. 2004). Because this often involves one or more questions (and, hence, measures), adding it to a study does not alter the quality of measures. Thus, it is listed as having no relevance for measures in Table 5. However, we listed it as having high relevance for verifying measurement accuracy. Specifically, it can help researchers reach more accurate conclusions regarding the reason for the relationship observed in the data, i.e., whether the relationship was due to the manipulation or not (Straub et al. 2004). In Boudreau's et al. (2001) review of the literature, however, only 22% of studies checked for manipulation validity and Mackenzie et al. (2011) lamented its infrequent use too.

Overall, our analysis above suggests that although confirmatory, quantitative researchers have many ways to assess validity, very few of them actually provide strong evidence regarding the quality of measures or measurements. Although good researchers do not assess the elements of validity in Table 5 mechanically, or in isolation (Straub et al. 2004), the fact that Table 5 contains very few activities of "high" relevance, coupled with the fact that activities that are highly relevant (such as content validity, common method bias, and manipulation validity) are addressed infrequently in IS research, leads us to the disarming conclusion that guidelines for confirmatory quantitative researchers appear to *focus more on ancillary issues, rather than the central issues of measures and measurements.* A shift in emphasis is required.

To clarify the significance of the required shift in emphasis, assume that we are reviewing a confirmatory, quantitative paper. A typical behavioral paper of this genre will justify convergent validity, discriminant validity, nomological validity, and reliability. Such tests do not, however, provide rigorous evidence of the quality of measures or measurements. We would propose that if the paper showed greater attention to content validity, method bias, and manipulation validity, more faith could be placed in its measures and measurements and, all things being equal, it should have a greater chance of acceptance.

**An Example from Qualitative Research**

The Grounded Theory Method (GTM) is an interesting method to consider because its signature claim is that it allows theory to "emerge" from data. Thus, one might assume that accuracy of data would be a key concern, because inaccurate data would lead to the emergence of an invalid theory; yet, GTM's founders downplayed its role:

- Glaser and Strauss (1967): "[Theory generation] …can be quickly killed by the twin critiques of accurate evidence and verified hypotheses (p. 28)…. . accuracy is not as important for generating theory …as it is for describing a … social unit or verifying a hypothesis" (p. 189).

- Glaser (1978 p. 54): "…the accuracy of the facts is not so 'crucial' (as we said over and over again in DISCOVERY) …."
- Glaser and Holton (2004): "…the conceptual hypotheses of GT [Grounded Theory] do not entail the problems of accuracy that plague [other qualitative data analysis methods]"
- Stern (2007): "Both Glaser and Strauss advised their students against paying attention to the accuracy of the interview data."
- Bryant and Charmaz (2007): "Glaser's stance implies that the researcher does not need to be concerned with quality of the data, range of data, amount of data, access to data, or accuracy…."

In Table 6, we show five key GTM principles and their relevance for issues of measures and measurement.[4] While these principles are commonly agreed to in GTM work, there are different GTM approaches, sometimes called the Glasserian and Straussian versions. Neither version emphasizes accuracy, but explicit admonitions *against* accuracy mainly appear in the Glasserian version. For instance, while Corbin and Strauss (2015) do not discuss accuracy in depth, they do occasionally discuss ways to improve it, namely, using multiple methods (pp. 37, 51), being sceptical (pp. 99, 146), checking data credibility (pp. 41, 199), and controlling biases (pp. 47, 66). However, they do not discuss differences between the Glasserian and Straussian perspectives on accuracy, and we have been unable to find a detailed explanation of the issue elsewhere in the GTM literature either. In this light, we focus on the five principles in Table 6 that have widespread emphasis in the GTM literature. We discuss each one in turn.

**Table 6: Analyzing Principles in Grounded Theory Research**

| Grounded theory principles | Relevance for measure issues (i.e., shared meaning between conceptualizations and operationalizations) | Relevance for measurement issues (i.e., accuracy of data) |
|---|---|---|
| 1. Theoretical sensitivity and management of preconceptions | High | High |
| 2. Constant comparison and different slices of data | High | Moderate |
| 3. Theoretical sampling | High | Not applicable |
| 4. Iterative, emergent, theoretical coding | High | Not applicable |

[4] Some researchers might question our examination of GTM in this section because GTM was originally developed as a non-positivist approach (Glaser and Strauss 1967, Strauss and Corbin 1994, p. 279). Nonetheless, GTM has evolved since that time and it is widely viewed as being epistemologically neutral and, therefore, appropriate for use by positivist, interpretive, critical, and mixed-paradigm research (see Urquhart et al. 2010, p. 361).

| | | |
|---|---|---|
| 5. Theory building, saturation, & integration | High | Low to moderate |

Source: (Birks et al., 2013 p. 2-3; Matavire and Brown, 2013 p. 120; O'Reilly et al., 2012 p. 248; Urquhart et al., 2010 p. 369)

*Theoretical sensitivity* was described by Glaser and Strauss (1967, p. 46) as follows:

> The sociologist should also be sufficiently *theoretically sensitive* so that he can conceptualize and formulate a theory as it emerges from the data. Once started, theoretical sensitivity is forever in continual development. It is developed as over many years the sociologist …[improves his] ability to have theoretical insight into his area of research, combining with an ability to make something of his insights [emphasis in original].

Theoretical sensitivity is indispensable in the act of coding—the assigning of a group of words to a researcher-defined category. This is what *measuring* is. The ability to perform such coding depends on the researcher's familiarity with existing theory and his/her experience in collecting and analyzing data. Thus, as Table 6 shows, theoretical sensitivity is highly relevant for *measure* issues.

Theoretical sensitivity also implies being sensitive to one's theoretical or personal preconceptions (Birks et al. 2013). For example, Van Maanen (1979) recalls a study in which he found that police had disdain for those who asked for help with family disputes. He found this hard to comprehend through his preferred lens of occupational theories. Only when he began viewing it differently, in terms of character, did it make sense. He wrote that he had been "blinded by [his] own perceptual screen" (p. 548). Thus, as Table 6 shows, theoretical sensitivity has high relevance for *measurement* issues too.

*Constant comparison* is the process of constantly comparing "instances of data labeled as a particular category with other instances of data in the same category" (Urquhart et al., 2010 p. 369) and comparing "concept to concept" to ensure that distinctly different instances of data are labeled under different categories (Matavire and Brown, 2013 p. 120). The product of constant comparison is a clear and well-justified link between one's data and one's theoretical concepts. Thus, it has high relevance for the *measure* issues we have outlined because it results in very well-justified codes.

The link between constant comparison and the accuracy of data is less clear. On one hand, Glaser and Strauss downplayed the role of accuracy and they took this line when discussing constant comparison (Glaser and Strauss 1967, p. 28):

> …generation of theory through comparative analysis … assumes … accurate descriptions, but only to the extent that the latter are in the service of generation. Otherwise they are sure to stifle it.

On the other hand, they argued that constant comparison could improve accuracy, especially when comparing *different slices of data*, where a slice simply refers to a specific type of data that offers a distinct view on the phenomena being studied (Glaser and Strauss 1967, p. 223). They argued (p. 68):

> … when different slices of data are submitted to comparative analysis… biases of particular people and methods tend to reconcile themselves as the analyst discovers the underlying causes of variation.

Overall, because Glaser and Strauss both highlighted and downplayed the influence of these two principles on accuracy, we indicate in Table 6 that the relevance of these principles for accuracy is 'moderate.'

*Theoretical sampling* refers to "the need to understand the nature and dimensions of emerging conceptualizations …by sampling data in a way that varies a particular set of dimensions…." (Birks et al., 2013 p. 3). Theoretical sampling supports constant comparison because if a researcher observes data that does not fit an existing category, he/she is motivated to collect more of such data, create codes to reflect it, and engage in further rounds of constant comparison to refine the codes. Because theoretical sampling ensures that codes are well-grounded in data, it has high relevance for measure issues, but not for measurement accuracy. Glaser and Strauss (1967, p. 30) wrote: "Since accurate evidence is not so crucial for generating theory, the kind of evidence, as well as the number of cases, is also not so crucial."

*Iterative, emergent, and theoretical coding* refers to the ongoing process through which researchers "identify the properties, dimensions, and boundaries" of categories (O'Reilly et al., 2012 p. 251) and then increase the abstraction level by relating "categories to each other" and scaling up "higher-level categories into broader themes" (Urquhart et al., 2010 p. 369). This process can be thought of as an extension of constant comparison, focusing on the development of the emerging theory. Theoretical sampling also supports this process because as researchers develop higher-level categories and relationships, they should constantly check whether additional data are required to support the new claims (Birks et al., 2013 p. 3).

Because this principle requires researchers to ensure that they have well-refined and justified codes and relationships among them, it is highly relevant for measure issues, for both individual variables *and* statement variables. However, the principle is silent on measurement accuracy—that is, it applies to issues of measures but not issues of measurement (per Table 4). Even so, while it is highly relevant for measure issues, it is often enacted poorly in practice; researchers often justify their categories (the *individual variables*), but not the relationships among them (the *statement variables*) (Urquhart et al., 2010 p. 370).

*Theory building, saturation, and integration*, the final GTM principle, involves researchers' articulation of their final grounded theory. The theory is considered saturated if "subsequent data incidents that are examined provide no new information, either in terms of refining the category or of its properties, or of its relationships to other categories" (Locke, 2001 p. 254). Saturation can only be met if one has followed the principles of constant comparison, theoretical sampling, and iterative theoretical coding. Thus, just as with those principles, this principle has high relevance for *measure* issues. However, because it focuses on refining and substantiating the final theory, rather than the accuracy of the data that the theory is based on, Table 6 shows that it has little relevance for issues of measurement. The only exception is the principle of *integration*. Glaser and Strauss (1967) argued that it "tends to correct inaccuracies of … data" (p. 223). Their rationale was that "If the index …is consistently related to a whole series of variables that, when put together, yield an integrated theory—this is … a more trustworthy validation … than the standard method of … showing that [the index measures]… what it is supposed to measure" (pp. 191-192). This is reminiscent of the view once held in the quantitative literature that construct validity is not so much a question of whether "one measures what one intends to measure … [but] whether the empirical relations … match [expected] theoretical relations" (Borsboom et al., 2004 p. 1061). In the quantitative literature, this is not considered strong evidence anymore because a researcher can obtain an integrated set of findings that is false (Borsboom et al., 2009; Borsboom et al., 2004). Thus, while we agree on the value of integration for issues of measures, its relevance for issues of measurement in Table 6 is low to moderate.[5]

Overall, our analysis shows that the GTM—especially the Glasserian version—stresses the quality of one's measures but not the accuracy of measurements. It offers three techniques for improving

[5] As one of our reviewers noted, we should re-emphasize that these arguments presume a commitment to the existence of an objective reality. The arguments would differ if GTM is used under strong interpretive assumptions.

accuracy: managing preconceptions, continually comparing different slices of data, and checking on pieces of data that do not fit the emerging, integrated account. Although these techniques can help improve measurement accuracy, the GTM founders argued against worrying about accuracy too much. This distinguishes GTM from many other qualitative methods, such as ethnography (Becker, 1996 p. 58; Van Maanen, 1979 p. 544) and the case study method (Lofland et al., 2006 p. 90; Yin, 1994 p. 81).

Glaser and Strauss (1967) gave two reasons for not worrying about accuracy: (1) focusing on accuracy can stifle "the creative energies required for discovering theory" (pp. 7; 28), and (2) "it is easier, faster, and considerably more economical to use [a] crude index" (p. 191). Glaser and Strauss did not justify either claim, but neither one is clear. Much like the quote that 'truth is stranger than fiction,' getting more accurate data could *spark* creativity by generating surprises. Likewise, it could be much more economical to develop an accurate theory because misguided theories can be costly. We do not make these arguments to critique GTM but to *strengthen* it. We believe there is an opportunity for a shift in emphasis in the GTM literature—to improve it by accounting for issues of *both* measures and measurement.

**Summary of Ramifications and the Limitations of our Proposal**

In summary, our work stresses researchers' need to: (1) be mindful of how errors with measures and measurements cascade through the research process; (2) engage in additional work to improve their measurement efforts, beyond current practices; and (3) focus on the fundamental issues of measures and measurement and recognize the extent to which current best-practice guidelines fail to address them.

Rather than seek to advance new techniques, the main theme of this essay has been the need to rethink and reemphasize fundamentals. Our definitions, framework, and analysis are new, and go beyond past work, but some of the practices we discussed for improving measures and measurement are not new; they are just rarely practiced. The value in our work lies not in the development of a new technique, but in revealing the full landscape of measurement issues that researchers must address, which is much broader than recognized in past work, and in offering a parsimonious account of the primary issues to be addressed in this landscape, irrespective of one's quantitative, qualitative, or mixed-methods orientation.

The prior sections showed that current practices—both enacted and recommended—inadequately address the fundamental issues of measures and measurement. There are only a small set of fundamental issues, but the practices that address them are too often overlooked.[6] A shift in emphasis is required. We believe the time is ripe for such a shift, given the problems we noted at the outset—the rapidly expanding number of principles that quantitative researchers must keep up with and the heterogeneous principles that mixed-methods researchers must handle. The point of our paper is not to add more principles and practices to the litany of existing ones, compounding the problem noted at the outset. Rather, our aim is to enable researchers to concentrate on and address the fundamental issues involved, irrespective of one's qualitative or quantitative orientation, and to enable reviewers to recognize and reward those who do.

The limitations of our work should also be noted. We offered a high-level guide, but not detailed instructions. For instance, we stressed the goal of shared meaning, but not how to implement it in, say, interval scales. Likewise, we did not discuss how to apply these ideas in particular research designs, such as in longitudinal or process-type studies in which shared meaning and accuracy may vary over time. We also provided detailed treatments of only some types of research. For instance, it would be useful to apply the insights offered here to other types of qualitative research beyond GTM (e.g., positivist case studies) and also examine if the ideas can be extended to interpretive research. A full treatment would also examine tradeoffs and constraints, e.g., how to allocate effort between theorizing, measuring, and topical relevance, and how to tradeoff measurement accuracy with speed or cost.

Finally, a limitation in our use of symbolic logic is that our paper reflects just a first step, as "the substitution of individual constants for individual variables is not the only way that propositions can be obtained from propositional functions…. [They] may also be obtained by the process called generalization or quantification" (Copi, 1986 p. 343). Having shown the value of using variables and constants in this paper, an appropriate next step in the use of symbolic logic would be to apply the lessons of quantification theory (which, despite its name, deals with qualitative logic) to measures and measurement too.

## CONCLUSIONS

[6] We are not saying that they are always overlooked. Some exemplar studies do address these issues very well, while other studies do not (see our Supplementary Appendix for examples).

Empirical research requires measures and measurements, but researchers do not have good definitions of measures and measurement, nor do they have clear agreement about how best to engage in or assess these activities (Lissitz, 2009). Our essay has contributed by defining these activities, clarifying their scope, identifying the primary issues that researchers should consider when engaging in them, and assessing how rigorously they are addressed in current practice, and in best-practice guidelines.

We addressed our topic in a general manner that can apply to quantitative, qualitative, and mixed-methods research alike. By taking this approach, we hope that our essay can provide a platform for shared discourse among researchers in our field. We hope this will help researchers to improve their measures and measurements, recognize the slow, ambitious, and programmatic work required to do so, and foster mutual appreciation of the common challenges that researchers face when they engage in empirical work.

**Acknowledgments:** We thank the SE, AE, and reviewers at *Information Systems Research* for their many helpful comments on the paper. Earlier versions benefited from presentations at the University of British Columbia, Queensland University of Technology, and the *Hawaii International Conference on Systems Sciences*. We thank Henri Barki, Sabine Matook, Tom Stafford, Natalia Levina, Thi Van Hau Trieu, and Heshan Sun for helpful advice at different stages. The research has also been made possible by the support of the UQ Business School and the Australian Research Council (FT130100942).

## REFERENCES

Bagozzi, RP (1984) A Prospectus for Theory Construction in Marketing. *J. of Marketing,* 48(Winter): 11-29.

Bagozzi, RP (2011) Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations. *MIS Quart.,* 35(2): 261-292.

Becker, HS (1996) The Epistemology of Qualitative Research. In R Jessor, A Colby, RA Shweder (Eds.), *Ethnography and Human Development* (pp. 53-72). Chicago: University of Chicago Press.

Bhaskar, R (1979) *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. New Jersey: Humanities Press.

Birks, DF, Fernandez, W, Levina, N, Nasirin, S (2013) Grounded Theory Method in Information Systems Research: Its Nature, Diversity, and Opportunities. *European J. of Infor. Syst.,* 22: 1-8.

Borsboom, D (2009) Book Review: Educational Measurement (4th Edition). *Structural Equation Modeling,* 16: 702-711.

Borsboom, D, Cramer, AOJ, Kievit, RA, Scholten, AZ, Franic, S (2009) The End of Construct Validity. In RW Lissitz (Ed.), *The Concept of Validity: Revisions, New Directions, and Applications* (pp. 135-170). Charlotte, NC: Information Age Publishing.

Borsboom, D, Mellenberg, GJ, van Heerdon, J (2004) The Concept of Validity. *Psychological Review,* 111(4): 1061-1071.
Boudreau, M-C, Gefen, D, Straub, DW (2001) Validation in Information Systems Research: A State-of-the-Art Assessment. *MIS Quart.,* 25(1): 1-16.
Bryant, A, Charmaz, K (2007) Grounded Theory in Historical Perspective: An Epistemological Account. In A Bryant K Charmaz (Eds.), *The SAGE Handbook of Grounded Theory* (pp. 31-58). London: Sage Publications.
Burton-Jones, A (2009) Minimizing Method Bias Through Programmatic Research. *MIS Quart.,* 33(3): 445-471.
Campbell, DT (1974) Evolutionary Epistemology. In PA Schilpp (Ed.), *The Philosophy of Karl Popper* (Vol. Book 1, pp. 413-463). La Salle, Illinois: Open Court.
Cenfetelli, R (2004) Inhibitors and Enablers as Dual Factor Concepts in Technology Usage. *J. of the Association for Infor. Syst.,* 5(11-12): 472-492.
Churchill, GA (1979) A Paradigm for Developing Better Measures of Marketing Constructs. *J. of Marketing Res.,* 16: 64-73.
Copi, I (1986) *Introduction to Logic* (7th ed.). New York: Macmillan.
Corbin, J, Strauss, A (2015) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. 4th Ed., Thousand Oaks, CA: Sage.
Davis, F (1989) Perceived Usefulness, Perceived Ease of Use, and End User Acceptance of Information Technology. *MIS Quart.,* 13(3): 318-339.
Davis, FD, Bagozzi, RP, Warshaw, PR (1989) User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science,* 35: 982-1003.
DeLone, WH, McLean, ER (1992) Information Systems Success: The Quest for the Dependent Variable. *Infor. Syst. Res.,* 3(1): 60-95.
Dimoka, A, Pavlou, PA, Davis, FD (2007) Neuro-IS: The Potential of Cognitive Neuroscience for Information Systems Research. *Proceedings of the 28th International Conference on Information Systems* Montreal, Canada, Dec 7-9, 1-20.
Easterby-Smith, M, Golden-Biddle, K, Locke, K (2008) Working with Pluralism: Determining Quality in Qualitative Research. *Organiz. Res. Methods,* 11: 419-429.
Edwards, J, R. (2001) Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytic Framework. *Organiz. Res. Methods,* 4(2): 144-192.
Eisenhardt, KM (1989) Building Theories from Case Study Research. *Academy of Management Review,* 14: 532-550.
Eisenhardt, KM, Graebner, ME (2007) Theory Building from Cases: Opportunities and Challenges. *Academy of Management J.,* 50(1): 25-32.
Ericsson, KA, Simon, HA (1993) *Protocol Analysis: Verbal Reports as Data (Revised edition)*. Cambridge, MA: The MIT Press.
Evermann, J, Tate, M (2011) Fitting Covariance Models for Theory Generation. *J. of the Association for Infor. Syst.,* 12(9): 632-661.
Feldman, JM, Lynch, JG (1988) Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *J. of Applied Psychology,* 73(3): 421-435.
Flick, U (2009) *An Introduction to Qualitative Research* (4th ed.). London, UK: Sage Publications.
Gefen, D, Karahanna, E, Straub, DW (2003) Trust and TAM in Online Shopping: An Integrated Model. *MIS Quart.,* 27(1): 1-40.
Glaser, BG, Holton, J (2004) Remodeling Grounded Theory. *Forum: Qualitative Social Research,* 5(2).
Glaser, BG, Strauss, AL (1967) *The Discovery of Grounded Theory*. Chicago: Aldine.
Glaser, R (1978) *Theoretical Sensitivity*. Mill Valley, CA: The Sociology Press.
Goertz, G (2006) *Social Science Concepts*. Princeton, NJ: Princeton University Press.
Grace, RC (2001) On the Failure of Operationism. *Theory & Psychology,* 11(1): 5-33.
Grover, V, Lyytinen, K (2015) New State of Play in Information Systems Research: The Push to the Edges. *MIS Quart.,* 39(2): 271-296.

Heider, KG (1988) The Rashomon Effect: When Ethnographers Disagree. *American Anthropologist,* 90(1): 73-81.
Hofweber, T (2004) Logic and Ontology. In EN Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Stanford University: available online at: http://plato.stanford.edu/entries/logic-ontology/.
Houston, MB (2004) Assessing the Validity of Secondary Data Proxies for Marketing Constructs. *J. of Business Res.,* 57: 154-161.
Im, G, Wang, J (2007) A TETRAD-based Approach for Theory Development in Information Systems Research. *Communications of the Association for Infor. Syst.,* 20(Article 22): 322-345.
Jarvis, CB, MacKenzie, SB, Podsakoff, NP (2003) A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *J. of Consumer Res.,* 30: 199-218.
Karabenick, SA, Woolley, ME, Friedel, JM, Ammon, BV, Blazevski, J, Bonney, CR, Groot, ED, Gilbert, MC, Musu, L, Kempler, TM, Kelly, KL (2007) Cognitive Processing of Self-Report Items in Educational Research: Do They Think What We Mean? *Educational Psychologist,* 42(3): 139-151.
Ketchen, DJ, Ireland, RD, Baker, LT (2012) The Use of Archival Proxies in Strategic Management Studies: Castles Made of Sand? *Organiz. Res. Methods,* 16(1): 32-42.
Kim, SS, Malhotra, NK (2005) A Longitudinal Model of Continued IS Use: An Integrative View of Four Mechanisms Underlying Post-Adoption Phenomena. *Management Science,* 51(5): 741-755.
King, WR, Liu, CZ, Haney, MH (2007) Method Effects in IS Survey Research: An Assessment and Recommendations. *Communications of the Association for Infor. Syst.,* 20: 457-482.
Klenk, V (1994) *Understanding Symbolic Logic*. NJ: Prentice Hall.
Law, KS, Wong, C-S, Mobley, WH (1998) Toward a Taxonomy of Multidimensional Constructs. *Academy of Management Review,* 23(4): 741-755.
Lee, AS (1994) Electronic Mail as a Medium for Rich Communication: An Empirical Investigation Using Hermeneutic Interpretation. *MIS Quart.,* 18(2): 143-157.
Lee, AS, Hubona, GS (2009) A Scientific Basis for Rigor in Information Systems Research. *MIS Quart.,* 33(2): 237-262.
Lissitz, RW (Ed.) (2009) *The Concept of Validity: Revisions, New Directions, and Applications*. Charlotte, NC: Information Age Publishing Inc.
Lissitz, RW, Samuelson, K (2007) A Suggested Change in Terminology and Emphasis Regarding Validity and Education. *Educational Researcher,* 36(8): 437-448.
Locke, K (2001) *Grounded Theory in Management Research*. Thousand Oaks, CA: Sage.
Lofland, J, Snow, DA, Anderson, L, Lofland, LH (2006) *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Belmont, CA: Wadsworth.
MacCorquodale, K, Meehl, PE (1948) On a Distinction Between Hypothetical Constructs and Intervening Variables. *Psychological Review,* 55: 95-107.
MacKenzie, SB, Podsakoff, PM, Jarvis, CB (2005) The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions. *J. of Applied Psychology,* 90(4): 710-730.
MacKenzie, SB, Podsakoff, PM, Podsakoff, NP (2011) Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Quart.,* 35(2): 293-334.
Markus, ML (1983) Power, Politics, and MIS Implementation. *Communications of the ACM,* 26(6): 430-444.
Matavire, R, Brown, I (2013) Profiling Grounded Theory Approaches in Information Systems Research. *European J. of Infor. Syst.,* 22: 119-129.
Mathieu, JE, DeShon, RP, Bergh, DD (2008) Mediational Inferences in Organizational Research: Then, Now, and Beyond. *Organiz.ational Res. Methods,* 11(2): 203-223.
Meade, AW, Craig, SB (2012) Identifying Careless Responses in Survey Data. *Psychological Methods,* 17(3): 437-455.

Messick, S (1989) Validity. In RL Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Michell, J (1997) Quantitative Science and the Definition of Measurement in Psychology. *British J. of Psychology,* 88: 355-383.

Mingers, J (2001) Combining IS Research Methods: Towards a Pluralistic Methodology. *Infor. Syst. Res.,* 12(3): 240-259.

Moore, GC, Benbasat, I (1991) Development of an Instrument to Measure the Perceptions of Adopting and Information Technology Innovation. *Infor. Syst. Res.,* 2: 192-222.

Nunnally, JC, Bernstein, IH (1994) *Psychometric Theory* (Third ed.). New York: McGraw-Hill Inc.

O'Reilly, K, Paper, D, Marx, S (2012) Demystifying Grounded Theory for Business Research. *Organiz.ational Res. Methods,* 15(2): 247-262.

Ortiz de Guinea, A, Webster, J (2013) An Investigation of Information System Use Patterns: Technological Events as Triggers, the Effects of Time, and Consequences for Performance. *MIS Quart.,* 37(4): 1165-1188.

Pare, G (2004) Investigating Information Systems with Positivist Case Research. *Communications of the Association for Infor. Syst.,* 13: 233-264.

Payne, S, Finch, JF, Tremble Jr., TR (2003) Validating Surrogate Measures of Psychological Constructs: The Application of Construct Equivalence to Archival Data. *Organiz.ational Res. Methods,* 6(3): 363-382.

Petter, S, McLean, ER (2009) A Meta-Analytic Assessment of the DeLone and McLean IS Success Model: An Examination of IS Success at the Individual Level. *Information & Management,* 46: 159-166.

Reichardt, CS, Rallis, SF (1994) Qualitative and Quantitative Inquiries are Not Incompatible: A Call for a New Partnership. In CS Reichardt SF Rallis (Eds.), *The Qualitative-Quantitative Debate: New Perspectives* (Vol. 61, pp. 85-91). San Francisco: Jossey Bass Publishers.

Rossiter, JR (2002) The C-OAR-SE Procedure for Scale Development in Marketing. *International J. of Res. in Marketing,* 19: 305-335.

Sarker, S, Lee, AS (2003) Using a Case Study to Test The Role of Three Key Social Enablers in ERP Implementation. *Information & Management,* 40(8): 813-829.

Sartori, G (1970) Concept Misformation in Comparative Politics. *The American Political Science Review,* 64(4): 1033-1053.

Schriesheim, CA, Cogliser, CC, Scandura, TA, Lankau, MJ, Powers, KJ (1999) An Empirical Comparison of Approaches for Quantitatively Assessing the Content Adequacy of Paper-and-Pencil Measurement Instruments. *Organiz.ational Res. Methods,* 2(April): 140-156.

Segars, AH, Grover, V (1993) Re-Examining Perceived Ease of Use and Usefulness: A Confirmatory Factor Analysis. *MIS Quart.,* 18(4): 517-525.

Shanks, G (2002) Guidelines for Conducting Positivist Case Study Research in Information Systems. *Australasian J. of Infor. Syst.,* 10(1): 76-85.

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

Shoemaker, PJ, Tankard, JW, Lasorsa, DL (2004) *How to Build Social Science Theories*. Thousand Oaks, CA: Sage Publications.

Spector, PE, Brannick, MT (2011) Methodological Urban Legends: The Misuse of Statistical Control Variables. *Organiz.ational Res. Methods,* 14(2): 287-305.

Spiggle, S (1994) Analysis and Interpretation of Qualitative Data in Consumer Research. *J. of Consumer Res.,* 21(3): 491-503.

Stern, PN (2007) On Solid Ground: Essential Properties for Growing Grounded Theory. In A Bryant K Charmaz (Eds.), *The SAGE Handbook of Grounded Theory* (pp. 114-127). London: Sage Publications.

Stevens, SS (1946) On the Theory of Scales of Measurement. *Science,* 103: 667-680.

Stock, B (1983) *The Implications of Literacy: Written Language and Models of Interpretation in the Eleventh and Twelfth Centuries*. Princeton, N.J.: Princeton University Press.
Straub, DW (1989) Validating Instruments in MIS Research. *MIS Quart.,* 13(2): 147-169.
Straub, DW, Boudreau, M-C, Gefen, D (2004) Validation Guidelines for IS Positivist Research. *Communications of The AIS,* Vol. 13(Article 24): 380-427.
Strauss, A (1987) *Qualitative Analysis for Social Scientists*. Cambridge, UK: Cambridge University Press.
Sun, H, Fang, Y, Hsieh, JP-A (2014) Consuming Information Systems: An Economic Model of User Satisfaction. *Decision Support Systems,* 57: 188-199.
Tambe, P, Hitt, LM (2014) Measuring Information Technology Spillovers. *Infor. Syst. Res.,* 25(1): 53-71.
Todd, PA, Benbasat, I (1987) Process Tracing Methods in Decision Support Systems Research: Exploring the Black Box. *MIS Quart.,* 11(4): 493-512.
Trenkler, G, Stahlecker, P (1996) Dropping Variables Versus Use of Proxy Variables in Linear Regression. *J. of Statistical Planning and Inference,* 50: 65-75.
Urquhart, C, Lehmann, H, Myers, MD (2010) Putting the "Theory" Back into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. *Infor. Syst. J.,* 20: 357–381.
Van Leeuwen, TM (1981) *Surplus of Meaning: Ontology and Eschatology in the Philosophy of Paul Ricoeur*. Amsterdam: Rodopi B.V.
Van Maanen, J (1979) The Fact of Fiction in Organizational Ethnography. *Administrative Science Quart.,* 24(Dec): 539-550.
Van Maanen, J, Pentland, BT (1994) Cops and Auditors: The Rhetoric of Records. In SB Sitkin RJ Bies (Eds.), *The Legalistic Organization* (pp. 53-90). Beverly Hills: Sage.
Venkatesh, V, Brown, SA, Bala, H (2013) Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quart.,* 37(1): 21-54.
Webb, E, Campbell, DT, Schwartz, RD, Sechrest, L (2000) *Unobtrusive Measures: Non-Reactive Research in the Social Sciences*. Thousand Oaks, CA: Sage Publications (Sage Classics Edition, Original Publication, 1966).
Wennberg, K (2005) Entrepreneurship Research Through Databases: Measurement and Design Issues. *New England J. of Entrepreneurship,* 8(2): 9-19.
Yin, RK (1994) *Case Study Research: Design and Methods* (2nd ed. Vol. 5). Thousand Oaks, CA: Sage Publications.
Zetterberg, HL (1965) *On Theory and Verification in Sociology* (3rd ed.). New York: The Bedminster Press.