Thinking about Measures and Measurement

Andrew Burton-Jones  
University of British Columbia  
andrew.burton-jones@sauder.ubc.ca

Allen S. Lee  
Virginia Commonwealth University  
allenslee@alum.mit.edu

Abstract

This essay situates measurement in its larger context. Relying on some basic concepts from formal logic, the essay shows that current conceptions of “measurement issues” comprise just a subset of the larger universe of issues concerning measurement. New issues, going beyond the current conceptions, are identified such as the need for attention also to be given to the measurement of relationships, not just the measurement of constructs. Our analysis of these issues has several ramifications for empirical research. In particular, it highlights the cascading or cumulative impacts that errors in measures and measurement have in successive building blocks of empirical scientific reasoning. Overall, the essay demonstrates the need for more emphasis to be placed on fundamental issues involved in measuring and measurement and provides a framework that can help researchers and reviewers consider and evaluate these efforts. Our arguments apply to all empirical research, whether quantitative or qualitative, confirmatory, or exploratory.

1. Introduction

What is measurement? And what philosophy is behind it?

Measuring is a process by which we, as empirical researchers, connect our theories to the empirical world we are studying. We do this whether we explicitly use the terms “measure” and “measurement” (as we do in quantitative research) or not (as in qualitative research). In this essay, instead of creating measures and using them to take measurements, we philosophically examine their logical foundations and offer a new way of thinking about them. We start by introducing four terms from formal logic that we use as the building blocks in our discussion: statement variables, statement constants, individual variables, and individual constants. We then use them to build four categories following from two dimensions: whether we are dealing with measures or with measurements, and whether we are measuring an empirical referent or a relationship between empirical referents. An innovation is to apply the terms measure and measurements not only to variables, but also statements. This allows us to proceed to identify primary issues – both “measure issues” and “measurement issues” – in each of the four categories. We then delineate ramifications of our analysis: the cascading of errors in measures and measurements from each of the four categories to the others and the need for researchers to engage in additional work to improve their measures and measurements.

2. Definitions

We begin this essay with definitions because an interesting feature of research on measurement is that there is no standard definition of measurement [5]. Perhaps the most common definition of that term is the assignment of symbols (such as numbers or words) to objects or events according to rules [32]. Nonetheless, this definition is not always accepted [24]. To help address the lack of a standard definition, we build on the following concepts from formal logic: statement variables, statement constants, individual variables, and individual constants. We will use these concepts to build definitions of the following related, but distinct, terms: measurement, measure (as a noun and a verb), measuring, and empirical referent.

In the proposition, “if \( p \) is true, then \( q \) is true,” \( p \) is a statement variable that, like an algebraic variable, can take a particular value, such as \( \text{all humans are mortal} \), where \( \text{all humans are mortal} \) is an example of a statement constant. In the same manner, the statement variable \( q \) can take the particular value \( \text{Socrates is mortal} \), for which \( \text{Socrates is mortal} \) is an example of a statement constant. Finally, \( \text{humans} \) itself is an individual variable that takes the particular value \( \text{Socrates} \), where \( \text{Socrates} \) is an individual constant. Note that the adjective “individual” does not necessarily refer to a human individual. “Return on investment” or “ROI” can be an individual variable, for which “7%” could be an individual constant.

To measure is to assign, to a variable, an actual 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
What, then, would a "measurement issue" be about? It therefore, whether or not BI is a good measure is more assigned to this measure. In our terminology, the process by which the score of 1.3 comes to be of BI, which we regard as a on a questionnaire. We regard 1.3 as a score that a person gives to several questions about BI This constant might reflect, for example, the average given person, such as 1.3, is an score that a person gives to several questions about BI.

3. Situating measures and measurements in four categories

Consider the construct, intelligence, and its well known measure, IQ. How well IQ captures or describes people’s intelligence is often called a “measurement issue.” However, because IQ itself is a measure and not a measurement, the term “measurement issue” would be more suitable. What, then, would “measurement issue” instead refer to? This question provides the motivation for situating measurement in its larger context. We use Table 1 for this purpose.

Using the original version of the technology acceptance model [9], we identify U (a person’s perceived usefulness of a technology), A (his/her attitude toward the technology), and BI (his/her behavioral intention to use this technology) as individual variables. Note that the researcher must decide how to operationalize a theoretical construct in the form of an individual variable, e.g., one researcher may use one set of questions to indicate perceived usefulness while another researcher may use a different set [6].

A particular numerical value that BI takes for a given person, such as 1.3, is an individual constant. This constant might reflect, for example, the average score that a person gives to several questions about BI on a questionnaire. We regard 1.3 as a measurement of BI, which we regard as a measure. Measuring is the process by which the score of 1.3 comes to be assigned to this measure. In our terminology, therefore, whether or not BI is a good measure is more so a “measure issue” than a “measurement issue.” What, then, would a “measurement issue” be about? It would be about, for instance, how accurate the measurement 1.3 is.

The relation “BI=β₀+β₁A+β₂U” is a statement variable. When applied to a population through a random sample, it can be instantiated as or measured with the statement constant, “BI=2.3+1.4A+2.2U.” In this way, we regard the general equation “BI=β₀+β₁A+β₂U” as a measure and we regard the fitted equation “BI=2.3+1.4A+2.2U” as a measurement taken with it. The coefficients in the general equation have positive signs because they operationalize TAM propositions, which specify directions (if TAM did not specify directions, we would show neutral operators instead). The signs on the coefficients in the fitted equation reflect the direction of relationships measured in the data. A different population would result in a different measurement’s being taken (e.g., for a different population, the measurement could be “BI=0.2+1.1A-0.1U”). As for what we regard the statement variable “BI=β₀+β₁A+β₂U” to be a measure of, it is a measure of the relationship, “behavioral intention to use a technology increases when attitude towards the technology increases and also when perceived usefulness of the technology increases.” Note that other measures for this relationship can also be selected (e.g., a different measure, or statement variable, can involve discontinuous or even non linear relationships among individual variables).

In this essay, we are therefore innovating a broad conception of measures and measurements, where these two terms can refer both to statements and individual variables, a usage recently introduced [20, p. 248]. We recognize that this might seem unusual for researchers who do not think of themselves as 'measuring' relationships and instead use terms such as 'estimate' or 'infer' for this activity. We extend the terms measure and measurement to relationships deliberately to emphasize the common issues involved. We are not claiming that the assignment of individual constants to individual variables is identical to the assignment of statement constants to statement variables. In fact, our classification highlights that they are different types of measurement. Our point is that the logical process of assigning constants to variables is sufficiently similar for both cases to be referred to as “measurement.” Stating that one type of assignment is “measurement” while the other is not, would seem arbitrary. More importantly, a narrow view of measurement would mask the similarities involved in both types of assignment. Understanding these similarities is important because both types of assignment involve similar issues, which we address next.
Table 1: Situating measurement in its larger context

<table>
<thead>
<tr>
<th>What is being measured?</th>
<th>What is the measure?</th>
<th>What is the measurement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>an empirical referent, such as a person’s behavioral intention to use a technology</td>
<td>I. an <strong>individual variable</strong></td>
<td>Example: Consider the construct referring to a person’s behavioral intention to use a technology. In technology acceptance research, a measure of this construct is the <strong>individual variable</strong> BI.</td>
</tr>
<tr>
<td>a relationship between empirical referents (for instance, the relationship that a person’s behavioral intention to use a technology is related to the person’s perception of the usefulness of the technology)</td>
<td>II. a <strong>statement variable</strong></td>
<td>Example: Consider the relationship that a person’s behavioral intention to use a technology increases when his or her attitude towards the technology increases and also when his or her perceived usefulness of the technology increases. In technology acceptance research, a measure of this relationship is the <strong>statement variable</strong> “BI=β0+β1A+β2U.” This statement variable is an operationalization in the form of a general equation.</td>
</tr>
<tr>
<td>an empirical referent</td>
<td>III. an <strong>individual constant</strong></td>
<td>Example: A measurement taken with the measure BI could be an <strong>individual constant</strong>, the number 1.3 (resulting from the scores given by a person on a questionnaire).</td>
</tr>
<tr>
<td>a relationship between empirical referents</td>
<td>IV. a <strong>statement constant</strong></td>
<td>Example: A measurement taken with the measure “BI=β0+β1A+β2U” could be a <strong>statement constant</strong>, the fitted equation “BI=2.3+1.4A+2.2U” (resulting from applying the general equation to a given population).</td>
</tr>
</tbody>
</table>

Table 2: Primary measurement and measure issues

<table>
<thead>
<tr>
<th>What is being measured?</th>
<th>What is the measure?</th>
<th>What is the measurement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>an empirical referent</td>
<td>I. an <strong>individual variable</strong></td>
<td>Primary “measure” issue: - Does the individual variable reflect the empirical referent as theorized (i.e., the construct)?</td>
</tr>
<tr>
<td>a relationship involving empirical referents</td>
<td>II. a <strong>statement variable</strong></td>
<td>Primary “measure” issue: - Does the statement variable reflect the relationship as theorized?</td>
</tr>
<tr>
<td>an empirical referent</td>
<td>III. an <strong>individual constant</strong></td>
<td>Primary “measurement” issue: - Does the individual constant reflect the particular instantiation of the empirical referent?</td>
</tr>
<tr>
<td>a relationship between empirical referents (for instance, the relationship that a person’s behavioral intention to use a technology is related to the person’s perception of the usefulness of the technology)</td>
<td>IV. a <strong>statement constant</strong></td>
<td>Primary “measurement” issue: - Does the statement constant reflect the particular instantiation of the empirical relationship?</td>
</tr>
</tbody>
</table>

Before addressing these issues, however, it is important to emphasize that our arguments should be interpreted with the maturity of a research program in mind. In the earliest stages of research, simply distinguishing among variables, statements, measures, and measurements can be an achievement. Over time, it typically becomes easier to make finer distinctions among them. The principles we offer in this paper should be applied by researchers in light of the research stage at which they find themselves.

4. Measurement and measure issues

The four logical concepts we have discussed so far allow us to propose a way to assess “rigor” in empirical research. It is the justification that researchers provide for the individual constants, individual variables, statement variables, and statements constants in their work. To facilitate such justifications, this section discusses the primary issues that researchers need to address for each
logical concept. These issues are summarized in Table 2 and described in the following subsections. In discussing these issues, our aim is not to suggest cut-off scores for construct loadings, reliability coefficients, or p-values. Rather, our aim is to lay out the issues that researchers need to address when they think about rigor in measures and measurement.

4.1 Primary issue with individual variables

The primary issue with individual variables is whether they reflect the empirical referent being theorized (i.e., the construct). This can be assessed by judging the extent to which the content expressed in the definition of the construct (its conceptualization) is shared by the content expressed in the empirical indicators for the construct (its operationalization). As Figure 1 shows, two general problems can occur. First, the empirical indicators may be imprecise, indicating something in addition to, or other than, the theoretical construct. Figure 1 depicts the case in which the indicators reflect something in addition to the construct, such as some other construct that the researcher did not intend to measure. Second, the empirical indicators may be incomplete, failing to indicate a relevant aspect of the theoretical construct. Note that this is not simply a problem of using, say, a relatively small number of indicators (which is often a pragmatic necessity). Rather, incompleteness occurs when the empirical indicators exclude relevant dimensions or aspects of the theoretical construct being studied. In the terminology of validity theorists [23], both of these problems would indicate a lack of “content validity” in the empirical indicators.

Obviously, a precondition to achieving shared meaning between empirical indicators and theoretical constructs is to ensure that the meaning of each one is clear. For example, describing weaknesses in many manuscripts, a reviewer for a top journal wrote: “[if] you only have a vague idea of what you are trying to measure, it is easy for the measures to become contaminated by unrelated factors and/or for them to underrepresent the conceptual domain” [22, p. 324].

The importance of achieving a shared meaning between empirical indicators and theoretical constructs is familiar to IS researchers. For example, IS researchers have created procedures (such as card sorting) to help create empirical indicators that have high content validity [26] and have critiqued the content validity of existing empirical indicators and proposed revised ones in their place [29].

Note that achieving shared meaning between theoretical concepts and empirical indicators is a relevant goal in all empirical research, not just some. For example, in quantitative research, a major distinction is often made between formative and reflective measures [4], but this issue applies irrespective of whether the empirical indicators reflect or form the construct. It also applies irrespective of whether a researcher creates the theoretical concept first and then creates empirical indicators (the typical sequence in confirmatory factor analysis) or whether a researcher creates the empirical indicators first and then defines the applicable theoretical concept (the typical sequence in exploratory factor analysis and some forms of cluster analysis) [14]. In any of these cases, the empirical indicators and the construct they measure should correspond to each other.

Qualitative researchers are also concerned with this issue. For example, qualitative researchers often assign codes to segments of data that they have recorded about a field setting (such as records of interview comments or observations). In this case, the segments of data are the empirical indicators and the codes are the constructs. Coding, in other words, requires the researcher to recognize “that a unit of data is an empirical indicator of a more general construct of interest” [31, p. 493]. To perform such coding rigorously, qualitative researchers often report the steps they took to ensure that their coding was complete (i.e., relevant segments of data were not missed) and that codes were applied consistently. As with quantitative research, these practices can be performed irrespective of whether codes are developed before or during data analysis, for even in the latter case, the goals is to achieve “a close fit between the research claims and the data that inspired them” [11, pp. 424-425].

![Figure 1: Problems in mapping empirical indicators to theoretical constructs](image-url)
4.2 Primary issue with individual constants

The primary issue with individual constants is whether they reflect the instantiation of the empirical referent. In other words, are the measurements of the instantiations accurate? When considering this issue, it is useful to consider how inaccuracies can arise in primary data vis-à-vis secondary data. Primary data are data that would not have been produced except for a given research study. Because such data are obtained through the use of a researcher’s data collection methods, the major reason that inaccuracies occur in such data is that the methods themselves lead to inaccuracies in the data—the general problem of “method bias” [8]. For example, a researcher may interview employees to ask them whether they will adopt a new technology in their work. Because of the social context of the interview, users may say that they will adopt the technology, even if they plan to resist it, because they may think that such an answer is more socially acceptable.

Secondary data are data that are produced independently of a given research study but that a researcher subsequently appropriates for his/her study. In such data, the problems of method bias can occur in two ways. First, method biases can occur in the original production of the data. For example, organizational records of an event might not reflect what transpired, but what stakeholders wished others to think transpired [37]. Second, even if the original records are accurate, method biases can occur if researchers make errors in interpreting or transforming the data in their research studies.

Researchers have long stated the importance of being sensitive to possible inaccuracies in primary and secondary data [3, 38]. In so far as qualitative research typically involves intensive, rather than extensive, data gathering, qualitative researchers typically more often than quantitative researchers get close to the context being observed and, thereby, reduce the likelihood of inaccuracies creeping into their data [3]. For example, if a quantitative researcher asks 1000 users to respond to a questionnaire regarding the quality of a system, he or she is unlikely to be able to check the truthfulness of each of these 1000 responses. In contrast, a qualitative researcher might intensively interview a handful of key users, but by immersing herself in the context of the study, she may be able to identify inaccuracies more easily. The following quotation provides an illustration, demonstrating Klein and Myers’ [19] principle of suspicion (not taking informants comments at face value) at work: “The first author’s training and hands-on experience … allowed her to understand interviewees’ comments about the system and to assess interviewees’ statements about the capabilities of the software. For instance, she reframed a user’s complaint about Compass’s inability to print financial transactions as the user’s lack of knowledge about the software (she knew that printing was permissible). [On some occasions, the first author realized that] interviewees … considered her to be a member of the development team because of her participant role. On these occasions, the first author sensed that interviewees were insincere rather than genuine” [17, p. 7].

A general strategy to ensure accuracy is to (a) understand the process by which data are produced and recorded and (b) identify and account for any instances in this process where inaccuracies occur [8]. In qualitative work, implementing this strategy involves taking steps to understand informants’ language and the context in which documents in a field setting are produced, and behaviors in that setting performed [3, 15]. Likewise, in quantitative work, implementing this strategy involves taking steps to understand how respondents interpret researchers’ language and how secondary data such as records of behavior are produced [17, 38]. In both research traditions, it also involves researchers understanding how their biases could improperly, and their pre-understandings could constructively, influence their interpretations [27].

4.3 Primary issue with statement variables

The primary issue with statement variables is whether they reflect the empirical relationship as theorized. When considering this issue, it is useful to recognize the very wide variety of relationships that can be theorized. For instance, consider the following relationship stated in Table 1: “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=β0+β1U. Note that this is just one of many ways that it could be operationalized. Even if we assume, for simplicity, that a researcher uses a regression model to operationalize it, he or she can consider: “the form of the relationship (linear, curvilinear, or power), the direction in which the variables are related (positive or negative), the coefficients (constant and slope), and the limits [such as continuous or limited to a particular range of values]” [30, p. 59].

The number of ways that relationships can be operationalized increases even further when more individual variables are added. For example, in TAM, there are two independent variables, attitude and perceived usefulness. The relationship between
these independent variables and the dependent variable, behavioral intention, was specified in TAM 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, over and above whatever positive or negative feelings may be evoked toward the behavior per se” [9, pp. 985-986].

If we consider this quotation in light of the options noted above, we would conclude that the authors have specified the direction of the relationships (i.e., positive), but have not specified the form of the relationships (linear, curvilinear, or power), the coefficients (such as high or low constant, and high or low slope), or the limits of the relationships (such as whether they hold for a particular range of A, U, or BI only). We make this point to bring attention to the additional landscape that researchers need to cover in order to improve the rigor with which they operationalize a theory.

Researchers can assess the rigor with which a theory is operationalized in the same way as they assess individual variables, as noted earlier: by assessing the extent to which the content expressed in the conceptualization of the relationship is shared by the content expressed in the operationalization of the relationship. Shoemaker et al. [30, pp. 54-56] provide a simple way to make such an assessment. 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 in which the operationalized relationship is described (which could be, but need not be represented as, an equation, such as a regression equation). A researcher can then examine the correspondence between the columns of the table.

Researchers can also take steps during theory development and testing to improve the link between the conceptualized and operationalized statements of a theoretical relationship. In confirmatory research, one way to do so is to conceptualize a theory as precisely as possible at the outset so that its operational form is clear [2]. Exploratory researchers can pursue the same aim working in the opposite direction. Using the statistical technique TETRAD, for example, researchers can analyze all of the possible relationships that can exist among a set of variables and discover the most plausible subset. The aim is to help the researcher understand the likely conceptual nature of a relationship by providing information about which operational relationships are supported or refuted in a data set [16].

The grounded theory method is a good example of a method that shares some qualities of both exploratory and confirmatory research because it requires researchers to constantly compare their emerging constructs and relationships to their data [36]. Recall from our discussion of individual variables that, in exploratory qualitative work, the operational statements are the segments of data (such as interview comments or records of observations) and the conceptual statements are the codes and categories applied to these segments of data [31]. A key element of grounded theory is the coding of relationships among categories. Urquhart et al. [36, p. 11] stress: “how much attention is paid to the precise nature of the association between constructs is critical to theorizing.” As with any part of grounded theory, the description of these relationships should be closely tied to the data because this allows the researcher to ensure that the conceptual relationships they abstract from the data are grounded in the operational relationships observed in the data, a state known as theoretical saturation: “… coding was considered complete when theoretical saturation was obtained, that is, when no new or relevant data informed a category, when the category development was densely populated, and when the relationships between categories were supported by ample data” [7, p. 8].

4.4 Primary issue with statement constants

The primary issue with statement constants is whether they reflect the actual, or true, instantiations of theorized empirical relationships. Simply, is the measurement of the instantiation accurate? To make this assessment, at least three approaches can be considered, which we refer to as verifying: predictions, processes, and honesty or truth. All three provide ways to assess the theory expressed in a statement constant; each assesses the extent to which it is a complete and/or honest reflection of reality.

The first approach is to make predictions and assess the accuracy of the theorized relationship by seeing whether the prediction holds. A recent exposition of this approach is offered in Lee and Hubona [20, pp. 246-255] and involves the use of prediction intervals. Consider, for example, the statement variable or “measure” in the form of the general equation $BI=\beta_0 + \beta_1 A + \beta_2 U$ (where BI is behavioral intention, A is attitude, and U is perceived...
usefulness) and the corresponding statement constant or “measurement” in the form of the fitted equation \(BI=2.3+1.4A+2.2U\), which results from applying the equation to a given population, consisting of the persons in a given organization. This statement constant makes predictions that the value of BI for any sample data point (person) in the population is a function of an intercept (2.3) and two coefficients (1.4 and 2.2). Researchers can test such a prediction by, first, selecting an individual who was part of the population studied, but not part of the sample from which the equation was estimated, measuring the values that this person has for A, U, and BI, and then comparing the predicted value of BI for this individual with its observed value. If the prediction falls outside the prediction interval—outside a predetermined “ball park” range considered acceptable for the predicted value—this will indicate that the relationships, as indicated in the statement variable or measure, are not accurate. Although this is a quantitative, statistical example, the same logic can also be used with qualitative and interpretive research, as Lee and Hubona [20] explain.

A second approach is to verify the underlying process that an empirical relationship is assumed to reflect. At this point, recall our earlier discussion of the technology acceptance model and our argument that many technology acceptance researchers assume that the relationships they estimate reflect real-world processes. For example, Kim and Malhotra [18, p. 743] write: “TAM presumes that [BI] is formed as a result of conscious decision-making processes.” One way to test the accuracy of the relationships identified in a set of data, therefore, is to verify whether they actually reflect the process that was presumed to give rise to that relationship. For example, in a study of TAM, this would involve determining whether users engage in conscious decision-making processes before forming their intention to use a system (or whether they, instead, simply use a system based on habit). Note that the verification of “accuracy” in this case has a slightly different meaning here than it did in the test of prediction intervals. In the case of verifying processes, accuracy refers to whether the researcher is making an accurate assumption about the process that gives rise to the observed relationship.

How can the accuracy of such an assumption be evaluated? One way is to use a data collection method that allows the researcher to obtain evidence about this process. For example, experimental researchers have long used protocol analysis and process tracing methods to verify their assumptions about the processes that individuals engage in [35]. These methods are not limited to any particular philosophical position. As Weber [39, p. vii] points out, “the data collected from protocol studies are often analyzed from an interpretive perspective as well as a positivist perspective.” Although we are not familiar with any process tracing studies of TAM, such an approach could be used. For example, researchers could use fMRI techniques to examine the extent to which individuals really do engage in conscious decision-making process when forming their intentions to use a system [10].

A third approach to testing the accuracy of statement constants is to verify the truth of relationships found in the data. This approach becomes relevant in studies where researchers have data regarding a relationship or process. For example, rather than collect process tracing data about how people decide to use a technology using fMRI, a researcher might obtain verbal protocols [13]. The accuracy of these verbal protocols clearly depends on whether subjects tell the truth about what they are thinking (rather than tell the researcher what they think s/he wishes to hear). The same issue can apply in field research. For example, Eisenhardt and Graebner [12, pp. 29-30] write: “It is … crucial to [determine] the underlying theoretical arguments that provide the logical link between the constructs …. These arguments can be drawn from case evidence (e.g., an informant explaining the logic) ….” In such cases, the accuracy of the data obtained depends on whether informants tell the truth or instead merely explain relationships in the way they would like to explain it, or in a way they think the researcher would like to hear. Such issues are simply more manifestations of the general problem of method bias referred to in our discussion of individual constants and can be addressed by understanding where potential biases can arise in data collection and using multiple methods (such as observation as well as interviews) to minimize the impact of these biases on the researcher’s conclusions [12, p. 28].

4.5 Summary of primary issues

Figure 2 summarizes our arguments. The figure distinguishes between theory, reality, and empirical research, where the last of these serves as a bridge between the first two. The figure makes no distinction between individual variables and constants and statement variables and constants because the basic issues apply to both logical concepts equally. Also, the figure makes no firm distinction between operationalizations and data. In confirmatory work, operationalizations and data can be clearly distinguished. For example, researchers might operationalize a construct by creating
questions for a structured interview script. Responses to these questions are then considered data. In exploratory work, however, a researcher might conduct unstructured interviews and transcribe all of the interview comments. He or she may, then, later use segments of these transcriptions as operationalizations of a theoretical concept [31]. The distinction between operationalization and data then disappears. In short, our conception of research in Figure 2 is designed to be applicable to any type of empirical research in which IS researchers engage. As the figure highlights, the primary “measure” issue in empirical research is the extent to which there is shared meaning or correspondence between researchers’ conceptions and operationalizations, and the primary “measurement” issue is the accuracy of data obtained about the world. Both issues need to be addressed. When accuracy is high, researchers can say that they have measured something successfully. When accuracy and shared meaning are high, researchers can say that they have measured successfully what they thought they measured.

The issues we have identified are sufficiently general that they can apply to all empirical research, irrespective of the method a researcher uses. Thus, they can be considered foundational criteria for assessing measures and measurements in empirical research. Researchers have called for such criteria [6, p. 13] but they have not been forthcoming, perhaps because of the vast differences in methodological techniques and terminologies across research communities. Our analysis helps to address this problem. It also does so in a parsimonious way. This is helpful because not only do researchers suffer from a lack of common frameworks for thinking about measures and measurement in all varieties of empirical research, but they also suffer from a multitude of different frameworks and ideas for improving measures and measurement in specific types of empirical research [see, e.g., 1, 21, 31]. This paper, serves as a counterpoint; it strips the problem down to its core and highlights the small set of fundamental issues that researchers should pay most attention to.

5. Major ramifications

A major ramification flows from our analysis that has implications for how empirical research is conducted in the IS discipline. We describe it as one of cascading errors. It refers to the fact that problems in accuracy and shared meaning that we have outlined can cascade through the process of empirical research. In one sense, they can be thought of as accumulating as researchers move horizontally through the realms in Figure 2. For confirmatory researchers, the problems accumulate as research moves from the left to the right of the figure. For example, if confirmatory researchers fail to ensure shared meaning between their theoretical constructs and their operationalizations, this will place an inherent limit on their ability to measure the constructs they theorized. Taking steps to obtain accurate measurements (for example, by reducing method bias) can reduce the likelihood of problems on the right side of Figure 2 affecting the data, but even if complete accuracy is obtained, the measurements will still not reflect the construct theorized because of the lack of shared meaning. A recent example is evident in [8]. After demonstrating his approach for addressing method bias, the author acknowledges that his demonstration was limited because he did not conclusively verify the content validity of his instruments: “Regrettably, strong evidence is lacking …. Additional studies would be needed to refine the instruments, particularly in terms of their content validity” [8, p. 463].

For exploratory researchers, problems accumulate moving from the right to the left of Figure 2. As a result, no matter how sophisticatedly such researcher explores their data, they cannot have faith that their conclusions reflect reality unless steps are first taken to ensure the accuracy of the data they used in the exploratory analysis.
In addition to problems accumulating horizontally in Figure 2, they can also be thought of as accumulating as research moves from the study of individual constants and variables to statement constants and variables, because statements involve relationships among concepts. In confirmatory quantitative research, Straub [33] emphasized this issue twenty years ago. We simply emphasize here its generality for all types of empirical research.

Overall, the cascading effect of problems with measures and measurements implies the benefit of moving back and forth between the three realms of Figure 2, checking one’s conceptualizations, operationalizations, and understanding of the empirical domain over time. Likewise, it implies the benefit of moving back and forth between one’s understanding of the constructs, and the relationships among constructs, in one’s research.

6. Conclusions

Creating measures and obtaining measurements are important activities that all empirical researchers engage in. Unfortunately, researchers do not have good definitions of these activities, nor do they have clear agreement about how best to assess them [21]. Our essay has contributed by defining these activities, clarifying their scope, and identifying the primary issues that researchers should consider when assessing how rigorously they are conducted.

We recognize, of course, that these activities are only a small subset of the many activities that researchers engage in. Moreover, while good empirical research requires acceptable measures and measurements, too much emphasis on these matters can sometimes stifle innovative ideas—and these ideas are researchers’ most important contributions [34]. When thinking of measures and measurement, therefore, researchers must balance the rigor that they ask of themselves, and others, with the need for innovation and relevance.

We addressed our topic in a general manner that can apply to all empirical research, regardless of methodological approach, as called for by Boudreau et al. [6, p. 13]. To do so, we did not choose the terminology of just one community (such as the language of confirmatory, quantitative work, or the language of interpretive work) and try to show how it can apply to all research. This approach has been tried in the past [1], but it is generally ineffective because the terminology used in any given community is typically a product of the interests and history of that community and, therefore, may not translate to other communities effectively [28, p. 42]. Instead, we adopted a language that all researchers can understand, the language of logic. Logic applies to all research because all researchers can engage in logical reasoning. By adopting such a common language, we hope that our essay can provide a platform for shared discourse among researchers in our field and will improve mutual understanding and appreciation of the common challenges that all researchers face when they engage in empirical work.

7. Acknowledgments

We thank our reviewers for very helpful comments.

8. References


Sandberg, J. "How Do We Justify Knowledge Produced Within Interpretive Approaches?," *Organizational Research Methods* (8:1), 2005, pp 41-68.


