Journal of the Association for Information Systems
JAIS

Editorial Note

Eight Obstacles to Overcome in the Theory Testing Genre

Shirley Gregor
Australian National University
shirley.gregor@anu.edu.au

Gary Klein
University of Colorado, Colorado Springs
gklein@uccs.edu

Abstract

Theory testing work is popular in information systems (IS), with many studies using questionnaires, experiments, or other methods to gather quantitative data and test hypotheses with statistical techniques. This editorial note highlights some of the obstacles that theory testing researchers face, and assists authors so that their papers will not be rejected outright on submission, nor slowed unnecessarily in the review process. In particular, we identify three obstacles relating to theorizing and five related to methods. We provide guidance on how authors can deal with each obstacle, and include examples of studies that have successfully addressed the obstacle. We hope that our editorial will encourage authors to take better note of these obstacles and to give further consideration to avoiding them when they plan and conduct their research studies.

Keywords: Theory-Testing Research, Quantitative, Obstacles, Research Rigor, Research Validity.

---

* Suprateek Sarker was the accepting editor.

Volume 15, Issue 11, pp. i-xix, November 2014

Eight Obstacles to Overcome in the Theory Testing Genre

1. Introduction

Theory testing work is popular in information systems (IS), with many studies using questionnaires, experiments, or other methods to gather quantitative data and test hypotheses with statistical techniques. There are, however, common problems that arise with work of this type. There are many worthy, carefully performed studies that never see the light of day because they fail to overcome unseen obstacles. This paper highlights some of the obstacles that theory testing researchers face, and assists authors so that their papers will not be rejected outright on submission, nor slowed unnecessarily in the review process. In particular, we identify three obstacles relating to theorizing and five related to methods. We provide guidance on how authors can deal with each obstacle, and include examples of studies that have successfully addressed the obstacle. We hope that our editorial will encourage authors to take better note of these obstacles and to give further consideration to avoiding them when they plan and conduct their research studies.

We do not replace or reproduce the in-depth guidance to rigorous theory development, research methods, and statistical analysis that many textbooks and journal papers provide. Rather, we highlight some problems that are seen relatively often in papers submitted to the *Journal of the Association for Information Systems* (*JAIS*) and other leading journals. Given the streamlined and timely review process used at *JAIS*, the obstacles described in this editorial are often the cause of immediate desk rejects in the *JAIS* reviewing process. Editors provide the same advice many times to different authors (often, sadly, with a rejection notice), and it would be preferable to provide this collection of advice in a single editorial that authors could read before submission. While the obstacles described are not necessarily new and while we draw on some editorial advice from other journals, this collection of advice provides needed guidance targeted toward a specific genre of research that is readily accessible in a single source. This editorial does, however, reflect the authors' views, rather than *JAIS*'s official policy.

There are, of course, many other problems that could affect a manuscript's perceived value. We present eight obstacles that deal with research rigour and validity. There are other aspects of research that are important but lie outside the scope of the present discussion, including the significance and interestingness of the problem that the study addresses (see Barley, 2006; Straub, 2009), novel approaches such as neuroscience (Riedl, Davis, & Hevner, 2014), and whether the study addresses meaningful phenomena that are related to information systems. No study is perfect, and excellent work that has not overcome all of our obstacles will still be published. In this paper, we consider only "theory testing genre" papers, that is, studies that gather quantitative data in surveys, experiments and quasi-experiments, archival analysis, and similar methods and use this data to test hypotheses (Neuman, 2006)[1]. We do not consider work that gathers qualitative data (e.g., from case studies). Further, we reserve the term "field studies" for field research with ethnographic or participant-observer research (Neuman, 2006), and we also do not consider this work. Some of the problems we discuss can, however, occur in other genres: for instance, poor theorizing and poorly (or un-) defined concepts.

We now discuss the eight obstacles themselves (summarized in Table 1). The first three obstacles are theorizing-type obstacles and the last five are methods-type obstacles. In our experience, problems with theorizing are the single largest cause of paper rejection.

[1] Some may use the label "positivist" for work of this type, but we avoid this label because it is not clearly and consistently used in the literature.

Table 1. Eight Obstacles to Overcome in Theory Testing Research

Obstacle		Short description
Theorizing type	1. Research model deficiency	Theory development is not sufficiently strong. A theoretical research model is not well conceived. The rationale or justification for the model, in part or in whole, is not sufficiently clear and/or convincing.
	2. Causality confusion	Causality is ascribed in correlational research in situations where it is not warranted.
	3. Construct clarity problems	The meaning or label of a theoretical construct is imprecise or changes throughout a paper.
Methods type	4. Common method bias	Bias introduced by a common method of measurement is not recognized.
	5. Formative construct unrecognized	Problems due to the use of formative constructs in statistical analysis are not recognized.
	6. Self-report data validity issues	Measurement problems with self-report data are not recognized and/or addressed sufficiently.
	7. Sample selection limitation	Limitations of the sample selection are not recognized.
	8. Analytic technique problems	Problems with the statistical techniques used for analysis, which include the choice of tool for analysing data and moderating-effects problems.

2. Theorizing-Type Obstacles

2.1. Obstacle 1: Research Model Deficiency

This obstacle arises when the theorizing behind a research model is inadequate. The rationale or justification for the model, in part or in whole, is not sufficiently clear and/or convincing. This obstacle causes editors to desk reject many papers and, because it arises so often, we devote more space to it than the other problems. Nevertheless, it is difficult to provide in-depth guidance for theory development and model building in a short paper, and our treatment here is necessarily limited[2].

Having received them ourselves, many (if not all) academic authors will recognize editor and reviewer comments that indicate a problem with theory development, such as:

- Insufficient theoretical contribution / the theoretical contribution is not sufficiently clear / the theoretical contribution is not articulated.
- The model is under-researched.
- It would be helpful to know what other prior empirical studies have looked at antecedents of this dependent variable and know what they found.

[2] By “model” or “research model”, we mean the set of concepts (constructs) and relationships among them, often represented in a “boxes and arrows” diagram, that are tested in a study. This research model should be accompanied by justification and/or explanation for the relationships proposed. By “theory”, we refer to theory for explanation and prediction (type IV theory in Gregor’s (2006) taxonomy), which is theory that can be tested. A theory should be specified to meet Weber’s (2012) and Whetten’s (1989) (among others) requirements. A research model is distinguished from the statement of a theory because a single research model in a single study is unlikely to meet these criteria. Specifically, it is unlikely to provide, with any degree of confidence, the boundaries of generalizability or range of the theory: Whetten’s (1989, p. 492) “who”, where”, and “when” conditions. By “strong” or “full” theory, we mean well-established theory that is supported by a cumulative body of work (i.e., supported by empirical studies employing varied methods and providing diverse evidence (Hempel, 1966, p. 34)).

- There is a lack of an overarching theoretical perspective to justify why all theories explained in this study are important and understanding them together is necessary in the research context.

- Why are these eight variables in the research model and not some other variables?

More picturesque comments include:

- A kitchen-sink model (i.e., many potential explanatory variables are included, some of which are not really needed).

- The conceptual development is more like "bead stringing" rather than a "closely knitted" argument (i.e., the model is not well integrated theoretically).

- Castle confections built on sand (i.e., the model appears artificial and is not well grounded in theory or data) (Paul, 2005, p. 209).

This obstacle is not easily surmounted because model and theory building is one of the most challenging tasks for researchers, and, somewhat paradoxically, is perhaps more of an art than a science. Theory and model building can be seen as more of an intuitive and imaginative process rather than an easily defined method. Feldman (2004b, p. 565), in an editorial in the *Journal of Management*, comments:

> *We tell authors that their papers do not make theoretical contributions, but often do not give authors much insight into what counts as a theoretical contribution or how to build stronger theories.*

Although theory building is not a simple matter with any research approach, there appear to be problems that occur more with theory testing and especially with cross-sectional survey-based research. It may be that the very ease with which a large number of variables can be included in a survey instrument (compared with, for example, an experimental manipulation) means that researchers do not exhibit as much care in choosing what to include in their research design. There may be some element of "casting a net" to see what turns out to be significant. However, the statistics used can easily identify spurious relationships and well-founded causal links (see Section 2.2). Moreover, "statistical packages promote promiscuous statistical relationships" (Paul, 2005, p. 209).

Figure 1 shows the processes that we see in theory development over a period of time. "Theory-testing" studies are an important part of this process. They are classified as theory testing because they test propositions based on theory. Nevertheless, our journals expect that some degree of theory building should also take place in addition to the theory testing. The studies should be based on a research model that has the strongest foundations possible in the existing knowledge base, which can include strong theory (perhaps from a reference field), findings from exploratory work to ground the model, professional knowledge that represents "wisdom from the field", and prior empirical studies relating to the problem at hand (inclusive of research approach). Please see Colquitt and Zapata-Phelan (2007) for a fuller discussion of the tensions between theory testing and theory building in a single empirical paper. For example, moderate levels of theory building in a theory testing paper can arise from including a new substantive mediator or moderator to an existing relationship. A theory can be tested in a new context in which its applicability is not currently known. If results are positive then the boundaries in which the theory holds can be extended.

In contrast to theory-testing papers, theory-building empirical papers develop theory by inductive means and typically conclude with a set of propositions that summarize the new theory (Colquitt & Zapata-Phelan, 2007). Overall, the empirical studies represent "interim struggles" that move a field towards stronger theory (Weick, 1995). The evidence that accumulates from a stream of empirical papers over a period can be synthesized in theory- and review-type papers. All papers should lead to new, refined or extended theory that can be returned to the field's knowledge base to inform future studies. Pure replication studies, in which an existing theory is tested without any extension or refinement, are unlikely to be accepted in leading journals (however, whether this is good scientific practice is sometimes debated).

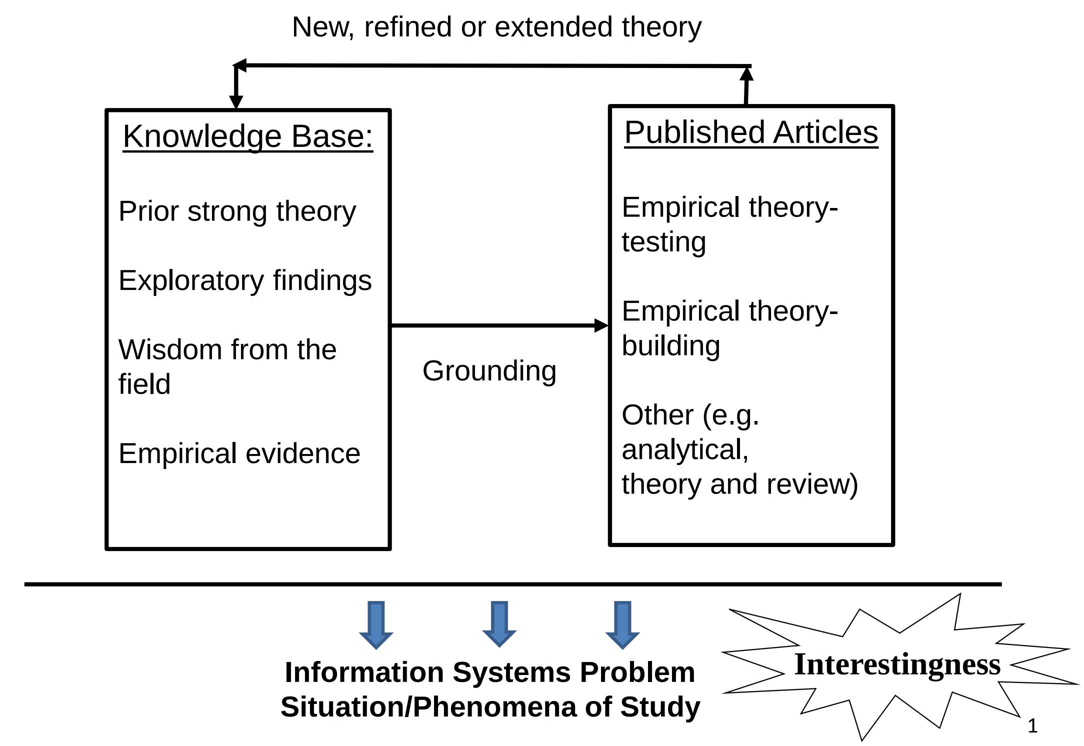

Figure 1. The Place of Empirical Theory-Testing Papers in the Theorizing Process

A good example of a theory-testing paper with some theory refinement is Xu, Benbasat, and Cenfetelli's (2011) paper on the effects of service and consumer product knowledge on online customer loyalty. This paper is founded on unified services theory and the service evaluation literature, and improves on prior research in that it proposes additional theory-based factors that influence online customer loyalty.

We suggest that authors:

1) Ensure that the literature review in their paper adequately identifies (from a variety of sources) prior theory that is relevant to the phenomena addressed and any prior relevant empirical work (see Webster & Watson, 2002). Prior empirical work may be summarized in a table (see Kim, Shin, Kim, & Lee (2011) for an example).

2) Perform some grounding work such as interviews or focus groups if prior theory does not give clear guidance on important explanatory factors. For example, Pavlou and Fygenson (2006) used an open-ended questionnaire to elicit the primary external beliefs relevant to e-commerce adoption before developing a detailed research model.

3) Read some of the excellent work on criteria for quality theory and theorizing by authors such as Weber (2012), Whetten (1989) and Feldman (2004b).

2.2. Obstacle 2: Causality Confusion

This obstacle arises when causality is ascribed in correlational research in situations when it is not warranted. This obstacle is linked to the first obstacle. It is often a concern about potential spurious causal relationships that underlies a reviewer's desire to see a well-justified research model with all important potential causal variables included in the model or included as controls. Existing strong theory or prior empirical studies can obviously suggest what these important potential causal factors are.

The problem is acute with survey methods. Neuman (2006, p. 276) says that "Survey researchers use control variables to approximate the rigorous test for causality that experimenters achieve with their physical control over temporal order and alternative explanations". The same problem occurs with relationships among several dependent variables in experiments.

Many papers have hypotheses that are cast in causal terms (i.e., X leads to Y, or X influences Y). However, with correlation- or covariance-based techniques (regression and SEM), it is rarely possible to test a cause-effect relationship. At most, what can be said is that there is support for an association (i.e., X is correlated with Y, or an increase in X is associated with an increase in Y).

While many published studies do attribute causality on the basis of correlational-type research, authors should heed methodologists' and statisticians' guidance (see Pearl, 2000), and more consistency and rigour is needed with claims of causality in survey research. In other fields, the problems are better recognized. In medicine, true experimental design is seen as the gold standard for inference of causality, whereas observational studies without controls (cross-sectional studies) provide one of the lowest grades of evidence (Geyman, 2000). Recognizing limitations of claims for causality should improve research practice overall.

Kline (2005, p. 95) states:

> *One should adopt the view that just as correlation does not imply causation, statistical causal modelling does not prove causation either. It is why Wilkinson and the Task Force on Statistical Inference (1999) emphasize that use of SEM computer programs "rarely yields any results that have any interpretation as casual effects".*

The requirements expected for attribution of causality are commonly those stated by John Stuart Mill (see Cook & Campbell, 1979, p. 18):

> *First, the cause has to precede the effect in time; second, the cause and effect have to be related; and third, other explanations of the cause-effect relationship have to be eliminated*.

Note the first requirement for attributing causality is time precedence. However when variables are concurrently measured, as in a cross-sectional survey, "it is not possible to demonstrate time precedence" (Kline, 2005, p. 95). Further, with measures of constructs representing emotions, attitudes, and human cognition, it is particularly difficult to specify a time sequence because emotion can influence cognition and cognition can influence emotion (Burton, Westen, & Kowalski, 2009), so extra care is needed.

A further requirement is that other potential explanations for the putative cause-effect relationship are eliminated. In non-experimental research, this requirement is difficult to meet. For example, a hypothesis could say "perception of a vendor's reputation influences intention to purchase from an online store". It is hard to tell in cross-sectional survey-based research if a perception of reputation influenced intention to purchase, or a third factor (e.g., the desire to avoid cognitive dissonance) led to the respondents giving suitably compatible answers to both the reported perception of reputation and the intention to purchase. (Note that many of the rival explanations for an outcome such as this one are also treated under common method bias in Section 3.1.)

As such, we suggest that authors:

1) Be aware of the limitations of correlational research with regard to the attribution of causality, and avoid claims supporting causal relationships when these claims are not warranted.

2) Work towards better approximations of causality by having well-founded research models that incorporate important alternative causal factors as far as possible (see Section 2.1).

3) If possible, supplement the quantitative data from a survey with some qualitative data that supports cause-effect linkages: for example, by including open-ended questions in a survey or by interviews. For example, Gregor, Lin, Gedeon, Riaz, and Zhu (2014) show that emotions influenced both cognitive evaluations and eLoyalty towards e-commerce websites with qualitative data that supplemented quantitative data.

2.3. Obstacle 3: Construct Clarity Problems

This obstacle arises when the meaning or label of a theoretical construct changes in a single paper or from its usage in other papers, or when the construct measure is deficient. Construct validity problems are the cause of a reasonable number of initial rejections of papers, where they constitute a "fatal flaw".

Constructs should not be used **loosely** and neither should there be **inconsistencies** in the references to these constructs. Theoretical claims that are made should be based on the constructs as defined and measured, not on other constructs that may appear similar, but represent different things and have different labels. See the advice from Feldman (2004a, p. 3):

> *Make sure you use consistent terminology… That is, if you label a term "strategic decision" in one place, then don't subtly change it to "strategic allocation" or "strategic direction" in another place. Consistency in language usage is valued highly by reviewers, but is probably not as salient an attribute to authors as it could be.*

This is called elegant variation—the practice of using different words for the same thing to avoid repetition.

Sometimes, this problem, more than just the loose use of language, creates a **logical disconnect** in the argument being made. For example, authors should not confuse human perceptions of a technology with properties of the technology itself. If a construct is defined to mean human perception, it should not be changed to mean a property of some physical thing. Human perception of ease of use is not the same thing as ease of use features possessed by a technology. Individual perceptions of ease of use will vary from one person to another depending on prior experience, expertise, and so on. A construct for a technology's ease of use denotes the possession by a specific technology of specific characteristics that contribute to its usability. Researchers can fall foul of this obstacle when there is a survey conducted and all respondents have experience of the same technology (e.g., one system in one company). In this example, the variation observed in responses is due to the variation in individuals' characteristics and perceptions, not to the technology's characteristics, which are the same for all.

Similarly, "behavioral intention" is not the same as "behavior", "intention to adopt" is not the same as "intention to continue using", and so on. These different constructs may have different theoretical backgrounds, and should take their meaning and understanding from the theory base. For example, "continuance intention" has different theoretical implications from "intention to use" (Li, Troutt, Brandyberry, & Wang, 2011). One might argue that two variables are closely correlated, but that is not the same thing as substituting one construct name for another without cause.

Problems can also occur when there is slippage between papers about the meaning of a construct. In reviewing treatment of construct validity, Boudreau, Gefen, and Straub (2001) note that metrics are sensitive to context, yet researchers tend to focus on whether a scale has been previously validated, rather than whether the metrics are fit for purpose in a new context.

Further, researchers include items that are poorly worded. Avoid double barreled questions such as: “Are you satisfied with the response time and error rate?”, where a single response is not adequate for both parts of the question. Also, avoid leading questions such as: “Because fewer people are involved in this community I am less likely to contribute”, which could suggest the hypothesis you want confirmed.

As such, we suggest that authors:

1) Heed advice for the necessity of clear construct definitions for good theorizing as in Weber (2012), Feldman (2004a), and Whetten (1989). See also Hovorka, Larsen, and Morachi’s (2009) work on concept convergence.

2) Heed advice on construct validity and use techniques that enhance validity, including pre-tests and pilot tests (Boudreau et al., 2001).

3) Re-validate any instrument used in a study no matter how established it is. A simple confirmatory factor analysis will do as a “goodness-of-fit” technique.

4) Include your survey instrument in the manuscript on submission (otherwise it will almost certainly be returned).

5) Take care in using labels for constructs: define each construct clearly and consistently use labels. Weber (2012) suggests providing a table that lists all construct names and their definitions. Do not confuse constructs: for example, do not substitute “perceptions of ease of use” for “ease of use” or “behavioral intention” for “behavior”.

6) Do not make claims that evidence supports a relationship between one set of constructs when a different set of constructs has been operationalized. That is, establishing that perceived ease of use is related to behavioral intention to use (psychological state research) is not the same as establishing that usability (referring to technology characteristics) leads to use (as in actual use).

An example of a paper that clearly defines constructs is Kim et al. (2011). This paper provides a table that defines its constructs in the paper body, and provides a table in the appendix with the scales that operationalize the constructs and references to the source of the scales.

3. Methods-Type Obstacles

3.1. Obstacle 4: Common Method Bias

Common method bias (CMB) can be a major threat to internal validity in social science research (Sharma, Yetton, & Crawford, 2009). In simple terms, CMB means that support for a hypothesized relationship is due to the methods used rather than an actual state of affairs. Though there are other method biases that adversely impact a study’s outcome, we focus on common method bias due to the frequency with which it is not given due consideration in the many submissions to *JAIS* that report results of cross-sectional surveys.

Scholars continue to warn of the degree, severity, and frequency of errors created by common method bias. The impact is at a measurement level, and leads to an increased correlation between variables and, in turn, to faulty conclusions about relationships. This can be quite severe, as shown in a meta-analysis of leadership studies that found a 23.9 percent inflation of correlations between independent and dependent variables in data collected from different sources as compared to data collected from a single source (Podsakoff, MacKenzie, & Podsakoff, 2012).

Authors should pay attention to the design and analysis of cross-sectional surveys. They should employ practices that minimize potential common method bias and statistically control for its effects. Failure to do so may call the results into question. Though not likely to result in a desk rejection, an editor or reviewer will likely require the authors to address concerns of bias, which can delay a merit-based decision.

Authors can mitigate common method bias in a variety of ways at multiple stages in the research process by avoiding its likely causes and examining the data set (Burton-Jones, 2009). During the design stages of the study, we suggest that authors:

1) Identify separate key informants for the dependent and independent variables. Having a common informant for both sets of variables is a major cause of common method bias.

2) Separate the measurement of the dependent and independent variables temporally. Allowing sufficient time between measurements can remove contextual cues.

3) Vary response formats and media across variables. This can provide the psychological break needed to prevent respondents from identifying links between variables.

4) Vary scale properties, guarantee anonymity, and eliminate question blocks (all items for one construct grouped together).

Sharma et al. (2009) provide examples of good practice in lowering the risk of CMB. They categorize TAM studies of the relationship between the constructs of perceived usefulness and usage in terms of their susceptibility to CMB. They rank studies that use scale items of the “system-captured” and “behavioral” types as less susceptible to CMB than items that are those “perceptually” anchored. System-captured data are from historical records or those recorded by a computer system, such as webpages accessed. Behavioral items ask questions about actions that the subjects have carried out, such as hours spent working on a system. Perceptually anchored items ask subjects to rate their professed value for items, such as regular usage of a system (Sharma et al., 2009).

Of course, these actions all require additional effort and expense in data collection. Further, constraints associated with unit of analysis (e.g., individual level), constructs intended for specific respondent classes (e.g., users), and environmental shifts (e.g., intervening implementation of agile methodologies) may not permit such flexibility.

Statistical approaches to mitigating concerns of common method bias should be applied to all cases where variables are measured by a common rater. Such approaches, however, do not detect common method bias with any consistency and only serve to lessen, not remove, concerns of CMB. Common approaches include:

1) Employing factor tests that rely on the property that method bias tends to inflate correlations. In such a situation, a common factor might encompass the entire set of items on a survey instrument or explain the vast majority of variance, and thus allow detection. Likewise, analysis can control for the effects of a measured or unmeasured latent “methods factor” (Padsokov, MacKenzie, Lee, & Podsakoff, 2003).

2) Including a marker variable in the study. A marker variable has no relationship to one or more variables in the study. Procedures exist to introduce a variable into the design of the study or identify one in a post hoc fashion. Correlations to the marker variable are used to determine and adjust for the impact of variance due to common

method bias. The adjusted correlations can be employed in causal analysis (Malhotra, Kim, & Patil, 2006). All current marker approaches, however, have limitations that must be recognized (Chin, Thatcher, & Wright, 2012).

However, the results of analytical tests have varied results and are far less convincing than a well-planned study. A combination of design and statistical analysis may be the more effective option. For example, Liang and Xue (2010) employ techniques of both design and post-collection analysis by varying the collection of data over time, then still follow with factor tests.

3.2. Obstacle 5: Formative Construct Unrecognized

A formatively measured construct is one whose observable indicators are assumed to “cause” a latent variable of interest. This reverses the assumed causality of reflectively measured constructs whose indicators are observable as a result of the latent variable of interest. Often, submissions to *JAIS* seem to consider that formatively measured constructs and reflectively measured constructs are freely exchangeable in models, with limitations decided more by available analytical software than by the properties, assumptions, and measurement issues of the formatively measured construct. The recent explosion of papers on this topic in the information system and management literatures indicates there is still a great deal of interest in modeling with formatively measured constructs, but much of the work highlights the many pitfalls that must be avoided (e.g., the entire issue of *SIGMIS Database*, Volume 44, Issue 4, November 2013).

The nature of the relationship between a set of formative indicators and a latent variable lead, by necessity, to implications that potentially create difficulties. To greatly oversimplify, think of a linear regression with multiple independent variables and a known dependent variable. Estimates of the relationships between the independent variables and the dependent are a simple calculation. However, how valuable is a regression with only minimal explained variance in terms of precisely predicting the dependent variable? Now consider a set of formative indicators and a latent variable. To have a good estimate of the latent variable, the formative indicators must explain much of the variance. If not, how good is the formative construct as a representation of the latent variable?

Avoiding problems of formative items requires that authors pay a great deal of attention to both the design and analysis stages of quantitative research (Kim, Shin, & Grover, 2010). As with any measurement model, errors made in design are difficult, if not impossible, to mitigate once data collection is complete. Thus, flaws of measurement design will often result in a rejected manuscript or the requirement that the measurement model be reconsidered followed by new data collection. Issues of analysis may often, but not always, be addressed through further investigation of the properties inherent in the data. The structural model can also suffer during analysis using formatively measured constructs as endogenous variables due to the inability to reach a determination of the latent variable in this case. The consolidation of multiple items may obscure interesting details. In all, the problems of formatively measured constructs can be difficult to overcome, and some researchers have called for them to be abandoned (Bergkvist & Rossiter, 2009; Edwards, 2011).

Still, there are significant benefits to employing well-designed formatively measured constructs (Diamantopoulos, Riefler, & Roth, 2008). This advantage is why there continues to be such heated discussion in the literature about their correct application. Formative indicators provide direction for practice in a way that reflective indicators cannot because the former likely specify manipulable attributes of an entity rather than perceptions of consequential attributes.

Given there are ample reasons for employing formatively-measured constructs, we suggest that authors take care to:

1) Follow good practice in designing any formatively measured construct (MacKenzie, Podsakov, & Podsakov, 2011). In particular, authors must be aware that the formative indicators need not be correlated, should exist in time prior to the latent

variable, and should fully explain variation in the latent variable. For these reasons, there should be a theoretical basis for defining the indicators.

2) Consider utilizing a common factors methodology during analysis (Treiblmaier, Bentler, & Mair, 2011).

3) Apply a sensitivity analysis to demonstrate stability of the independent variables (Kim et al., 2010). Formatively defined exogenous variables are subject to interpretational confounding where the independent variables are unstable and depend on one or more of the endogenous variables in the model.

4) Remember that placing a formative construct in an endogenous position implies that additional explanatory items for the latent variable exist, which violates the assumption that a formative construct is fully defined by its measurement items. When a formative construct is rooted in theory, then placing the construct endogenously implies a theoretical change that must be justified.

5) Address multi-collinearity among the indicators while being aware that the removal of an indicator has theoretical and measurement implications.

Advances in formative measurement will likely continue at a rapid pace in the near future. Staying abreast of these advances will be crucial to convincing an editor that any applications of formatively measured constructs is well considered, appropriate, properly validated, and controlled.

3.3. Obstacle 6: Self-Report Data Validity Issues

Self-report studies collect data directly from respondents without researcher interference or interaction. Self-report data typically result from a survey that measures feelings, personal traits, attitudes, perceptions, or beliefs. A large percentage of submissions to *JAIS* report studies containing self-report data. In many cases, this is entirely appropriate (Burton-Jones, 2009). For example, with a relatively objective piece of data such as an individual's age, a respondent may respond accurately, especially if they were assured confidentiality. It may be difficult or impossible to measure the internal feelings or perceptions of an individual without self-report. For example, it is hard to imagine determining career satisfaction by other means. However, such techniques are frequently abused. Schminke (2004, p. 312) mentions one problematic situation:

> *Your data on both the independent and dependent variable sides of the equation are exclusively self-reported, entirely perceptual, collected by a single method, and provided by a single informant.*

Not only does this raise the issue of common method bias described earlier, but, for each variable, respondents may give less-than-accurate representations of the external or internal world. Biased responses will result from a lack of knowledge about the variable of interest and natural impulses to shade the world in a more favorable light.

Impulses that influence self-report bias are most often considered in the light of socially desirable responding (Donaldson & Grant-Vallone, 2002). When the true state of affairs is socially undesirable, respondents are inclined to bias the response. For example, when someone is stuck in a dysfunctional organization and is unable to seek other employment, they may evaluate the true state more favorably. When a construct includes sensitive items, respondents are motivated to bias their response (e.g., when students are asked about time spent doing homework). In general, some individuals have a greater propensity to bias responses than others. Situational pressures might lead respondents to bias, especially if there is no trust in the guarantee of privacy. All of these factors require that the researcher consider the most accurate data sources when planning a study.

As such, for studies with self-reported data, we suggest that authors:

1) Replace some of the variables with independent data. Financial information serves as firm performance data and is often readily available. Neuropsychological data is gaining traction for complex emotional variables. Be aware, however, that some sources are not well established or cleaned of data imperfections and can be fraught with measurement errors. Trade-offs should be carefully considered and choices justified.

2) Employ other collection methods that include observation, archives, implicit measures, and behavioral traces. Such surrogates must be validated against the desired psychological constructs being replaced; however, such validation may be well documented in prior studies.

3) Select constructs that have been shown to be minimally biased in past studies. Add a variable of socially desirable responding to control concerns of self-report bias (Paulhus, 1998).

4) Recognize possible bias and discuss how it may influence the interpretation and generalization of results. Bias may be due to situational aspects, sensitivity of the information, and the disposition of the subject(s) (Schmidt, 1994).

An example of a study that obtains independent, computer recorded data is Horton, Buck, Waterson, and Clegg (2001). These authors employ system-measured usage rather than self-reported usage. Note that Sharma et al. (2009) categorize this measure as "very low" in terms of common method bias, in addition to it being an alternative to self-report problems (such as imperfect memory).

3.4. Obstacle 7: Sample Selection Limitation

The sample selected for a study is the main determinant of external validity. When we formulate a research question, there is typically a population we wish to understand. To understand the population of interest, we draw a sample we believe representative in order to make generalizations. Should the sample not be representative, then we do not have the external validity required to generalize. Sampling is fraught with potential error and can lead to serious concerns in a decision to accept or reject a manuscript. For example, if we wish to consider a question of intention to adopt a new technology and the sample consists of individuals who are already adopters and those who are yet to adopt, then we have a contaminated sample that prohibits generalization.

In designing a study, authors need to consider several sampling steps. First and foremost is a clear indication of the population to which the research question applies. Just which population draws our interest? If we want to address questions about user participation in the development of software, then a sample of undergraduate students hearing about the importance of user participation via their first class in analysis and design is not appropriate. The sample we use to generalize about the population at large must be representative, else there is no ability to form the generalization (see Seddon & Scheepers, 2012). Further sampling complications can occur because a sample may be limited to place and/or time. Place may matter in terms of differences in culture, organizational processes, or industrial conditions that might impact the generalization. Time plays a factor, too. Certainly, responses to questions of software maintenance issues would differ in today's climate compared with the days of Y2K.

Thus, we see that authors must identify their research questions' population of interest. Authors should then select a target sample to reflect the population of interest. In itself, this is a varied task. Sometimes lists of professional societies or third party databases provide the ideal sample frame; other times, it can be very difficult to identify a list of candidates to complete the survey. Sampling becomes even more difficult when one is trying to identify sets of respondents to avoid problems of common method bias. To add to confidence in external validity, we suggest that authors:

1) Apply random sampling whenever possible.

2) Do a comparative analysis of the final sample against the population. Are parameters that may make a difference in judgments the same in your sample and the population at large?

3) Should random sampling prove difficult, attempt to attain sufficient diversity and key expertise required to provide responsible responses to the questions through a purposive sample (e.g. an appropriate professional society). A snowball technique will attain qualified experts, though authors should recognize potential limitations due to a limited network.

4) Seek a sufficient sample size to guarantee ample power of the analysis.

5) Apply techniques of limited response rates to add confidence to the sample (Sivo, Saunders, Qing, & Jiang, 2006).

6) Split the sample should populations be mixed to show that the results hold more generally. For example, do the thoughts of developers and users hold equally in behavioral models?

7) Use controls for time, place, and personal characteristics to add confidence in generalization. Note, however, that these controls can also inflate measures of fit. It is advisable to dismiss the influence of demographic traits from analysis early if possible, then remove them from further analysis.

8) Give details of when and where your sample was gathered so readers can make inferences themselves about generality. Strangely, these details are often lacking in published papers. Are authors concerned about revealing the age of their data or that it was gathered in a specific region?

We believe editors and reviewers should insist on these vital details (as does the American Psychological Association; see APA, 2008). Many times, our data will be gathered some years before publication given the length of our review processes, but we have to live with this situation and we should be open about it.

A sound example of generalizing from a limited sample is Markus, Steinfield, and Wigand's (2006) study. The authors generalize a case based on specific findings to other cases where the same conditions are likely to occur. Generalization must be argued based on context, methodology, and sample. Effective guidelines for a wide variety of research settings and samples may be found in Seddon & Scheepers (2012).

3.5. Obstacle 8: Inappropriate Analytic Techniques

Perhaps the least detrimental obstacle is incorrectly selecting, applying, or simply altogether omitting an appropriate analytic technique. Methods selection is often a curious process because many authors choose a method based on the success of data collection instead of the study's objective(s). Admittedly, this may at times be unavoidable due to difficult data gathering or unexpected deviations from the underlying assumptions of preferred techniques. Errors in application seem to be less common, perhaps due to increased attention by reviewers and vast improvements in the statistical software available. Omissions tend to comprise a lack of reporting. Authors often fail to justify why they chose a particular technique, discuss the processes employed in the analysis, or include critical information that characterize the strength of the results used to draw conclusions, such as simple correlations. A number of submissions avoid evident issues of common method bias, proper examination of mediation, validity of formative constructs, and consideration of type II errors. Complete information is essential to reviewers and future researchers.

There has recently been ongoing discussion about the appropriate use of the partial least squares (PLS) as the analytical tool of choice for structural equation modelling (SEM). PLS has gained widespread acceptance in the IS literature, likely due to the increase in software capability and extension to a variety of research models (Esposito Vinzi, Chin, Henseler, & Wang, 2010). Compared to covariance-based SEM techniques, researchers claim that PLS is more robust in the face of smaller sample sizes and deviations from normality, and adapts more readily to formatively measured constructs (Ringle, Sarstedt, & Straub, 2012). However, these claims continue to be questioned (Treiblmaier et al., 2011). What may be more problematic is that the choice should be based on one's research aims: PLS may be more appropriate for research goals that are exploratory or predictive, while covariance-based approaches are likely better for theory testing, confirmation, and comparison (Hair, Ringle, & Sarstedt, 2011).

Since so many techniques, tools, and statistics exist, providing a global set of guidelines for their choice is nigh impossible. Further, the research community seems to only stochastically accept analytic techniques, with advances in modeling altering the landscape almost continually. As such, we suggest that authors:

1) Justify your selections: explain the reasons for choosing any particular technique. Be certain to include the research question, the characteristics of the measurement model, and the peculiarities in the data.

2) Apply the selections credibly: follow acceptable practices according to current thought while recognizing limitations. Make certain that common problems are avoided, such as the incorrect testing of moderation effects (Carte & Russell, 2003) and the application of difference scores (Klein, Jiang, & Cheney, 2009). Do not use the chosen technique to cover data deficiencies. If sample size is cited as a reason for choosing a method, provide a credible argument as to why no further data can be readily collected or the response rate increased.

3) Avoid omissions through greater inclusiveness: provide sufficient information necessary for a thorough evaluation without being excessive. Focus on novelty to determine what information must be in the paper. For example, far more extensive information about instrument validation is required when new measures are employed. Consider how to strengthen an argument. For example, holdout sample analysis can fortify predictive and theory building studies.

4) If not certain about what to include, err on the side of too much reporting in the paper. However, make certain that the additional material does not detract from the flow of the paper. Use appendices for reporting analytical details that are peripheral to the study or about which you are uncertain. Referees and editors are more than willing to indicate what should be incorporated or eliminated, but require details essential for evaluation.

5) Follow good practice in reporting the results of your study (e.g., APA, 2008). A point of concern is that, with SEM techniques, authors often do not give basic descriptive statistics (e.g., means and correlations) even though recognized authorities advocate this practice (e.g., Kline, 2005, p. 480). Pee, Kankanhalli, and Kim (2010) illustrate a good example of reporting descriptive statistics with SEM. Further, authors should also be aware that *JAIS* has its own data policies[3], as do many academic journals.

Remaining current in analytic techniques is at times problematic. Acceptable practices are frequently discussed in major academic journals, but they often fail to bring newer approaches to light or debate

[3] See http://aisel.aisnet.org/jais/policies.html#data

the efficacy of popular techniques. Books on multivariate approaches can provide solid foundations, but can lag acceptable practice. Better approaches may even gain traction while a paper is under review. However, *JAIS*, for example, seldom rejects a paper on the basis of analytic techniques alone. Analysis can be readily redone if the editors deem it appropriate. Initial concerns regarding the analytic techniques can be overlooked if the research questions are sound and interesting and the data collection produces a solid basis for analysis.

4. Conclusions

This editorial note represents personal distillations of the experiences of the first author as editor-in-chief of *JAIS* and the second author as a senior editor. As these views are personal they are no doubt debatable and debate is encouraged. With statistical approaches and methods, we know that debate is vigorous and ongoing, and we certainly do not present a final word here.

References

APA. (2008). *Reporting standards for research in psychology*. Retrieved November 12, 2013, from www.apa.org/pubs/authors/jars.pdf

Barley, S. (2006). When I write my masterpiece: Thoughts on what makes a paper interesting, *Academy of Management Journal, 49*(1), 16-20.

Bergkvist, L., & Rossiter, J. R. (2009). Tailor-made single-item measures of doubly concrete constructs. *International Journal of Advertising*, *28*(4), 607-621.

Boudreau, M., Gefen, D., & Straub, D. (2001). Validation in information systems research: A state-of-the art assessment, *MIS Quarterly, 25*(1), 1-16.

Burton, L., Westen, D., & Kowalski, R. (2009). *Psychology* (2nd ed.). Milton, Qld: John Wiley & Sons.

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

Carte, T., & Russell, C. (*2003*). In pursuit of moderation: Nine common errors and their solutions. *MIS Quarterly*, *27*(3), 479-501.

Chin, W., Thatcher, J., & Wright, R. (2012). Assessing common method bias: Problems with the ULMC technique, *MIS Quarterly,* 36(3), 1003-1019.

Colquitt, J., & Zapata-Phelan, C. (2007). Trends in theory building and theory testing: A five-decade study of the *Academy of Management Journal*. *Academy of Management Journal*, *50*(6), 1281-1303.

Cook, T., & Campbell, D. (1979). *Qasi-experimentaion design and analysis issues for field settings*. Boston: Houghton Mifflin.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203-1218.

Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, *17*(2), 245-260.

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, *14*(2), 370-388.

Feldman, D. (2004a). The devil is in the details: Converting good research into publishable papers (editorial). *Journal of Management*, *30*(1), 1-6.

Feldman, D. (2004b). What are we talking about when we talk about theory? *Journal of Management*, *30*(5), 565-567.

Geyman, J. (2000). Evidence-based medicine in primary care: An overview. In J. Geyman, R. Deyo, & S. Ramsey (Eds.), *Evidence-based clinical practice: Concepts and approaches* (pp. 1-12). Woburn, MA: Butterworth Heinemann.

Gregor, S. (2006). The nature of theory in IS. *MIS Quarterly*, *30*(3), 611-642.

Gregor, S., Lin, A., Gedeon, T., Riaz, A., & Zhu, D. (2014). Neuroscience and a nomological network for understanding emotions in information systems research. *Journal of Management Information Systems*, *30*(4), 13-47.

Hair, J., Ringle, C., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory & Practice*, *19*(2), 139-152,

Hempel, C. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice-Hall.

Horton, R., Buck, T., Waterson, P., & Clegg, C. (2001). Explaining intranet use with the technology acceptance model. *Journal of Information Technology, 16*(4), 237-249.

Hovorka, D., Larsen, K., & Morachi, D. (2009). *Conceptual convergences: Positioning information systems among the business disciplines.* Paper Presented at the European Conference on Information Systems, Verona.

Kim, G., Shin, B., & Grover, V. (2010). Investigating two contradictory views of formative measurement in information systems research. *MIS Quarterly*, *34*(2), 345-364.

Kim, G., Shin, B., Kim, K. K., & Lee, H. (2011). IT capabilities, process-oriented dynamic capabilities, and firm financial performance. *Journal of the Association for Information Systems*, *12*(7), 487-517.

Klein, G., Jiang, J., & Cheney, P. (2009). Resolving difference score issues in information systems research. *MIS Quarterly*, *33*(4), 811-826.

Kline, R. (2005). *Principles and practice of structural equation modelling* (2nd ed.). New York, NY: Guildford Press.

Li, X., Troutt, M., Brandyberry, A., & Wang, T. (2011). Decision factors for the adoption and continued use of online direct sales channels among SMEs. *Journal of the Association for Information Systems*, *12*(1), 1-31.

Liang, H., & Xue, Y. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. *Journal of the Association for Information Systems*, *11*(7), 394-413.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293-334.

Malhotra, N. K., Kim, S. S., & Patil, A. (2006). Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science*, *52*(12), 1865-1883.

Markus, M. L., Steinfield, C. W., & Wigand, R. T. (2006). Industry-wide information systems standardization as collective action: The case of the US residential mortgage industry. *MIS Quarterly*, *30*(2), 439-465.

Neuman, W. (2006). *Social research methods*. Boston, MA: Pearson.

Paul, R. (2005). Editor's view: An opportunity for editors of IS journals to relate their experiences and offer advice. The editorial view of Ray J Paul. First in a series. *European Journal of Information Systems*, *14*(3), 207-212.

Paulhus, D. L. (1998). *Manual for the balanced inventory of desirable responding.* Toronto, Canada: Multi-Health Systems.

Pavlou, P. A., & Fygenson, M. (2006). Understanding and predicting electronic commerce adoption: an extension of the theory of planned behavior. *MIS Quarterly, 30*(1), 115-143.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.

Pee, L., Kankanhalli, A., & Kim, H.-W. (2010). Knowledge sharing in information systems development: A social interdependence perspective. *Journal of the Association for Information Systems*, *11*(10), 550-575.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*(1), 539-569.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903.

Riedl, R., Davis, F., & Hevner, A. (2014). Towards a neuroIS research methodology: Intensifying the discussion on methods, tools, and measurement (special issue editorial). *Journal of the Association for Information Systems*, 15(10), pp. i-xxxv.

Ringle, C., Sarstedt, M., & Straub, D. (2012). Editor's comments: A critical look at the use of PLS-SEM in *MIS Quarterly*. *MIS Quarterly*, *36*(1), iii-xiv.

Schmidt, N. (1994). Method bias: The importance of theory and measurement. *Journal of Organizational Behavior*, *15*(5), 393-398.

Schminke, M. (2004). From the editors raising the bamboo curtain. *Academy of Management Journal*, *47*(3), 310-314.

Seddon, P., & Scheepers, R. (2012). Towards the improved treatment of generalization of knowledge claims in IS research: Drawing general conclusions from samples. *European Journal of Information Systems*, *21*(1), 6-21.

Sharma, R., Yetton, P., & Crawford, J. (2009). Estimating the effect of common method variance: The method-method pair technique with an illustration from TAM technique. *MIS Quarterly*, *33*(3), 473-490.

Sivo, S. A., Saunders, C., Qing, C., & Jiang, J. J. (2006). How low should you go? Low response rates and the validity of inference in IS questionnaire research. *Journal of The Association For Information Systems*, *7*(6), 351-413.

Straub, D. (2009). Why top journals accept your paper (editorial). *MIS Quarterly*, 33(3), iii-x.

Treiblmaier, H., Bentler, P., & Mair, P. (2011). Formative constructs implemented via common factors. *Structural Equation Modeling*, *18*(1), 1-17,

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (Eds.) (2010). *Handbook of partial least squares: Concepts, methods, and applications.* Berlin: Springer.

Weber, R. (2012). Evaluating and developing theories in the information systems discipline. *Journal of the Association for Information Systems*, *13*(1), 1-30.

Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii-xxiii.

Weick, K. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, *40*(3), 385-390.

Whetten, D. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490-495.

Xu, J., Benbasat, I., & Cenfetelli, R. (2011). The effects of service and consumer product knowledge on online customer loyalty. *Journal of the Association for Information Systems, 12*(11), 741-766.

About the Authors

Shirley GREGOR is Professor of Information Systems at the Australian National University, Canberra, where she is Director of the National Centre for Information Systems Research and Associate Dean Research in the College of Business and Economics. Her current research interests include the innovative and strategic use of information and communications technologies, knowledge systems, and the philosophy of technology. She has published in journals such as *MIS Quarterly, Journal of the Association of Information Systems*, *International Journal of Electronic Commerce*, *International Journal of Human Computer Studies*, *European Journal of Information Systems,* and *Information Technology & People*. Professor Gregor was made an Officer of the Order of Australia in the Queen's Birthday Honour's list in June 2005 for services as an educator and researcher in the field of information systems and for work in e-commerce in the agribusiness sector. She is a Fellow of the Australian Computer Society and a Fellow of the Association for Information Systems. She was a Senior Editor of *MIS Quarterly* 2008-2010 and *Editor-in-Chief* of the Journal of the Association of Information Systems 2010-2013.

Gary KLEIN is the Couger Professor of Information Systems at the University of Colorado, Colorado Springs. His research interests include project management, technology transfer, and mathematical modeling with over 200 academic publications in these areas. He served as Director of Education for the American Society for the Advancement of Project Management, is an active member of the Project Management Institute and the International Project Management Association, and is a Fellow of the Decision Sciences Institute. He serves on the editorial board of the *International Journal of Information Technology Project Management*, as a departmental editor for the *Project Management Journal*, as an SE for the *Journal of the Association of Information Systems* and the *Pacific Asia Journal of the Association of Information Systems*.