MIS Quarterly

RESEARCH ESSAY

MINIMIZING METHOD BIAS THROUGH PROGRAMMATIC RESEARCH[1]

By: **Andrew Burton-Jones**
Management Information Systems Division
Sauder School of Business
University of British Columbia
2053 Main Mall
Vancouver, BC V6T 1Z2
CANADA
andrew.burton-jones@sauder.ubc.ca

Appendix A

The Empirical Consequences of Method Bias

Method bias has conceptual and empirical consequences. Conceptually, it results in a lack of construct validity, or a failure to measure the intended trait. Although this is a serious problem, it will not always alter a study's results. For example, method bias will have no empirical effect if the measured construct is perfectly correlated with the theorized construct. Thus, it is important to know when and how method bias can have empirical consequences.

As Table A1 shows, several factors determine the empirical consequences of method bias. The primary factor is whether method bias is consistent or inconsistent. Classical test theory uses the terms *systematic* and *random* rather than *consistent* and *inconsistent*, but the latter are better terms because method bias results from the method used; it is not purely random. Method bias is *consistent* when it occurs the same way each time the trait is measured. For example, a researcher may wish to measure how much a person actually uses a new information technology. A consistent bias will occur if, when measuring the trait, the researcher measures a different construct each time, for example, asking each user to complete a self-report questionnaire, thus measuring perceived rather than actual usage. As Table A1 shows, the effect of consistent method bias depends on whether it is additive or correlational.

- Bias is *additive* when the extent of bias is constant across the sample. For example, every respondent's self-rating may be higher on a trait than that respondent's actual score.

- Bias is *correlational* when the extent of bias is correlated with another trait. For example, men may prefer to rate themselves higher on some variable than women do.

Method bias is *inconsistent* when it occurs to varying degrees each time the trait is measured, and can be assumed to average out to zero in a large number of measurements. For example, suppose that a researcher asks users to complete a self-report questionnaire about their use of an IT and that the response items on the questionnaire are placed too close together. The poorly constructed questionnaire might lead respon-

Table A1. Empirical Consequences of Method Bias*

		Nature of effect			
		Consistent		*Inconsistent*	
		Additive	*Correlational*	*Idiosyncratic*	*Generic*
Empirical consequences of method bias	*Consequence for measurement model*	No effect or decrease in reliability	Increase or decrease in reliability	No effect	Decrease in reliability
	Consequence for structural model	No effect or increase in Type 2 error	Increase in Type 1 or Type 2 error	No effect	Increase in Type 1 or Type 2 error

*Based on Viswanathan (2005)

dents to place their responses higher or lower on the scale than they would wish. However, these errors might average out to zero in a large sample of measurements. As Table A1 shows, the effect of an inconsistent method bias depends on whether the effects are idiosyncratic or generic.

- Effects are *idiosyncratic* when method bias occurs in a negligible proportion of tests. For example, an interviewer may read out a question incorrectly in a few interviews.

- Effects are *generic* when method bias occurs in a large proportion of tests. For example, a survey may contain ambiguous questions that every respondent finds difficult to answer.

These consistent and inconsistent effects can alter results in many ways (Viswanathan 2005); Table A1 gives only a brief summary. The key lesson is that in addition to resulting in a lack of construct validity, method bias can have numerous empirical consequences.

Table A1 lists the impact of method bias on measurement models and structural models because this is in accord with the emphasis given to these issues in empirical research. However, method bias can lead to incorrect classifications in qualitative research and to incorrect point estimates and effect sizes in quantitative research—critical concerns when making practical recommendations.

In Figure A1, the two sources of method bias—knowledge bias and rating bias (see "Sources of Method Bias" in the main paper)—are mapped to the consistent and inconsistent empirical consequences. As the figure shows, knowledge bias yields consistent effects, but rating bias can yield consistent and inconsistent effects. Researchers should be concerned with both sources of method bias, however, because both consistent and inconsistent errors can have major empirical consequences.

The observation about knowledge bias is based on the reasonable assumption that a person's knowledge will remain fairly stable during the measurement process. For example, a user may give a consistently different rating of IS quality relative to that user's manager because the user has greater first-hand knowledge of the system's daily operation. In some instances, a rater may not have knowledge of the trait before the measurement process begins; for example, a user might not have constructed an opinion about IS quality until asked about it. However, even in these cases, it is reasonable to assume that the rater's knowledge, once constructed, remains fairly constant during the measurement process, rather than changing over time with the changes averaging out to zero.

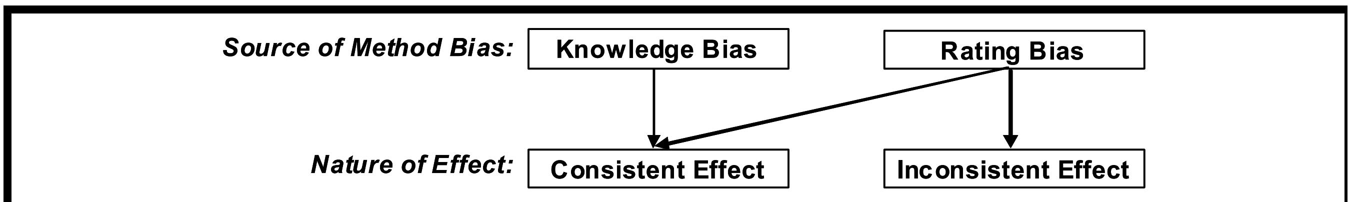

Figure A1. Mapping Sources of Method Bias to Their Effects

In contrast, rating bias can lead to consistent or inconsistent effects. For example, asking users a question that threatens their privacy will tend to lead to consistent rating errors, but having too few response options will tend to lead to inconsistent rating errors, since with insufficient options, respondents sometimes rate themselves higher and sometimes lower than their true rating.

Reference for Appendix A

Viswanathan, M. 2005. *Measurement Error and Research Design*, Thousand Oaks, CA: Sage Publications.

Appendix B

Meta-Theoretical Perspectives on Method Bias

As is true of any methodological concept, the meaning of method bias depends on certain meta-theoretical assumptions. Method bias is an issue in positivist research, with *positivist* being used broadly to refer to research interested in hypothesis testing, measurement, and validity (Hovorka et al. 2008, p. 36). This type of research has involved different assumptions at different points in time.

Operationalism

Operationalism was the initial meta-theoretical position. Operationalists (for example, logical positivists) rejected unobservable intellectual notions such as research constructs. To operate without constructs, operationalists assumed that operations defined real-world traits and that measures defined trait scores, so a person's measured IQ *is* that person's intelligence (Trout 2000). Thus, for the operationalist, method bias did not exist because data obtained through different methods was viewed as different traits. Although operationalism was initially influential, most scientists rejected it long ago, along with logical positivism (Grace 2001).

Realism

Realism was an early alternative to operationalism. Realists embraced unobservable notions and assumed that constructs could mirror real-world phenomena. To support this assumption, realists made strong assertions about the empirical realm, including that real-world phenomena, such as humans, have inherent traits, such as anxiety, with an inherent scale. The scale might be nominal, ordinal, or interval; and members of a population vary in their true scores on this scale (Essex and Smythe 1999). "True score" in this context is not a statistical notion. Rather, the true score of a trait refers to a phenomenon's real score on a particular property (Borsboom 2005).

The aim of realist research was to build constructs that correspond to real-world traits and to use methods to obtain the true scores of these traits (Borsboom 2005; Trout 2000). Thus, method bias, according to the realist, reflects an unwanted deviation from the trait score stemming from the use of an imperfect method.

Constructivism

Constructivists rejected the realists' assumption that the true nature of the world could be observed (Kuhn 1996). To operate without this assumption, constructivists made three assumptions.

- Constructs do not correspond to reality; they are merely useful fictions that researchers *attribute to* phenomena.
- Scales do not reflect the *true* nature of a trait; they merely enable researchers to obtain interesting or useful insights about phenomena.
- Scores are not "true;" they are just fictions that allow researchers to discuss the phenomena's relative standing on a socially constructed construct and scale (Messick 1989).

Although constructivism overcame some of realism's limitations, it also left the position of measurement and progress rather vacuous, since nothing really exists to measure or to progress toward (Borsboom 2005). Method bias is not a problem according to this view; it is just another instance of the social construction of measurement.

Critical Realism

Critical realism—also referred to as transcendental realism, constructive realism, and hypothetical realism—melded constructivism and realism in an attempt to overcome each one's limitations (Cook and Campbell 1979). Although critical realism has many aspects (Archer et al. 1998, Mingers 2004), this paper focuses on its measurement aspects. The greatest challenge for measurement is that critical realists embrace both ontological realism—the view that the real world exists—and epistemological fallibilism—the view that knowledge of the world will always remain partial and fallible. This dual view implies that critical realists are faced with somewhat of a paradox: Measurement is possible, but perfect, true measurement is not.

Different researchers respond in different ways to this paradox. Some lean toward the realist position, arguing that researchers must try to obtain true measures of real-world traits, because to do otherwise results in invalid measures and poor science (Borsboom 2005). Others lean more toward the constructivist position, arguing that it is not possible to obtain true measures of real-world traits, so why make that a goal? For example, "I reject the notion that I try to measure reality….I have no way of knowing reality, so how can I know whether my measure of reality, whatever reality might be, is valid? …The best I can do…is build constructs that I find useful in understanding the world" (Weber 2004, pp. vii-ix).

Middle-Ground and Evolutionary Perspectives

Still others are more comfortable with a middle-ground position, arguing that a single study cannot obtain true measures of real-world traits, but true measures might be possible over time. For example, "unforeseen technological or conceptual developments can occur that might make possible the measurement of variables that previously were treated as unmeasurable" (Bollen 2002, p. 608).

Likewise, Campbell advocated a Darwinian perspective, arguing that even though theories and measures are fallible, scientists can arrive at true measures and theories over time by continually testing them against the real world and rejecting those that fail (Campbell and Russo 2001). This perspective is evolutionary critical realism, and it is the perspective taken in the main paper: Researchers can measure real-world traits, but only if the scientific community engages in programmatic research over time.

References for Appendix B

Archer, M., Bhaskar, R., Collier, A., Lawson, T., and Norrie, A. (eds.). 2002. *Critical Realism: Essential Readings*, London: Routledge.

Bollen, K. A. 2002. "Latent Variables in Psychology and the Social Sciences," *Annual Review of Psychology* (53), pp. 605-634.

Borsboom, D. 2005. *Measuring the Mind,* Cambridge, UK: Cambridge University Press.

Campbell, D. T., and Russo, M. J. (eds.). 2001. *Social Measurement*, Thousand Oaks, CA: Sage Publications.

Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Setting*s, Boston: Houghton Mifflin Company.

Essex, C., and Smythe, W. E. 1999. "Between Numbers and Notions: A Critique of Psychological Measurement," *Theory & Psychology* (9:6), pp. 739-767.

Grace, R. C. 2001. "On the Failure of Operationism," *Theory & Psychology* (11:1), pp. 5-33.

Hovorka, D. S., Germonprez, M., and Larsen, K. R. 2008. "Explanation in Information Systems," *Information Systems Journal* (18:1), pp. 23-43.

Kuhn, T. S. 1996. *The Structure of Scientific Revolutions* (3rd ed.), Chicago: University of Chicago Press.

Messick, S. 1989. "Validity," in *Educational Measurement* (3rd ed.), R. L. Linn (ed.), New York: Macmillan, pp. 13-103.

Mingers, J. 2004. "Real-izing Information Systems: Critical Realism as an Underpinning Philosophy for Information Systems," *Information and Organization* (14:2), pp. 87-103.

Trout, J. D. 2000. "Measurement," in *A Companion to the Philosophy of Science*, Volume 18, W. H. Newton-Smith (ed.), Malden, MA: Blackwell, pp. 265-277.

Weber, R. 2004. "Editor's Comments: The Rhetoric of Positivism Versus Interpretivism: A Personal View," *MIS Quarterly* (28:1), pp. iii-xii.

Appendix C

Applicability of the Approach to Information Systems Research

This research aims to contribute to behavioral research in general. However, it also aims to contribute to IS research in two ways. First, past studies show that IS researchers have not dealt with method bias effectively (King et al. 2007; Woszczynski and Whitman 2004). An analysis of recent issues of *MIS Quarterly* and *Information Systems Research* found that this is still a significant problem. Table C1 gives the details. Strikingly, although 80 percent of the constructs in these articles were not defined in terms of a rater, 90 percent were measured with only one rater class, and 90 percent were measured with just one instrument. This contrasts sharply to the proposed approach. These figures might be more acceptable if the monoview strategy was always appropriate and if researchers took other steps to reduce method bias, but in most of the papers examined, this was not the case. Thus, the proposed approach meets a need in IS research.

Second, the approach reveals a rich opportunity for substantive IS research. Many researchers rely on retrospective self-reports only because there are few alternatives. The IS field is uniquely positioned to develop creative IT-based data collection methods. Given the value of metrics in practice (Chidambaram et al. 2005; Pfeffer and Sutton 2006), the creation of IT-based measurement methods to collect real-time self-reports (Stone et al. 2007) or more objective measures (Hilbert and Redmiles 1998), could become an important IS research domain.

Table C1. Dealing with Method Bias in IS Research: Recent Evidence*

	Issues	Number of Articles	No. of Positive Empirical Articles	Avg. No. of Constructs per Article	Avg. No. of Constructs Rater-Defined	Avg. No. of Constructs Measured by One Rater Class	Avg. No. of Constructs Measured by One Instrument
MISQ	3/05 – 6/06	43	30	10	2	10	9
ISR	3/05 – 6/06	30	21	9	2	8	8
Total	3/05 – 6/06	73	51	10	2	9	8

*Data reflects the scores from two independent coders after they resolved all coding differences, indicating 100% agreement among raters. Results are rounded to whole numbers to facilitated interpretation.

References for Appendix C

Chidambaram, L., Zmud, R. W., and Karahannas, M. 2005. "Measuring the Business Value of Information Technology (IT): A Review and Analysis of IT Metrics," in *Business Performance Measurement: Towards Organizational Excellence*, S. S. Kambhammettu (ed.), Chennai, India: Le Magnus University Press, pp. 1-16.

Hilbert, D. M., and Redmiles, D. F. 2000. "Extracting Usability Information from User Interface Events," *ACM Computing Surveys* (32:4), pp. 384-421.

King, W. R., Liu, C. Z., and Haney, M. H. 2007. "Method Effects in IS Survey Research: An Assessment and Recommendations," *Communications of the Association for Information Systems* (20), pp. 457-482.

Pfeffer, J., and Sutton, R. I. 2006. "Evidence-Based Management," *Harvard Business Review* (84:1), pp. 62-74.

Stone, A., Shiffman, SS ., Atienza, A., and Nebeling L. (eds.). 2007. *The Science of Real-Time Data Capture: Self-Reports in Health Research*, New York: Oxford University Press.

Woszczynski, A. B., and Whitman, M. E. 2004. "The Problem of Common Method Variance in IS Research," in *The Handbook of Information Systems Research,* M. E. Whitman and A. B. Woszczynski (eds.), Hershey, PA: Idea Group Inc., pp. 66-77.

Appendix D

Additional Results from the Empirical Demonstration

Table D1 shows the descriptive statistics. There were no significant violations of normality, no evidence of multicollinearity, and no major outliers. A total of 171 students completed the survey but 5 of them did not submit their final Excel files successfully and 5 recordings of time could not be coded. One case was also deleted because of nonsensical responses.

Table D1. Descriptive Statistics

Construct	Rater	Item	N	Mean	Std. Dev.
Deep structure usage	Self-report[1]	DS1	171	6.11	1.73
		DS2	171	6.08	1.68
		DS3	171	6.09	1.59
		DS4	171	6.98	1.56
	Observer[2]	DI1	45	4.40	1.12
		DI2	45	4.31	1.82
Focused immersion	Self-report[1]	FS1	171	5.96	1.89
		FS2	171	5.78	1.80
		FS3	171	5.94	1.65
		FS4	171	5.73	1.68
	Observer	FI1[2]	45	4.15	0.89
		FI2[3]	40	55.88	11.15
Performance	Self-report[1]	PS1	171	6.40	2.30
		PS2	171	5.89	1.97
	Observer[3]	PI1	166	81.01	15.87

Key: DS (deep structure self-report); DI (deep structure independent observer); FS (focused immersion self-report); FI (focused immersion independent observer); PS (performance self-report); PI (performance independent observer).
[1]Uses a 1 to 9 Likert scale.
[2]Uses a 1 to 7 Likert scale.
[3]Uses a 0 to 100% scale.

Table D2 presents evidence for construct reliability and convergent and discriminant validity. The data from the reflective scales converged and discriminated in the expected way and the reliability scores met accepted guidelines (Nunnally and Bernstein 1994). Table D3 shows that all the formative items were highly predictive of their respective traits. Overall, the data suggests that the scales have adequate convergent and discriminant validity and reliability. Further, results of testing the structural model (reported in the main paper) suggest that the scales have adequate nomological validity because they all related in the expected direction.

Table D4 shows the correlations, which were in line with expectations. First, each usage measure was significantly correlated with each performance measure. Second, the correlations showed some evidence of rating bias, since the self-report scales correlate highly among each other and with the common method factor. Finally, the presence of knowledge bias was evident in the varying strength of correlations between the self-report and observer items (from 0.29 for focused immersion to 0.81 for performance), indicating that different raters may have been rating different traits (as in the study by Straub et al. 1995).

Table D2. Instrument Validity and Reliability for Reflective Scales				
	Item	**DS**	**FS**	**Reliability[2]**
Loadings and cross-loadings[1]	DS4	**0.86**	0.53	DS: Cron. α = 0.82 CR = 0.70
	DS3	**0.82**	0.31	
	DS2	**0.81**	0.39	
	DS5	**0.73**	0.38	
	FS4	0.45	**0.84**	CS: Cron. α = 0.81 CR = 0.69
	FS2	0.36	**0.81**	
	FS5	0.36	**0.81**	
	FS1	0.45	**0.73**	
Correlations and AVE[3]	DS	**0.80**	–	
	FS	0.51	**0.81**	

Key: DS (deep structure self-report); FS (focused immersion self-report).
[1]Obtained from PLS. All loadings are significant (p<.05).
[2]Cronbach's alpha obtained from SPSS, CR obtained from PLS.
[3]The values on the diagonal are the square root of each construct's AVE (Average Variance Extracted).

Table D3. Weights for Formative Scales			
Construct	**Rater**	**Item**	**Weight**
Deep structure usage	Observer	DI1	.995
		DI2	.841
Focused immersion	Observer	FI1	.995
		FI2	.715
Performance	Self-report	PS1	.818
		PS2	.973

Key: DI (deep structure independent observer); FI (focused immersion independent observer); PS (performance self-report).

Table 4. Weights for Formative Scales						
	DI	**FS**	**FI**	**PS**	**PI**	**CM**
FS	0.43*					
FI	0.48*	0.29				
PS	0.52*	0.63*	0.49*			
PI	0.67*	0.54*	0.55*	0.81*		
CM	0.52*	0.85*	0.39*	0.82*	0.67*	
DS	0.43*	0.51*	0.26	0.59*	0.49*	0.85*

Key: DS (deep structure self-report); DI (deep structure independent observer); FS (focused immersion self-report); FI (focused immersion independent observer); PS (performance self-report); PI (performance independent observer); CM (common method factor).
*Significant at p < .05 two-tailed.

References for Appendix D

Nunnally, J. C., and Bernstein, I. H. 1994. *Psychometric Theory* (3rd ed.), New York: McGraw-Hill.
Straub, D., Limayem, M., and Karahanna-Evaristo, E. 1995. "Measuring System Usage: Implications for IS Theory Testing," *Management Science* (41:8), pp. 1328-1342.