3-1-2004

Minimizing Method Variance in Measures of System Usage

Andrew Burton-Jones
abjones@cis.gsu.edu

Detmar W. Straub Jr.

MINIMIZING METHOD VARIANCE IN MEASURES OF SYSTEM USAGE

Andrew Burton-Jones
Department of Computer Information Systems
J. Mack Robinson College of Business
Georgia State University
abjones@cis.gsu.edu

Detmar W. Straub Jr.
Department of Computer Information Systems
J. Mack Robinson College of Business
Georgia State University
dstraub@gsu.edu

Abstract

Although the system usage construct has a central place in information systems (IS) research, there has been little discussion to date about how it should be measured. Usage is most commonly measured by self-reported questionnaire data, and, on occasion, interview reports, computer logs, or independent observation. Some researchers have found that the relationship between usage and other constructs differs depending on its method of measurement. Consequently, there is a need to determine how to obtain 'true' measures. The objective of this paper is to present a framework for measuring systems use. Specifically, the paper distinguishes two components of systematic variance—common methods bias and distance bias. These can lead to inaccurate measures of usage and inaccurate measures of its relationships with other constructs. The paper then presents an approach to minimizing these two sources of systematic bias and operationalizes this approach in the context of studying the relationship between usage and performance.

Keywords: measurement theory, validity, bias, system usage, objectivity, self-reports

Introduction

Systems usage is among one of the most widely studied phenomena in information systems (IS) research. Despite the many studies of system usage, there have been repeated calls over the years to study it more closely. Early on Ginzberg (1978) called for researchers to distinguish between 'how' and 'how much' an IS was used. Later, Trice and Treacy (1986) called for a focused reconceptualization of usage. In 1993, Zigurs renewed the call when she says: "system usage is an example of a deceptively simple construct that needs to be looked at more carefully" (p. 117). Still more recently, DeLone and McLean (2003, p. 16) examined the usage literature and concluded that: "the problem to date has been a too simplistic definition of this complex variable." To contribute to a deeper understanding of system usage, this study proposes a measurement theoretic framework for studying usage and its relationship with other constructs.

A Framework for Measuring System Usage

To develop a framework for measuring system usage, we first outline some basic measurement theory principles that apply to measuring constructs in general. We then describe the theoretical nature of systems usage. Finally, we apply the measurement theory principles to system usage.

The framework we offer in this paper rests on two assumptions: 1. accurate measures are possible, and 2. accurate measures are practical. Although both assumptions are reasonable, they require clarification. The first assumption speaks to our position on the nature of social reality. In this paper, we follow Cook and Campbell (1979) by subscribing to a critical realist position in which one believes that an external reality exists, but acknowledges that all constructs are fictions and all measures of constructs are made with error. Thus, we assume that 'true' measures only measure constructs, not reality, and we never actually achieve 'true' measures, we only approximate them. The second assumption speaks to the issue of pragmatism. Researchers must make tradeoffs during every stage of a research project. Thus, our framework does not recommend a single, 'true' measurement approach. Instead, it provides an intellectual framework for understanding the implications of different measurement approaches and making an informed tradeoff for the research question at hand.

Measurement Theory

In classical measurement theory, any measure is assumed to be composed of a true score and random error. Across a set of scores, this is represented in the true-score equation in which a measure's variance is composed of variability due to true score and variability due to error:

$$var(x) = var(T) + var(e_X) \quad (1)$$

True score theory is the foundation for reliability theory and underlies tests such as Cronbach's alpha. However, it assumes that error is random, and thus does not account for *systematic* error such as method effects (Fiske, 1982). Thus, a more complete conception of measurement distinguishes true variance from systematic variance and random error variance:

$$var(x) = var(T) + var(eSys) + var(eRandom) \quad (2)$$

Equations 1 and 2 emphasize the tenuous nature of measurement. Many researchers justify the *reliability* of their measures using tests such as Cronbach's alpha and the *validity* of their measures using tests of convergent and discriminant validity. Equations 1 and 2 remind us that these tests are insufficient for measuring a construct's true score. For instance, reliability theory assumes that method variance is true variance. Thus, method variance tends to inflate reliability scores such as Cronbach's alpha (Tepper and Tepper, 1993). Equation 2 allows one to separate method bias from random error, but a researcher must use multiple methods to separate the effects (Campbell and Fiske, 1959). If all measures are obtained by a single method, high convergent and discriminant validity do *not* indicate that one's measures are close to their true scores (Podsakoff and Organ, 2003).

Is method variance serious? To address this question, Cote and Buckley (1987) studied every social science study that had used Campbell and Fiske's (1959) multiple traits/multiple methods (MTMM) design. Using confirmatory factor analysis, they then studied the extent of true score variance, method variance, and error variance in past research. They found that on average across all the studies, measures contained 42% true variance, 26% method variance, and 32% random error variance. Using these numbers, Podsakoff et al. (2003) showed that where two constructs have a true correlation of 1.00, researchers could conclude that the correlation was only 0.52, and where two constructs have a true correlation of 0.00, researchers could conclude that the correlation was 0.23. These numbers can also vary across contexts. Spector (1992) estimated that many self-reported measures in organizational behavior capture only 10-20% of a construct's true variance.

Despite method bias reducing researchers' ability to truly measure a construct, few researchers control for its effects. In IS research, Woszczynski and Whitman (2004) analyzed 116 empirical studies in the field's top journals and found that 58% collected all of their data via just one instrument and only 10 studies (8.6%) explicitly mentioned potential method variance.

In measurement theory, method variance is referred to as systematic bias. However, it is useful to distinguish between two reasons why a rater may provide biased measures: lack of *ability* and lack of *intent*. In this paper, we use 'distance bias' to refer to the former and we reserve 'common method bias' for the latter (Figure 1).

'Distance' refers to the rater's mental distance from the construct in construct space. If a rater does not have access to the focal construct (e.g., A in Figure 1), their measurement of it may be restricted to a *proxy* (e.g., B in Figure 1). One example of distance bias occurs when a rater scores an attribute that belongs to something other than themselves, e.g., an information system, an organization, or another person. In this case, the individual may lack full knowledge of the construct being rated. Another example occurs when an individual scores a personal attribute of which they have little recollection, e.g., past behavior (Blair and Burton, 1987).

In contrast to distance bias, 'common method bias' refers to contexts where a rater *could* give a construct's true score, but because of a common method, s/he gives a different rating (e.g., 'C' in Figure 1). Table 1 shows many well-known sources of common method bias.

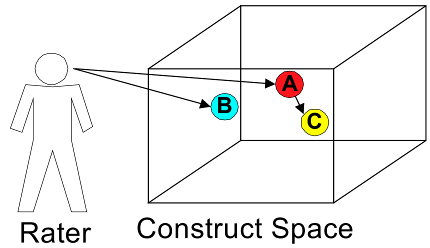


Figure 1: Two Sources of Method Variance: Distance and Bias

Table 1: Sources of Common Methods Bias (Podsakoff et al., 2003)

Common rater effects	**Item characteristic effects**	**Item context effects**
Consistency motif	Item social desirability	Item priming effects
Implicit theories	Item demand characteristics	Item embeddedness
Social desirability	Common scale formats	Context-induced mood
Acquiescence bias (yea-saying, nay-saying)	Common scale anchors	Scale length
Mood state	Positive and negative item wording	Grouping items

Although distance bias and common methods bias are two distinct sources of systematic bias, past studies have tended to emphasize only one or the other. For example, in studies of systems usage, Straub et al. (1995) found that system usage had different relationships with other constructs depending on the method used to measure it. Since then, several studies have used different methods to measure antecedents to use (e.g., questionnaires) and use (e.g., computer logs) (e.g., Venkatesh et al., 2003). This procedure tackles common method bias, but it does not address distance bias (e.g., whether computer logs truly capture system usage).

Therefore, in this paper we emphasize the distinction between common methods (CM) bias and distance (Dist) bias as two separate elements of systematic variance, per equations 3 and 4:

$$eSys = eCMBias + eDistBias \quad (3)$$

$$var(x) = var(T) + var(eCMBias) + var(eDistBias) + var(eRandom) \quad (4)$$

As equation 4 suggests, researchers must minimize both common method bias and distance bias to obtain (or at least approximate) true measures of their constructs.

Defining System Usage

To measure systems usage, we must first define it. In this research, we build on Goodhue (1995) and define IS usage as the interaction between a user, an IS, and a task. 'Users' can exist at any level of analysis (e.g., individual, group, organization; Lamb and Kling, 2003), but we limit our analysis here to the individual level. At an individual level, constructs describing human action typically consider a person's cognition, affect, and/or behavior (Kozlowski and Klein, 2000). Therefore, we define individual IS usage as an *individual's cognition, affect, and/or behavior when using an IS in a task*. Cognition here refers to the users' cognitive state (e.g., level of attention); affect refers to the users' affective state (e.g., level of anxiety); and behavior refers to the users' observable actions (e.g., the IS features that they use).

Measuring System Usage

Given the multidimensional nature of usage, measuring its true score requires one to measure each dimension of usage with minimum common method bias and distance bias. Figure 3 presents a framework for measuring usage in this way. The aim is to use multiple methods both within usage and between usage and its related constructs. This is not to say that each dimension of use must always be measured by different methods. On the contrary, a measurement method may participate in the meaning of the construct to such an extent that any other method may fail to tap into the construct. For example, to some executives, independently rated measures may be more relevant than users' self-reports, but for others, the reverse might be true. Consequently, we use the term 'validation' in Figure 3, below, to indicate that the approach provides the necessary data to allow a validation. Following this reasoning, the first level of validation ensures that one chooses methods that reduce the *distance* between the rater and each dimension of use. The second level of validation ensures that one reduces the effect of *common methods* bias on the measured relationship between constructs. By using both levels of validation, one can obtain truer measures of usage and its relationships with other constructs. The aim is to use multiple methods to reduce both sources of bias to a minimum, and then use triangulation, where necessary, to confirm the overall pattern of results.

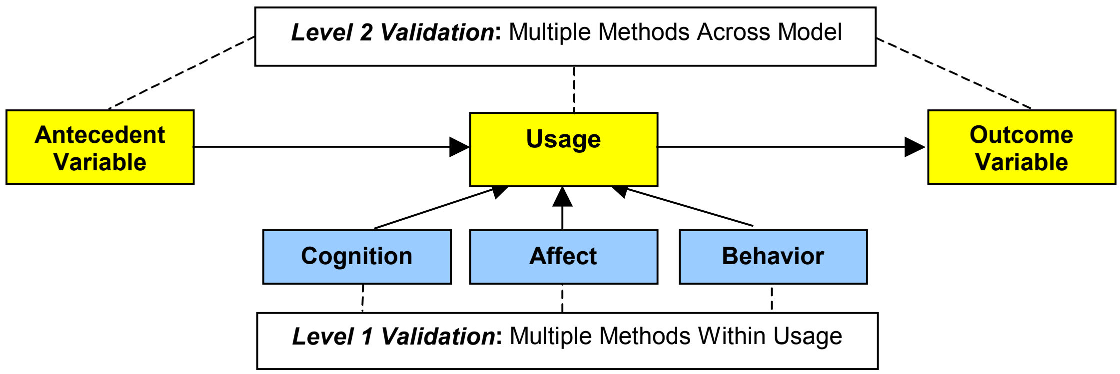


Figure 3: A Framework for Measuring System Usage

To operationalize the framework in Figure 3, one must place system usage in a nomological net. To demonstrate how this would work, Figure 4 operationalizes the framework in the context of the use→performance relationship. This is a useful context for studying the framework given recent calls for such research (Venkatesh, et al. 2003). Because our aim is to provide an example of how one would operationalize the framework, rather than provide a full test, we have not included antecedents of use in Figure 4, nor the affective dimension of use. For the purposes of illustration, we instead concentrate on a parsimonious model of cognitive and behavioral use and performance.

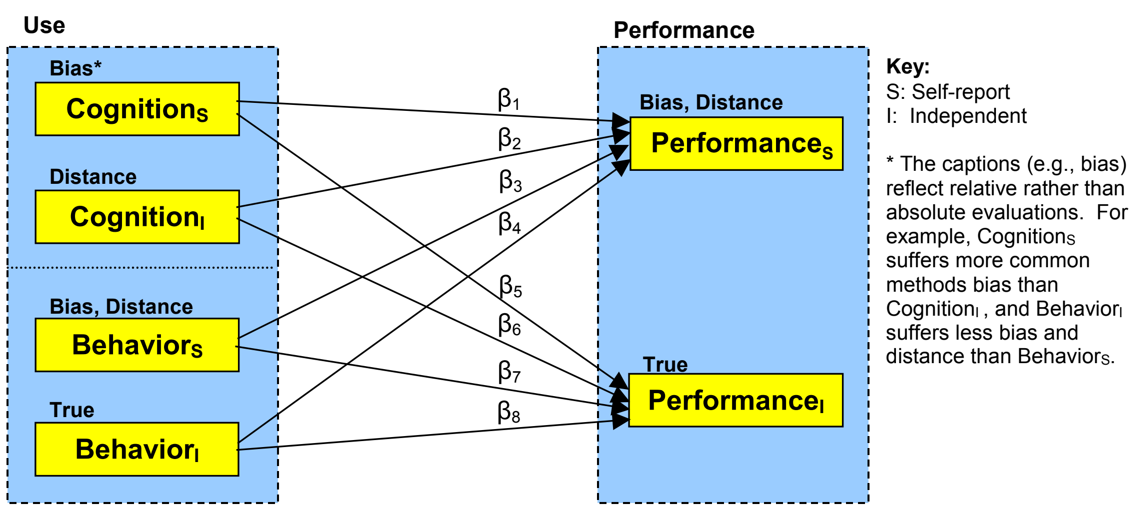


Figure 4: Measuring System Use and Performance

Figure 4 describes how one could use two methods — self reports and independent observation — to test the use→performance relationship. Following an MTMM design, this strategy employs self-reports and independent observation to rate each construct. These methods are chosen because they differ in their degree of distance from each construct and their likelihood of bias.

In terms of *distance*, when one measures cognition, self-reported measures suffer less distance than independent observations. If the cognitions being reported are salient in a person's mind, s/he has access to the 'true' level of the construct (i.e., low distance). However, an independent observer does not have access to the cognitions of others, so his/her measures will be subject to high distance bias (i.e., s/he will be restricted to measuring a proxy). Conversely, self-reported measures of behavior are more subject to distance bias than independent observation (Figure 4). This is for two reasons.

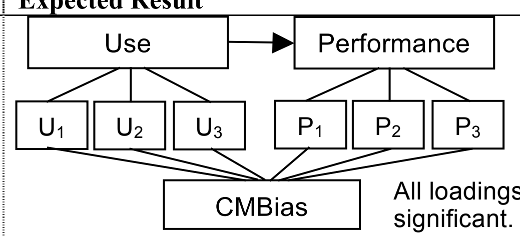
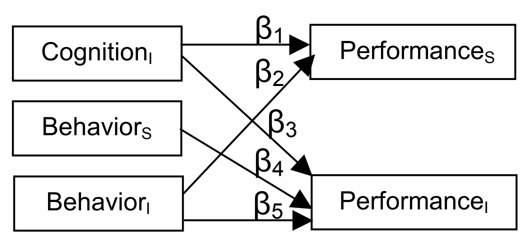
Rating one's own behavior requires that one (a) recall the behavior being rated, and (b) rate the behavior once recalled. The first weakness with self-reported ratings of behavior is that individuals have difficulty recalling past behavior (Blair and Burton, 1987). Independent ratings, however, can accurately record the true level of behavior (e.g., a computer log can record *exactly* how many times a user invoked an IT feature). The second weakness occurs with rating the behavior once recalled. If the rating requires an evaluation against some external standard (e.g., level of sophistication or appropriateness), individuals might not have a full understanding of the benchmark whereas an independent rater, if qualified, can rate the behavior more accurately.

Self-reported and independent measures also differ in their likelihood of *common methods bias*. In general, independent measures should be less influenced by common methods bias. Nevertheless, the degree to which independence reduces common methods bias depends on the type of rater. If the rater is a human, common method effects such as those in Table 1 can still occur, because they stem from general human tendencies. If the rater is an objective instrument (e.g., a video, computer program, etc), behavior can be recalled without these biases.

In summary, if a researcher wishes to measure system usage with a minimum of method variance, s/he should measure the *behavioral* dimension via independent observation. The implication for studying the relationship between use and performance is that performance is itself a behavior and should, therefore, be measured independently. Measures of *cognition*, on the other hand, suffer common methods bias when one uses self-report data and distance bias when one uses observation. Thus, for user cognitions, researchers are limited to using a measure of use in which method variance participates or using multiple methods to triangulate on the 'true' relationship between use and performance.

Of course, obtaining 'true' measures of use and performance does not mean that one will obtain high correlations between constructs. Distance bias could inflate or suppress the true relationship among constructs, depending on the nature of the proxy used to measure the construct. Similarly, common method bias could inflate or suppress relationships between constructs, depending on the social desirability of the relationship and the degree to which the relationship meets respondents' implicit theories. Therefore, in the context of measuring use→ performance, the effects of common method and distance biases depend on the particular measures (and proxies) of use rated by the respondent. Following this reasoning, Table 3 presents two predictions that can be used to provide an initial test of the framework in Figures 3-4.

Table 3: Predictions Derived from Figures 3-4

#	Prediction	Test	Expected Result
1	Common method bias will be significant when use and performance are measured using the same method	If one measures use and performance with the same method and includes a factor for common methods bias in the model, the common methods factor will have significant factor scores.	
2	Distance bias influences the strength of the relationship between use and performance	The strength of the relationship between use and performance will be significantly influenced by the degree of distance bias in the model, i.e., β (with $e_{DistBias}$) $\neq$ β (with $e_{No\ DistBias}$) Specific Hypotheses:	
		H1A: $Behavior_I$ will have a significantly different relationship with $Performance_I$ than will $Behavior_S$	$B_5 \neq B_4$, I.E., $B_5 - B_4 \neq 0$
		H1B: $Behavior_I$ will have a significantly different relationship with $Performance_I$ than with $Performance_S$.	$B_5 \neq B_2$, I.E., $B_5 - B_2 \neq 0$
		H1C: $Cognition_I$ will have a significantly different relationship with $Performance_I$ than with $Performance_S$.	$B_3 \neq B_1$, I.E., $B_3 - B_1 \neq 0$

Discussion

This study provided a measurement-theoretic framework to guide choices about how one should go about measuring systems usage in the context of a nomological network. It then demonstrated how one could do this in the context of studying the relationship between systems usage and performance. The contributions of the research are: 1. that it extends the field's understanding of method variance by decomposing it into two distinct elements, namely distance bias and common methods bias, and 2. that it provides a useful intellectual framework for researchers and practitioners who study systems usage and need to obtain accurate measures of system usage and its relationship with other constructs such as performance.

References

Blair, E., & Burton, S. (1987) Cognitive processes used by survey respondents to answer behavioral frequency questions, *Journal of Consumer Research,* 14, 280-288.

Cook, T.D. & and Campbell, D.T. (1979) Quasi-Experimentation: Design & Analysis Issues for Field Settings, Boston: Houghton Mifflin Company.

Campbell, D.T., & Fiske, D.W. (1959) Convergent and discriminant validity by the multitrait-multimethod matrix, *Psychological Bulletin,* 56, 81-105.

Cote, J.A., & Buckley, R. (1987) Estimating trait, method, and error variance: Generalizing across 70 construct validation studies, *Journal of Marketing Research,* 24, 315-318.

DeLone, W.H., & McLean, E.R. (2003) The DeLone and McLean model of information systems success: A ten-year review, *Journal of Management Information Systems,* 19(4), 9-30.

Fiske, D.W. (1982) Convergent -discriminant validation in measurements and research strategies, In D. Brinbirg & L. Kidder (Eds.), *Forms of Validity in Research* (pp. 77-92), California: Jossey Bass.

Ginzberg, M.J. (1978) Finding an adequate measure of OR/MS effectiveness, *Interfaces*, 8(4), 59-62.

Goodhue, D.L. (1995) Understanding user evaluations of information systems, *Management Science*, 41(12), 1827-1844.

Kozlowski, S.W.J., & Klein, K.J. (2000) A multilevel approach to theory and research in organizations, In K. J. Klein and S. W. J. Kozlowski (eds.), *Multilevel Theory, Research, and Methods in Organizations* (pp. 3-90), California: Jossey-Bass.

Lamb, R., & Kling, R. (2003) Reconceptualizing users as social actors in information systems research, *MIS Quarterly,* 27(2), 197-235.

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

Spector, P.E. (1992) A consideration of the validity and meaning of self-report measures of job conditions, In C. L. Cooper and I. T. Robertson (eds.), *International Review of Industrial and Organizational Psychology*, 7, (pp. 123-151). New York: Wiley.

Straub, D., Limayem, M., & Karahanna-Evaristo, E. (1995) Measuring system usage: Implications for IS theory testing, *Management Science,* 41, 1328-1342.

Tepper, B.J. & Tepper, K. (1993) The effects of method variance within measures, *The Journal of Psychology,* 127(3), 293-302.

Trice, A.W., & Treacy, M.E. (1986) Utilization as a dependent variable in MIS research, *Proceedings of the Seventh International Conference on Information Systems*, San Diego, CA, 1986, pp. 227-239.

Venkatesh, V., Morris, M.G., Davis, G.B., & Davis, F.D. (2003) User acceptance of information technology: Toward a unified view, *MIS Quarterly,* 27(3), 425-478.

Woszczynski, A.B., & Whitman, M.E. (2004) The problem of common method variance in IS research, In A. B. Woszczynski and M. E. Whitman (Eds.), *The Handbook of Information Systems Research* (pp. 66-77), Hershey, PA: Idea Group Inc.

Zigurs, I. (1993) Methodological and measurement issues in group support system research, In L. M. Jessup and J. S. Valacich (Eds.), *Group Support Systems: New Perspectives* (pp. 112-122), New York: Macmillan Publishing Company.