ICBBR Working Paper Series No 2010_06

**Self-generated Validity, Framing Effects,
and Survey Research in IT**

Wendy Hui

Abstract

Based on the accessibility-diagnostics cognitive framework, we discuss self-generated validity and framing effects as two method biases that can threaten the validity of survey research results. Two empirical studies are used to establish their presence in IS research. We discuss their research implications, with a focus on model comparison and the testing of new instruments or theories. We also suggest a number methodological remedies that can control these method biases, including randomization or partial randomization of question items, temporal separation of construct measurements, replication of research studies, experimental methods and multilevel analysis with split-sample design.

Keywords

IT, Questionnaire surveys, User Acceptance of IT, IT diffusion and adoption

Authors' affiliations

Nottingham University Business School Ningbo, China

Address for correspondence

wendy.hui@nottingham.edu.cn Tel: +8657488180197

Research Note:
Self-generated Validity, Framing Effects, and Survey Research in IS

Wendy Hui
Nottingham Business School in China
University of Nottingham Ningbo China
199 Taikang East Road
Ningbo 315100, P. R. China
wendy.hui@nottingham.edu.cn

31 March 2010

Research Note: Self-generated Validity, Framing Effects, and Survey Research in IS

Abstract

Based on the accessibility-diagnostics cognitive framework, we discuss self-generated validity and framing effects as two method biases that can threaten the validity of survey research results. Two empirical studies are used to establish their presence in IS research. We discuss their research implications, with a focus on model comparison and the testing of new instruments or theories. We also suggest a number of methodological remedies that can control these method biases, including randomization or partial randomization of question items, temporal separation of construct measurements, replication of research studies, experimental methods and multilevel analysis with split-sample design.

Keywords: Questionnaire surveys, User Acceptance of IT, IT diffusion and adoption

Research Note: Self-generated Validity, Framing Effects, and Survey Research in IS

1. Introduction

Managers and behavioral researchers in Information Systems (IS) often rely on self-reports to investigate user perceptions, attitudes and intentions related to the adoption and usage of information technology (IT). The literature suggests that self-reports are subject to various method biases, including social desirability, common method variance, consistency motif and implicit theories, among others (Podsakoff et al. 2003). These biases may systematically inflate or deflate correlations between measures and constructs, and thus affect inferences. Although method biases can affect the validity of research findings, they have not received much discussion in the IS literature until recently (e.g. Molhotra (2006), King et al. (2007), Burton-Jones (2009) and Sharma et al. (2009)).

Studies in applied psychology, organizational behavior and marketing have assessed method biases and have arrived at different conclusions about the size of their effects (e.g. Cote and Buckley 1988, Spector 1994, Doty and Glick 1998). To reconcile these findings, Crampton and Wagner (1994) call for continued “domain specific” research to identify areas that are most susceptible to method biases. In IS, common method variance has generated the most research attention among other method biases. (e.g. Molhotra, 2006, Sharma et al. 2009). However, as pointed out by Burton-Jones (2009), common method variance is only one of many method biases associated with self-reports. This paper focuses on two related method biases: (1) *self-generated validity* and (2) *framing effects*. We provide empirical evidence of their presence in IS research and discuss their significance for research that involves the development of new instruments or new theories and the comparison of alternative or competing models.

Self-generated validity refers to the reactive effects of the measurement of beliefs, attitudes and intentions on the observed correlations between measures and constructs (Feldman and Lynch 1988). Based on developments in social cognition and cognitive psychology, the theory of self-generated validity suggests that at the time of measurement, if a belief, attitude, or intention has not yet existed in the long-

term memory, the construct can be created by the measurement. Depending on their accessibility and perceived diagnosticity¹ in the respondent's mind, answers to earlier questions in a survey may be used as inputs to compute the answer to the current question. More importantly, if earlier questions in a survey were not answered, the respondent might rely on other past experiences and external factors to compute a different answer. Hence, questions in a survey tend to draw the respondent's attention to a limited set of constructs based on the researcher's theoretical model; if these constructs are highly diagnostic for each other, the respondent may ignore relevant past experiences and external factors which would normally be used in the absence of the measurement operation. The increased reliance on earlier answers thus inflates correlations between measures. The effect of self-generated validity is not limited to surveys. Self-reports from interviews are also subject to the same problem. Furthermore, the effect of self-generated validity can persist and affect behavior after the measurement operation (Chandon et al. 2005).

Questions in a survey may also create a frame that consistently induces more or less favorable attitudes across all respondents. These effects are known as *framing effects* or *directional context effects* (Tourangeau et al. 2000). In the IS discipline, several senior researchers have called for research to open the black box of perceived usefulness in order to provide practical design guidelines (Benbasat and Barki 2007, Straub and Burton-Jones 2007). However, if the antecedents of usefulness as well as other constructs in a theoretical model are all measured using the survey method, questions about the antecedents of usefulness may prime the respondents to answer other questions in a specific way. While self-generated validity is mostly concerned with the *correlational* effects on measured constructs (Chandon et al. 2005), framing effects are *directional* such that they shift the overall mean of a construct (Tourangeau et al. 2000, p. 198). As we shall discuss in the following sections, the presence of framing effects may make it difficult to compare competing models.

¹ The perceived diagnosticity of the first judgment for a second (later) one is the degree to which the respondent perceives that the answer to the first question correctly identifies how the second should be answered (Feldman and Lynch 1988).

Self-report data collected through a survey are inherently prone to various types of measurement errors which are difficult to measure or control for effectively. As a result, many researchers, especially those in Economics, are skeptical about survey data (Bertrand and Mullainathan 2001). However, when the theoretical construct of interest resides in the mind of survey respondents, which is often the case in organization psychology, marketing and IS, self-reports may be the best available source of data (Spector 1994). The purpose of this paper is to discuss some of the method biases that are relevant to but have not been adequately addressed in the IS discipline, and to develop specific guidelines to effectively control these biases when we continue our quest to produce research of both theoretical significance and practical relevance.

This paper is organized as follows. First, the relevant literature is reviewed. This is to establish the theoretical basis of our research design and data analysis. Then, two empirical studies are presented that highlight self-generated validity and framing effects and discuss their research implications. Several methodological remedies are proposed to control for their effects. Finally, we conclude the paper with a summary of findings.

2. Literature Review

2.1 Cognitive Aspects of Survey Methodology (CASM)

Theories of human behavior often seek to explain the relationships between variables that reside within an individual's mind, e.g. emotion, perception, attitude and intention. In order to empirically validate such theories, researchers typically measure these variables using self-report data collected through surveys. Some issues inherent in the survey method are yet to be resolved (Edwards 2008) though recent research in social measurement has provided tools for researchers to control for some of these issues. By conceptualizing a measurement to be composed of the trait component (i.e. the real-world property of interest), the method component and the random error, some of the effects of the method component can be estimated using techniques such as multitrait-multimethod procedures (Spector 1987), confirmatory

factor analysis and structural equation modeling (e.g. Williams et al. 1989, Bagozzi and Yi 1990). However, the estimated size of the effects depends very much on the assumptions made about the nature of the method component, thus different studies may yield different conclusions about the extent to which the method component confounds measurements (Williams and Brown 1994). Some even doubt whether it is possible to separate the trait, method and random error components without a strong theory of method bias that delineates the various types of biases and explains how each might affect measurements under different settings (Schmitt 1994). Therefore, before survey data are collected, researchers are wise to avoid survey design and procedures that have been identified to be susceptible to method biases (Podsakoff et al. 2003).

Studies in the cognitive aspects of survey methodology (CASM) have provided insights on how to improve the quality of survey research. According to Sirken and Schechter (1999), the cognitive paradigm started in the early 1980s. Research in this area has since broadened and deepened our understanding of the cognitive processes underlying survey method biases through the use of experiments and laboratory methods such as think-aloud interviews and intensive probing. Research in CASM has identified survey structures, questions and topics that are likely to be error-prone and has shown how cognitively burdensome questions can be detected and corrected (Schwarz 1999, Tourangeau et al. 2000). Results from CASM research have produced guidelines on how to develop measurement scales, use appropriate question wording and position question items in a survey. The aim of this paper is to discuss some of the CASM research findings relevant to IS research. In response to the recent call by Benbasat and Barki (2007) for opening the black box of perceived usefulness and the constant need for IS researchers to develop new instruments to investigate emerging IT phenomena, this paper focuses on self-generated validity and framing effects.

2.2 Self-Generated Validity

Feldman and Lynch (1988) propose the accessibility-diagnostics model to explain the cognitive process underlying the responses to survey questions. Although method bias in survey research is a major focus of

Feldman and Lynch's (1988) work, the model itself is quite general. It can be applied to other human decision making situations, e.g. brand evaluation (Ahluwalia and Gurhan-Canli 2000) and persuasion (Aaker 2000). In the accessibility-diagnostics model, *accessibility* refers to the ease with which an input in memory is brought to mind, whereas the *diagnostics* of an input in the memory refers to the degree to which an individual believes that the decision implied by that input alone would accomplish his or her decision goals (Lynch 2005). In a survey context, an answer to an earlier question A is diagnostic for a subsequent question B if the respondent thinks that the answer to A correctly identifies how B is answered. The diagnostics of A for B may be different from the diagnostics of B for A. The accessibility-diagnostics model suggests that an input in the memory can be used to determine a related judgment or behavior as a positive function of its own accessibility and diagnostics and an inverse function of the accessibility and diagnostics of alternative inputs (Lynch 2006).

Feldman and Lynch (1988) note that very often in a survey context, constructs of interest to the researcher do not exist in the memory of the respondent before questions are asked. Answers to earlier questions about a construct can increase the construct's accessibility in the memory, making it more likely to be used to compute answers to later relevant questions about constructs that do not exist in the memory. As a result, the likelihood that the respondent will use other relevant information in the memory is reduced. Information that is only moderately diagnostic and that would normally be used if the earlier questions had not been asked may even be ignored. Figure 1 is a graphical representation of the situation. Figure 1(a) depicts the situation where X is not measured. In this situation, a moderately diagnostic input Y may be used to compute Z, which does not exist before it is asked about. Figure 1(b) depicts the situation where X is asked before Z. In this case, X becomes accessible to the respondent when Z is asked and the respondent may ignore Y and rely solely on X to compute Z. Thus, the measurement of constructs in a theory through a survey can have a distorting effect that tends to inflate the correlations among measures and constructs. This distorting effect is known as *self-generated validity*. It can lead to artificially high internal consistency when a construct is measured by multiple items. The problem may be

lessened by having the items interspersed throughout a survey, depending on the level of similarity among the items and the distance between the items. Harrison and McLaughlin (1993) offer some empirical support for this proposition.

Figure 1. Possible Effect of Survey Questions on the Use of Moderately Diagnostic Inputs in Computing an Evaluation

The emergence of innovative technologies and new applications create interesting social phenomena that generate new ideas for IS researchers. Very often, new theories involve constructs for which no developed instrument exists in the literature. Even if a developed instrument exists, researchers may need to modify question items to make them relevant to a specific research context. Relationships between constructs are undoubtedly of interest to researchers. However, if the auxiliary theory for a measurement model does not exist, the structural level relationships cannot be meaningfully tested (Edwards and Bagozzi 2000). Therefore, when IS researchers include newly developed instruments in their survey, a questionnaire structure that minimizes self-generated validity should be used. For instance, question items can be randomized interspersed across the survey or part of the survey (Budd 1987), an irrelevant cover story can be inserted in the middle of a survey, and different response formats such as semantic differentials, Likert scales, open-ended questions, etc. can be used (Podsakoff et al. 2003).

Despite being equally susceptible to self-generated validity, most established instruments in IS have been developed with methodological rigor (e.g. Technology Acceptance Model (TAM) constructs by Davis (1989)) and tested extensively under a wide variety of research settings using different questionnaire formats (e.g. Davis and Venkatesh 1996). In such a way, researchers can accept the reliability and validity of established instruments with confidence. Newly developed instruments that have not been subject to such widespread scrutiny should ideally be developed to include additional steps such as the generation of a comprehensive item pool and review by experts (DeVellis 2003). In IS, it has not been a common practice to report the details of how a new instrument is developed if it is not the major focus of the research study. Often, it is not indicated whether or not question items are randomized or partially randomized. If Cronbach's alpha values are already high, after subtracting an appropriate amount to account for any self-generated validity, they may still exceed 0.6 (or 0.7), the commonly suggested minimum requirement in exploratory (or confirmatory) research (Nunnally 1967). However, if self-generated validity is present and Cronbach's alpha values are close to the minimum acceptable requirement, the constructs may actually not be reliable for further structural level analysis.

Besides inflating the correlations between the measures of the same construct, self-generated validity can also inflate the correlations between constructs. For example, if construct A is highly diagnostic for construct B and a question about A is directly followed by a question about B, then the respondent is likely to use the answer to the question about A to compute the answer to the question about B, if B does not exist in the respondent's memory. Such correlation inflation will lead to inflated path significance and affect the amount of variance explained. The likelihood of falsely rejecting the null hypothesis or falsely accepting a poor theory may be high if path significance is only moderate even under the influence of self-generated validity. Therefore, when testing a new theory, it is important to ensure that self-generated validity is minimized through good survey procedures.

Study 1 of this paper investigates the effects of item randomization and temporal separation on construct reliability and between-construct correlations, and discusses the advantages and disadvantages of their use.

2.3 Framing Effects

The effects of self-generated validity are mostly correlational (Chandon et al. 2005). Suppose an earlier question about the processing speed of a technology is followed by a question about perceived quality of the technology. According to the accessibility-diagnostics model, respondents who find processing speed to be diagnostic for quality will use their answer to the first question to compute the answer to the second question, if it does not exist in the long-term memory. A respondent who finds the technology slow is likely to perceive a low quality, whereas a respondent who finds the technology fast is likely to perceive a high quality. Hence, the correlation between processing speed and quality is strengthened due to the increased accessibility of processing speed at the time when the question about quality is answered. However, the overall sample mean of perceived quality may not be affected.

On the other hand, framing effects refer to the priming effects of earlier questions that alter responses to later questions regardless of the responses to the earlier questions. In a study conducted by Tourangeau et al. (1989), respondents were divided into three groups. The first group was asked about military threats posed by the Soviet Union prior to the question on defense spending, whereas the second and third groups were asked about the need for arms control and other unrelated questions. It is found that regardless of how the Soviet threat questions were answered, respondents in the first group showed significantly greater support for an increase in defense spending. In other words, even for respondents who did not see much military threat from the Soviet Union, the questions about such threats induced more favorable attitudes towards increased defense spending.

One of the major research interests in the IS discipline is to identify factors that can increase IT productivity. Technology adoption research in the past two decades has provided ample evidence to

support the perceived ease of use and the perceived usefulness of a technology as significant predictors of a user's intention to use the technology. While much less research attention has been paid to investigate the antecedents of perceived usefulness to provide practical design guidelines (Benbasat and Barki 2007), the quest is likely to involve the investigation of task and system characteristics which that can influence a user's perception of usefulness. If constructs in a model that attempts to open the black box of perceived usefulness are all measured by self-reports, questions about task and system characteristics may prime respondents to answer the questions about perceived usefulness and intention more or less favorably in a systematic manner.

Suppose a researcher is interested in studying the factors that affect the adoption of a security technology using the protective motivation theory (PMT) (Rogers 1975), which models protective motivation as a function of perceived severity, perceived vulnerability and response efficacy. In the context of adopting a security technology, protective motivation refers to the intention to use the security technology, perceived severity refers to the perceived damage from a security incident, perceived vulnerability refers to the perceived likelihood of a security incident, and response efficacy refers to the effectiveness of the technology in minimizing the effects of a security incident. Asking respondents about their perceived severity and perceived vulnerability of a security threat may induce a greater intention to use the security technology, thus distorting the observed correlations between intention to use and other constructs in the theory.

In using a latent model of intention, Chandon et al. (2005) show that the measurement of purchase intention not only strengthens the correlation between purchase intention and actual purchase behavior, but also increases actual purchase behavior. The implications of framing effects are twofold. First, a predictive model (for example, of user satisfaction) based on self-reports may not be as accurate as desired when applied to the target population who have not received the measurement operation. The second implication relates to model comparison. There is a need for model comparison when a researcher wants to compare various alternative models in an exploratory study to produce a final model that is

parsimonious and, at the same time, explains a good amount of data variance. As shown in Tourangeau et al. (1989), a survey question that brings framing effects can affect the response to a target question even when the question does not significantly correlate with the target question. A model that omits this question on the basis of an insignificant path coefficient may not accurately reflect the reality despite demonstrating high variance explained in the final model. This is because without responding to the omitted question, the respondents may have answered the other survey questions differently. For example, in testing the model in Figure 2(a), it is possible that although questions about system characteristics do not significantly explain perceived usefulness, the thinking process involved reduces the relative accessibility of perceived ease of use when the respondent computes the answer to perceived usefulness, leading to an insignificant path between perceived ease of use and perceived usefulness. As a result, the final model (Figure 2(b)) does not show an arrow between perceived ease of use and perceived usefulness even though it might appear if the system characteristics had not been measured (Figure 2(c)).

Figure 2. An Example of Hypothetical Model Development

A model comparison is also required when a researcher wants to critically assess competing theories (e.g. Davis et al. 1989, Mathieson 1991, Taylor and Todd 1995). The researcher may include all constructs of the competing theories in a survey and evaluate each theory without incorporating constructs that are not part of the theory. However, this approach will face the same validity threat as in the hypothetical model development example illustrated in Figure 2. Alternatively, the researcher may divide respondents into groups so that each group is exposed to constructs belonging to only one theory. The problem with this approach is that different sets of questions may change the respondents' opinions in different ways, thus preventing the researcher from making a meaningful comparison between groups. Framing effects should not be problematic if none of the questions in a survey significantly alter the respondents' mind. However, probing into the respondents' perceptions about task and system characteristics is likely to produce priming effects in the measurement. This suggests that methodological remedies effective in controlling framing effects may be needed in future research that seeks to open the black box of perceived usefulness using the survey method.

To account for framing effects (as well as self-generated validity) in predictive models, researchers can develop a latent model following the procedure described in Chandon et al. (2005). The drawback of this method is that it requires the identification of questions that cause the method biases and the inclusion of a control group who are not asked these questions. The implementation of such a research design may be costly. Other ways to avoid framing effects for a predictive or theoretical model includes avoiding the use of self-reports, complementing the survey method with experimental methods, and conducting multilevel analysis as opposed to individual level analysis. More discussion on these methodological remedies is presented in the Discussion section for Study 2, which shows that there is a significant difference between respondents who answer questions about constructs in TAM and respondents who answer questions about constructs in PMT in their intention to use a security technology.

3. Study 1 – Self-Generated Validity

3.1 Hypotheses

Randomization of question items (Budd 1987) and temporal separation (Podsakoff et al. 2003) are the two suggested ways to reduce the effects of self-generated validity. Using a 1 x 6 experiment design, Study 1 provides empirical evidence of the presence of self-generated validity in survey research in IS by showing that randomization of question items can significantly reduce the Cronbach's alpha values of constructs and that temporal separation of construct measurement can weaken the between-construct correlations. Specifically, this study tests the following hypotheses:

- H1: The Cronbach's alpha values of constructs measured in a randomized questionnaire design are smaller than the Cronbach's alpha values of constructs measured in a nonrandomized questionnaire design.
- H2: Construct correlations between constructs measured with temporal separation are smaller than the construct correlations between constructs measured without temporal separation.

3.2 Research Methodology

The subjects involved in this study were students of an introductory IT course. A total of 400 students were randomly selected and invited to take part in the study at the beginning of the Fall semester of 2009. They were randomly assigned to six groups and each group was asked to complete an online survey that was divided into three parts. The online survey system, which was part of the course website, placed a 12-hour separation between Part 1 & Part 2 and Part 2 & Part 3. Student participation was not compulsory but those who took part in the survey would have their print accounts topped up by an amount equivalent to 100 A4-sized pages. In addition, they received a chance to win some fashion accessories.

The theory of interest is a version of TAM that consists of four constructs: perceived ease of use (PEOU), perceived usefulness (PU), attitude (ATT) and intention (INT) (Figure 3). Measures are adapted from Venkatash et al. (2003) with changes in wordings to suit the research context (Appendix A).

Responses to questions on PEOU, PU and INT are rated on a 7-point Likert scale, with 1 representing “strongly disagree” and 7 representing “strongly agree”. ATT is measured by 7-point semantic differentials using bipolar adjectives “bad-good,” “foolish-wise”, “dislike-like” and “unpleasant-pleasant”. The technology examined is the course blog. The subjects were told in the first lecture that the instructor would like to experiment with the use of a course blog as a platform for knowledge sharing after the third week of the course. The course consisted of three sections. At the beginning of the first lecture for each section, the course blog was activated so that the instructor could present the blog to the students and give them some idea of what the blog would be like. At the end of the lecture, the blog was inactivated. Throughout the data collection period, the subjects could not access the blog.

Figure 3. Research Model in Study 1

Besides the TAM constructs, other irrelevant constructs such as gender, age, emotional intelligence, tendency to trust, individualistic orientation, and locus of control were also measured. These constructs were used in another research study. All the six groups answered 25 questions in each part of the survey. Groups 1 to 3 answered all the TAM questions in Part 1 but in a different order, while Groups 4 to 6 answered the TAM questions in different parts of the survey (Table 1).

Table 1. Experimental Design for Study 1

	Questions in Part 1	Questions in Part 2	Questions in Part 3
Group 1	PEOU, PU, ATT and INT (in this order) followed by 9 irrelevant questions in random order	25 irrelevant questions in random order	25 irrelevant questions in random order
Group 2	INT, ATT, PU and PEOU (in this order) followed by 9 irrelevant questions in random order	25 irrelevant questions in random order	25 irrelevant questions in random order
Group 3	INT, ATT, PU and PEOU and 9 irrelevant questions (all in random order)	25 irrelevant questions in random order	25 irrelevant questions in random order
Group 4	PEOU, PU and 17 irrelevant questions (all in random order)	ATT and 21 irrelevant questions (all in random order)	INT and 21 irrelevant questions (all in random order)
Group 5	INT and 21 irrelevant questions (all in random order)	ATT and 21 irrelevant questions (all in random order)	PEOU, PU and 17 irrelevant questions (all in random order)
Group 6	PEOU, PU, ATT, INT and 59 irrelevant questions (all in random order across three parts of the survey)		

3.3 Data Analysis

The number of respondents who answered all TAM questions in each group is shown in Table 2. The lower number of respondents in Groups 4 to 6 compared to Groups 1 to 3 is due to attrition since the subjects in Groups 4 to 6 completed all parts of the survey. Probably due to network problems, a number of records were not properly recorded. Therefore, the total number of usable records is slightly less than the number of respondents who answered all TAM questions.

Table 2. Sample Size for Each Group (Respondents Who Answered All TAM Questions)

	Group (Order)	Number of Respondents Who Answered All TAM Questions	% of Total Invitations Sent to Group	Number of Usable Records
Constructs Measured without Temporal Separation	Group 1 (Logical)	51	75%	51
	Group 2 (Reverse logical)	50	74%	49
	Group 3 (Random)	50	75%	50
Constructs Measured with Temporal Separation	Group 4 (Logical)	28	42%	26
	Group 5 (Reverse logical)	30	46%	29
	Group 6 (Random)	31	47%	30

3.3.1 Construct Reliability

As discussed in the Literature Review section, the accessibility-diagnostics model suggests that randomization of question items can weaken construct reliability. Table 3 presents the Cronbach's alpha values of the TAM constructs under different experimental conditions in our study. Group 3 and Group 6 who answered all the TAM questions in a random order² are highlighted.

Table 3. Cronbach's Alpha Values of TAM Constructs in Each Experimental Group

	Constructs Measured without Temporal Separation			Constructs Measured with Temporal Separation		
	Group 1 (Logical)	Group 2 (Reverse logical)	Group 3 (Random)	Group 4 (Logical)	Group 5 (Reverse logical)	Group 6 (Random)
Perceived Ease of Use (PEOU)	0.869	0.912	0.868	0.864	0.831	0.679
Perceived Usefulness (PU)	0.896	0.938	0.812	0.756	0.797	0.786
Attitude (ATT)	0.896	0.933	0.905	0.918	0.884	0.815
Intention (INT)	0.935	0.929	0.888	0.908	0.910	0.838

As shown, all Cronbach's alpha values are greater than 0.7, except that for PEOU in Group 6, providing further support for the robustness of TAM instruments. The purpose of Study 1, however, is to show that the Cronbach's alpha values decrease *significantly* in a randomized questionnaire design compared to a nonrandomized design. Thus, if an instrument has not been properly developed, a nonrandomized questionnaire design may not be able to detect items that should be dropped.

The Fisher-Bonett test proposed by Bonett (2002, 2003) and evaluated by Kim and Feldt (2008) is used to test whether the Cronbach's alpha values between the randomized and non-randomized questionnaire designs are statistically different after taking into account the difference in sample size. Table 4 presents the test results, where α denotes the Cronbach's alpha value and the subscript denotes the experimental group. As shown, the Cronbach's alpha values associated with the randomized designs are

² Group 4 and Group 5 answered the PEOU and PU questions randomized among 17 other irrelevant questions, the ATT questions randomized among 21 other irrelevant questions, and the INT questions among 21 other irrelevant questions. However, the order of presentation at the construct-level is not completely randomized (See Table 1).

in general smaller than those associated with the non-randomized designs. Some of the differences are statistically significant. Hence, Study 1 provides empirical support for H1.

Table 4. Fisher-Bonett Test Results

	Constructs Measured without Temporal Separation		Constructs Measured with Temporal Separation	
	$H_0 : \alpha_3 < \alpha_1$	$H_0 : \alpha_3 < \alpha_2$	$H_0 : \alpha_6 < \alpha_4$	$H_0 : \alpha_6 < \alpha_5$
Perceived Ease of Use (PEOU)	$p = 0.495$	$p = 0.115$	$p = 0.029^*$	$p = 0.072$
Perceived Usefulness (PU)	$p = 0.036^*$	$p < 0.001^{***}$	$p = 0.615$	$p = 0.454$
Attitude (ATT)	$p = 0.600$	$p = 0.149$	$p = 0.035^*$	$p = 0.142$
Intention (INT)	$p = 0.048^*$	$p = 0.087$	$p = 0.108$	$p = 0.092$

* Significant at the 0.05 level. ** Significant at the 0.01 level. *** Significant at the 0.001 level.

3.3.2 Between-Construct Correlations

According to the accessibility-diagnostics model, when a construct in a theory is diagnostic for other constructs in the theory, responses to questions about the construct can be used to compute responses to questions about the other constructs, thus resulting in inflated construct correlations. A lack of temporal separation between construct measurements makes it easy for respondents to recall responses to questions about one construct and use them to compute responses to questions about other constructs. Therefore, construction correlations from designs that do not have a temporal separation between construct measurements are expected to be stronger than those from designs that do.

Table 5 presents the construct correlation matrices for the experimental groups with the diagonal entries in each matrix being the square root of the average variance extracted (AVE) by the constructs. Although the purpose of Study 1 is not to assess TAM, as an aside, we note that the AVE of each construct exceeds 0.5, the benchmark for convergent validity (Fornell and Larcker 1981). The square root of the AVE of each construct is greater than the correlations between the construct and other constructs in the model, suggesting adequate discriminant validity (Fornell and Larcker 1981).

Table 5. Construct Correlation Matrices for Experimental Groups

Constructs Measured without Temporal Separation					Constructs Measured with Temporal Separation				
Group 1 (Logical order)					Group 4 (Logical order)				
	PEOU	PU	ATT	INT		PEOU	PU	ATT	INT
PEOU	0.846				PEOU	0.841			
PU	0.734	0.873			PU	0.587	0.760		
ATT	0.631	0.716	0.874		ATT	0.272	0.541	0.895	
INT	0.593	0.631	0.757	0.915	INT	0.408	0.651	0.632	0.888
Group 2 (Reverse logical order)					Group 5 (Reverse logical order)				
	PEOU	PU	ATT	INT		PEOU	PU	ATT	INT
PEOU	0.890				PEOU	0.820			
PU	0.736	0.919			PU	0.692	0.799		
ATT	0.683	0.764	0.912		ATT	0.585	0.529	0.862	
INT	0.652	0.660	0.752	0.908	INT	0.282	0.250	0.349	0.873

NB. Diagonal Entries are Average Variance Extracted (AVE) by Constructs.

Correlations between constructs measured at different times are highlighted in Table 5. Using the formula provided in Cohen et al. (2002), tests were performed to determine whether the highlighted construct correlations between Group 4 and Group 5 are the same as the corresponding construct correlations between Group 1 and Group 2, respectively. The test results are presented in Table 6. Probably due to the small sample size of Group 4 and Group 5 and thus insufficient statistical power, some of the null hypotheses are not rejected despite seemingly big differences observed in construct correlations between groups with and without temporal separation. The difference in $r_{PEOU, ATT}$ between Group 1 and Group 4 and the difference in $r_{PU, ATT}$ between Group 2 and Group 5 are marginally significant (it would be significant at the 0.05 level on the basis of a one-tailed test). The differences in $r_{PEOU, INT}$, $r_{PU, INT}$, $r_{ATT, INT}$ between Group 2 and Group 5 are significant at the 0.05 level. Overall, the present study provides support for H2.

Table 6. Tests for Equivalence of Correlations

	Group 1 (Logical order)	Group 4 (Logical order)	Sig.	Group 2 (Reverse logical order)	Group 5 (Reverse logical order)	Sig.
$r_{\text{PEOU, ATT}}$	0.631	0.272	0.067 ⁺	0.683	0.585	0.506
$r_{\text{PU, ATT}}$	0.716	0.541	0.246	0.764	0.529	0.089 ⁺
$r_{\text{PEOU, INT}}$	0.593	0.408	0.327	0.652	0.282	0.046*
$r_{\text{PU, INT}}$	0.631	0.651	0.893	0.660	0.250	0.029*
$r_{\text{ATT, INT}}$	0.757	0.632	0.338	0.752	0.349	0.013*

⁺ Significant at the 0.1 level. * Significant at the 0.05 level. ** Significant at the 0.01 level. *** Significant at the 0.001 level.

3.3.3 Assessing the Effects of Attrition

Subjects in Group 4 and Group 5 completed all three parts of the survey, while some subjects in Group 1 and Group 2 completed only one or two parts. It is possible that those who completed the entire survey tend to answer the questions more (or less) consistently than those who did not. To evaluate the effects of attrition, we perform the same set of analyses including only subjects in Group 1 and Group 2 who completed all parts of the survey. The analysis results are presented in Appendix B. One-tailed Fisher-Bonett Tests for constructs measured with vs. without item randomization yield three significant differences at the 0.05 level and two at the 0.01 level (Table A1). Two-tailed tests for equivalence of correlations between constructs measured with vs. without temporal separation yield two significant differences at the 0.05 level and three at the 0.10 level. Overall, the test results seem to have lost some statistical significance due to decreased sample size, but they continue to allow us to draw the same inferences.

3.4 Discussion

An artificially high correlation between two question items may result if one item that is highly diagnostic for another is placed in proximity in a survey. Hence, when question items measuring the same construct are grouped together in a survey, the construct reliability will be exaggerated. A randomized questionnaire design can reduce the effects of self-generated validity by creating distance between question items of the same construct. Study 1 empirically shows that such design can significantly lower

the construct reliability in terms of Cronbach's alpha and hence alleviate the problem of inflated construct reliability due to self-generated validity.

In a similar fashion, when constructs that are highly diagnostic for each other are placed close to each other, self-generated validity may inflate their correlations, thus increasing the likelihood of a false rejection of the null hypothesis and an exaggerated R^2 . Temporal separation is shown in this study as a promising method that can alleviate the problem of inflated construct correlations resulting from self-generated validity. It is observed that the design in Group 4 and Group 5 characterized by temporal separation between construct measurements *and* randomization of questions among other irrelevant questions at each time of data collection seems to be least affected by self-generated validity. However, the introduction of temporal separation may result in high attrition. The technique will inevitably make the data collection exercise more costly and time-consuming.

Another disadvantage of temporal separation is the potential contamination effects (e.g. temporary mood state) of intervening factors between measurements. The intervening factors may mask the construct relationships that really exist. Thus, temporal separation may increase the likelihood of falsely accepting the null hypothesis. Podsakoff et al. (2003) suggest that careful calibration of the time lag between measurements can minimize the contamination effects of intervening factors. Ultimately, the researcher needs to weigh the cost of the false rejection of the null hypothesis against the cost of its false acceptance. In this paper, we argue that the false acceptance of the null hypothesis is probably the lesser of the two evils in the testing of new instruments and new theories. This is because the rejection of the null hypothesis may produce poor business guidelines which, if relied upon, may result in inefficient allocation of resources. Yet the false acceptance of the null hypothesis should not prevent researchers from further investigation of moderating conditions under which the relationship of interest might be heightened. It is believed that the false acceptance of the null hypothesis will eventually be corrected by continued exploratory research efforts.

On a different note, within the academic research community, greater faith in the theory can be engendered if the theory can stand even under methodologically adverse conditions. Therefore, when measurements are to be made using a survey, the use of a randomized questionnaire design is recommended, especially when the survey consists of new instruments. If the questionnaire is long and complete randomization is likely to cause fatigue or even annoyance to the respondents, partial randomization can be adopted such that items of a newly developed instrument are intermixed with items of one or two well-established instruments in the model. The use of temporal separation can be used if it is feasible under the constraints of time and cost. Last but not least, other techniques that are not empirically evaluated in this study can also be considered, e.g. the use of different response formats and data collection media for construct measurements (Podsakoff et al. 2003).

4. Study 2 – Framing Effects

4.1 Hypothesis

Opening the black box of perceived usefulness has been identified as an important research question in the IS discipline. In this study, we discuss some concerns about using the survey method to address this question. Study 2 is a simple experiment that illustrates the potential intrusive effects of the measurement operation on survey respondents. It tests the following hypothesis:

H3: The mean value of a measurement may be different depending on what other measurements are made in the survey.

Specifically, it shows that individuals' intention to use a technology can be affected by the questions presented in the survey.

4.2 Research Methodology

Around the middle of the Fall semester 2009, 681 students from the same introductory IT course who had indicated earlier that they would be interested in further online survey studies were invited to participate in Study 2. The survey lasted for 3 days and students were encouraged to participate by offering them

chances to win some stationery products. Participants were alternately assigned to one of two experimental groups based on their order of participation. The first participant was assigned to Group 1, the second to Group 2, the third to Group 1, the fourth to Group 2, and so on. Group 1 answered randomized questions about constructs in a simple version of TAM2, an extension of the original TAM (Venkatesh and Davis 2000); whereas Group 2 answered randomized questions about constructs in PMT (Rogers 1975). The research model is presented in Figure 4. Intention to use is the common construct for both groups and is measured by the same question items in both groups. Subjects in both groups were asked to answer a total of 12 questions (Appendix C) which are rated on a 7-point Likert scale, with 1 representing “strongly disagree” and 7 representing “strongly agree”. The technology examined was Hushmail, a web-based email service that protects email privacy using the PGP technology. Both groups were given a brief description of the purpose of the technology before they were presented with the survey questions.

Figure 4. Research Model in Study 2

4.3 Data Analysis

Altogether 369 subjects participated in the survey, representing 54% of the target population. 22 of the surveys were not complete. The final dataset consists of 176 records from Group 1 and 171 records from Group 2. Table 7 presents the Cronbach's alpha values of constructs in both TAM2 and PMT. All of them exceed 0.7, the threshold for reliability suggested by Nunnally (1967). Table 8 presents the construct correlation matrix for each model, with the diagonal entries being the square root of the AVE of the constructs. The AVE of each construct exceeds 0.5, the benchmark for convergent validity (Fornell and Larcker 1981). The square root of the AVE of each construct is greater than the correlations between the construct and other constructs in the model, suggesting adequate discriminant validity (Fornell and Larcker 1981).

Table 7. Cronbach's Alpha Values of Theoretical Constructs in Each Experimental Group

Group 1 (TAM2)		Group 2 (PMT)	
Perceived Ease of Use (PEOU)	0.805	Perceived Severity (PS)	0.710
Perceived Usefulness (PU)	0.883	Perceived Vulnerability (PV)	0.831
Subjective Norm (SN)	0.819	Response Efficacy (RE)	0.827
Intention (INT)	0.878	Intention (INT)	0.842

Table 8. Construct Correlation Matrices for Experimental Groups

Group 1 (TAM2)					Group 2 (PMT)				
	EOU	PU	SN	INT		PS	PV	PRE	INT
EOU	0.847				PS	0.798			
PU	0.726	0.900			PV	0.079	0.865		
SN	0.682	0.753	0.857		PRE	0.361	0.086	0.862	
INT	0.682	0.839	0.765	0.897	INT	0.318	0.161	0.460	0.871

NB. Diagonal Entries are Average Variance Extracted (AVE) by Constructs.

The path analysis results are presented in Figure 5. The findings indicate that paths in PMT in general show a lower level of statistical significance than paths in TAM2. The R^2 for INT in PMT is much lower than that in TAM2. On the surface, it seems to suggest that PMT is a weaker theory. However, the mean INT in PMT (4.37) is higher than that in TAM2 (4.10), and the difference is statistically significant ($t = 2.151, p = 0.032$).

Figure 5. Results for Path Analysis in Study 2

4.4 Discussion

Hushmail is a technology designed to protect email privacy. PMT explores questions related to security threats that the technology is designed to prevent. When answering these questions, information relevant to the decision about INT is likely to become more salient in respondents' working memory, resulting in more positive attitudes towards the technology. TAM2 explores general questions relevant to a typical technology adoption decision. Answering these questions is likely to invoke a different set of information in the working memory, which affects responses to INT in a different way. As a result, the PMT group shows a significantly higher INT than the TAM2 group does. Interestingly, the correlations between INT and its antecedents in PMT are lower than those between INT and its antecedents in TAM2. Hence, in the PMT group, even for respondents who do not rate the PMT antecedents highly, the exercise of answering questions about them is sufficient to increase the rating on INT. This finding resonates with that of Tourangeau et al. (1989), where questions about the Soviet threat systematically increase attitudes towards increased defense spending regardless of the respondents' perception about the Soviet threat. A

possible explanation for the empirical paradox observed in this study is that although thoughts about the PMT antecedents tend to induce a higher INT than thoughts about the TAM2 antecedents do, the PMT antecedents are not perceived to be as diagnostic as the TAM2 antecedents for INT.

The intrusive framing effects of survey measurement have two implications. First, the measurement may alter the beliefs and attitudes of the respondents, albeit temporarily, such that at the time of data collection they are not representative of the target population, and thus hurting the accuracy of the resultant predictive models. This is a well-known problem in commercial marketing research where researchers often rely on self-reports to develop predictive models for future purchasing behavior. A simple solution to improve the performance of predictive models of behavior is to apply appropriate discounting factors to the predictions. A rate conversion system similar to that of the ACNielsen's BASES system can also be used to adjust forecasts made by predictive models (Chandon et al. 2005).

The second implication relates to model comparison. For research that focuses on the testing of a standalone model, framing effects may not cause a practical concern. However, when framing effects are expected to be significant (e.g. when questions probe into task and system characteristics of an information system), exploratory researchers assessing alternative models based on the path significance of factors may end up omitting factors that actually induce thoughts highly relevant to the target variable. Omission of such factors may hurt the overall quality of the final model. Therefore, replication of the study without measuring the omitted constructs can be conducted to ensure that none of the statistical properties of the remaining constructs change significantly and that the final model remains satisfactory. Similarly, when framing effects are expected to be significant, the comparison of competing theories based on survey data may not be meaningful no matter if all constructs in the competing theories are collected from a single survey or from multiple experimental groups who are exposed only to questions about one single theory. This is because, if all constructs in the competing theories are measured in a single survey, measurements for one theory may be confounded by measurements for another theory; and

if respondents only answer one set of questions related to one theory, the questions in different theories may change the respondents in different ways, making it difficult to perform a fair comparison.

One way around the problem is to complement the survey with experimental methods by running both designs in parallel. Suppose a researcher is interested in comparing two competing theories. Theory 1 predicts Y with X_1 and Theory 2 predicts Y with X_2 . Instead of having only one group answering all questions about X_1 , X_2 and Y , or two groups with one answering questions about X_1 and Y and the other answering questions about X_2 and Y , the researcher can have three groups, with Group 1 answering all questions, Group 2 answering questions about X_1 and Y , and Group 3 answering questions about X_2 and Y . Using the measurements of X_1 and Y from both Group 1 and Group 2, the researcher can determine the effects of the measurement of Theory 2 on the observed relationships between the constructs in Theory 1 as shown in Figure 6. Similarly, using the measurements of X_1 and Y from both Group 1 and Group 3, the researcher can determine the effects of the measurement of Theory 1 on the observed relationships between the constructs in Theory 2.

Figure 6. Controlling for the Measurement of Theory 2 in the Estimated Model for Theory 1

Obviously, supplementing a survey with experimental methods this way is cumbersome and time-consuming. Before engaging in such an exercise, it is probably wise to first find out through a pilot study if there is such a need, i.e., if any questions will likely cause framing effects; and if so, whether it is

possible to avoid the use of self-reports. In fact, in order to move the research on technology adoption forward, various senior researchers have called for the measurement of key IS theoretical constructs such as usefulness and system usage beyond user perceptions (Benbasat and Barki 2007, Straub and Burton-Jones 2007).

Finally, if multiple technologies are investigated and if system characteristics are identified to be the constructs that cause framing effects, multilevel analysis can be conducted at both the individual as well as the technology levels. System characteristics form the technology level variables and individual beliefs, attitude and intention form the individual level data. Technology level variables can be aggregated from one group of respondents, while the individual-level variables are obtained from another group of respondents. This split-sample design prevents measurement of system characteristics to affect the measurement of individual beliefs, attitudes and intentions.

5. Conclusion

The validity of empirical research can be jeopardized when the measurements involved are unreliable. Many social phenomena and properties of objects of interest to social scientists are complex, difficult to operationalize, and prone to measurement errors. Although several fundamental issues related to self-reports and subjective data remain unresolved, past research efforts on causes of method biases and ways to control them have provided researchers reasonably reliable tools to quantify variables of interest and to study them scientifically. Research in IS often seeks to theorize relationships between IT, use of IT and IT productivity. Some of the variables in IS theories can be fairly easily operationalized and measured objectively. However, like other social sciences, there are situations where objective measures are expensive or where subjective measures are the best data source. In such situations, researchers' awareness of potential measurement errors needs to be heightened so that methodological remedies can be effectively devised and adopted to control for their effects.

Adopting a cognitive model for answering questions enables us to understand the nature of some of the measurement errors associated with the survey method. Based on the accessibility-diagnostics model, this paper has analyzed the method bias of self-generated validity, which has the tendency to inflate the correlations between question items and between constructs that are highly diagnostic for each other. Study 1 of this paper presents an empirical study that shows how these correlations can be reduced by a randomized questionnaire design and temporal separation between construct measurements. Besides artificially inflated correlations, survey questions can induce framing effects that systematically alter the mind of the survey respondents, hurting the accuracy of the resultant predictive model and preventing a fair model comparison. Study 2 shows that that framing effects can be directional such that favorable or unfavorable attitude is induced by the measurement operation.

This work suggests self-reports may need to be approached with caution for certain types of research problems. When testing new instruments and theories, randomization or partial randomization of question items and temporal separation can be used to improve the validity of research findings. The effects of framing effects can be reduced by replication of exploratory survey studies. When comparing competing models, experimental methods can be used to complement the survey method. When multiple technologies are investigated, multilevel analysis using a split-sample design can eliminate the problem of framing effects. The choice on which methodological remedy to adopt is dependent on the specific research context.

In the IS discipline, perceptions of a technology, typically measured by self-reports, have been established as significant predictors of technology adoption and usage. In an attempt to extend our understanding of adoption decision and usage behavior by exploring factors such as system characteristics and management practices, there may be a need to approach self-reports with caution. It is hoped that this paper has raised some valid concerns with regard to the survey method and suggested some useful ways to control for some of biases associated with the method. In the future, continued efforts in research

methods, not limited to survey, are needed to recommend best approaches to study emergent economic, social and management issues created by new technologies and applications.

Appendix

Appendix A: Instruments for Constructs of Interest in Study 1

Construct	Question Items
Perceived Ease of Use (PEOU)	<ol style="list-style-type: none"> 1. Learning to use the course blog would be easy for me. 2. My interaction with the course blog would be clear and understandable. 3. It would be easy for me to become skilful at using the course blog. 4. I would find the course blog easy to use.
Perceived Usefulness (PU)	<ol style="list-style-type: none"> 1. The course blog can help me learn the course materials more effectively. 2. The course blog can enhance my learning experience. 3. I can learn from the comments posted by instructors, tutors and students on the course blog. 4. The course blog can facilitate the sharing of knowledge among instructors, tutors and students.
Attitude (ATT)	<ol style="list-style-type: none"> 1. Using the course blog is a bad/good idea. 2. Using the course blog is a foolish/wise idea. 3. I dislike/like the idea of using the course blog. 4. Using the course blog would be unpleasant/pleasant.
Intention (INT)	<ol style="list-style-type: none"> 1. I intend to use the course blog in the future. 2. I predict I would use the course blog in the future. 3. I plan to use the course blog in the future. 4. I will use the course blog in the future.

Appendix B: Data Analysis for Subjects who Completed All Parts of Survey in Study 1

Table A1. Fisher-Bonett Test Results

	Constructs Measured without Temporal Separation		Constructs Measured with Temporal Separation	
	$H_0 : \alpha_3 < \alpha_1$	$H_0 : \alpha_3 < \alpha_2$	$H_0 : \alpha_6 < \alpha_4$	$H_0 : \alpha_6 < \alpha_5$
Perceived Ease of Use (PEOU)	0.075	0.171	0.029*	0.079
Perceived Usefulness (PU)	0.007**	0.002**	0.614	0.454
Attitude (ATT)	0.146	0.084	0.037*	0.152
Intention (INT)	0.044*	0.193	0.097	0.089

* Significant at the 0.05 level. ** Significant at the 0.01 level. *** Significant at the 0.001 level

Table A2. Tests for Equivalence of Correlations

	Group 1 (Logical order)	Group 4 (Logical order)	Sig.	Group 2 (Reverse logical order)	Group 5 (Reverse logical order)	Sig.
$r_{PEOU, ATT}$	0.670	0.272	0.061 ⁺	0.596	0.585	0.951
$r_{PU, ATT}$	0.666	0.541	0.485	0.819	0.529	0.042*
$r_{PEOU, INT}$	0.737	0.408	0.072 ⁺	0.605	0.282	0.138
$r_{PU, INT}$	0.693	0.651	0.787	0.689	0.250	0.033*
$r_{ATT, INT}$	0.829	0.632	0.121	0.678	0.349	0.096 ⁺

⁺ Significant at the 0.1 level. * Significant at the 0.05 level. ** Significant at the 0.01 level. *** Significant at the 0.001 level.

Appendix C: Instruments for Constructs of Interest in Study 2

Construct	Answered by	Question Items
Perceived Ease of Use (PEOU)	Group 1	1. My interaction with Hushmail would be clear and understandable 2. It would be easy for me to become skillful at using Hushmail. 3. I would find Hushmail easy to use.
Perceived Usefulness (PU)	Group 1	1. Using Hushmail would benefit me. 2. Hushmail would be useful to me. 3. I would find Hushmail useful.
Subjective Norm (SN)	Group 1	1. People who influence my behavior would support me using Hushmail. 2. People who are important to me would support me using Hushmail. 3. People who are important to me would agree that it is a good idea to use Hushmail.
Perceived Severity (PS)	Group 2	1. It is a very offensive invasion of my privacy if other people can read my personal email. 2. I will be very unhappy if other people can read my personal email. 3. I will be very much bothered if other people can read my personal email.
Perceived Vulnerability (PV)	Group 2	1. It is likely that some people will try to read my personal email without my knowledge. 2. It is likely that some people will try to gain unauthorized access to my personal email. 3. It is likely that some people will try to look into my email without my permission.
Response Efficacy (RE)	Group 2	1. Hushmail can protect my email privacy. 2. Hushmail can prevent unauthorized access to my personal email. 3. Hushmail can prevent others from reading my email without my knowledge.
Intention (INT)	Groups 1 & 2	1. I intend to use Hushmail in the future. 2. I plan to use Hushmail in the future. 3. I will use Hushmail in the future.

References

- Aaker, J. L. 2000. The Influence of Culture on Persuasion Processes and Attitudes: Diagnosticity or Accessibility? *Journal of Consumer Research* **26**(4) 340-357.
- Ahluwalia, R., Z.Gurhan-Canli. 2000. The Effects of Extensions on the Family Brand Name: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research* **27**(3) 371-381.
- Benbasat, I., H. Barki. 2007. Quo Vadis, TAM? *Journal of the Association for Information Systems* **8**(4) 211-218.
- Bertrand, M., S. Mullainathan. 2001. Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review* **91**(2) 67-72.
- Bogozzi, R. P., Y. Yi. 1990. Assessing Method Variance in Multitrait-multimethod Matrices: The Case of Self-reported Affect and Perception at Work. *Journal of Applied Psychology* **75**(5) 547-560.
- Bonett, D. G. 2002. Sample Size Requirements for Testing and Estimating Coefficient Alpha.” *Journal of Educational and Behavioral Statistics* **27**(4) 335-340.
- Bonett, D. G. 2003. Sample Size Requirements for Comparing Two Alpha Coefficients. *Applied Psychological Measurement* **27**(1) 72-74.
- Budd, R. J. 1987. Response Bias and the Theory of Reasoned Action. *Social Cognition* **5**(2), 95-107.
- Burton-Jones, A. 2009. Minimizing Method Bias through Programmatic Research. *MIS Quarterly* **33**(3) 445-471.
- Chandon, P., V. G. Morwitz, W. J. Reinartz. 2005. Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research. *Journal of Marketing* **69**(2) 1-14.
- Crampton, S. M., J. A. Wagner. 1994. Percept-percept Inflation in Microorganizational Research: An Investigation of Prevalence and Effects. *Journal of Applied Psychology* **79**(1) 67-76.
- Cohen, J., P. Cohen, S. G. West, L. S. Aiken. 2002. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd Edition, Psychology Press.
- Cote, J. A., R. Buckley. 1987. Estimating Trait, Method, and Error Variance: Generalizing across 70 Construct Validation Studies. *Journal of Marketing Research* **24**(3) 579-582.

- Davis, F. D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* **13**(3) 319-340.
- Davis, F. D., R. P. Bagozzi, P. R. Warshaw. 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science* **35**(8) 982-1003.
- Davis, F. D., V. Venkatesh. 1996. A Critical Assessment of Potential Measurement Biases in the Technology Acceptance Model: Three Experiments. *International Journal of Human-Computer Studies* **45**(1) 19-45.
- DeVellis, R. F. 2003. *Scale Development: Theory and Application*, 2nd Edition, Sage Publications.
- Doty, D. H., W. H. Glick. 1998. Common Methods Bias: Does Common Methods Variance Really Bias Results? *Organizational Research Methods* **1**(4) 374-406.
- Edwards, J. R., R. P. Bagozzi, 2000. On the Nature and Direction of Relationships between Constructs and Measures. *Psychological Methods* **5**(2) 155-174.
- Feldman, J. M., J. G. Lynch, Jr. 1988. Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior.” *Journal of Applied Psychology* **73**(3) 421-435.
- Fornell, C., D. F. Larcker. 1981. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research* **18**(3) 382-388.
- Harrison, D. A., M. E. McLaughlin. 1993. Cognitive Processes in Self-Report Responses: Tests of Item Context Effects in Work Attitude Measures. *Journal of Applied Psychology* **78**(1) 129-140.
- Kim, S., L. S. Feldt. 2008. A Comparison of Tests for Equality of Two or More Independent Alpha Coefficients. *Journal of Educational Measurement* **45**(2) 179-193.
- King, W. R., C. Z. Liu, M. H. Haney. 2007. Method Effects in IS Survey Research: An Assessment and Recommendations. *Communications of the Association for Information Systems* (20) 457-482.
- Lynch, J. G., Jr. 2005. Accessible but Nondiagnostic Memories about Memory and Consumer Choice. In Griffin, A., C. Otnes (eds.). *16th Paul D. Converse Symposium*. Chicago: American Marketing Association, 88-115.

- Lynch, J. G., Jr. 2006. Accessibility-Diagnosticity and the Multiple Pathway Anchoring and Adjustment Model. *Journal of Consumer Research* **33**(1) 25-27.
- Malhotra, N. K., S. S. Kim, A. Patil. 2006. Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. *Management Science* **57**(12) 1865-1883.
- Mathieson, K. 1991. Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior. *Information Systems Research* **2**(3) 173-191.
- Nunnally, J. C. 1967. *Psychometric theory*. New York: McGraw Hill.
- Podsakoff, P. M., S. B. MacKenzie, J.-Y. Lee, N. P. Podsakoff. 2003. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology* **88**(5) 879-903.
- Rogers, R. W. 1975. A Protection Motivation Theory of Fear Appeals and Attitude Change. *Journal of Psychology* **91**(1) 93-114.
- Schmitt, N. 1994. Method Bias: The Importance of Theory and Measurement. *Journal of Organizational Behavior* **15**(5) 393-398.
- Schwarz, N. 1999. Self-Reports: How the Questions Shape the Answers. *American Psychologist* **54**(2), 93-105.
- Sharma, R., P. Yetton, J. Crawford. 2009. Estimating the Effect of Common Method Variance: The Method-Method Pair Technique with an Illustration from TAM Research. *MIS Quarterly* **33**(3) 473-490.
- Sirken, M. G., S. Schechter. 1999. Interdisciplinary Survey Methods Research. In Sirken, M. G., D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, R. Tourangeau (eds.). *Cognition and Survey Research*, John Wiley & Sons, Inc.
- Spector, P. E. 1987. Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology* **72**(3) 438-443.
- Straub, D. W., A. Burton-Jones. 2007. Veni, Vidi, Vici: Breaking the TAM Logjam. *Journal of the Association for Information Systems* **8**(4) 223-229.

- Straub, D., M.-C. Boudreau, D. Gefen. 2004. Validation Guidelines for IS Positivist Research. *Communications of the AIS* (13) I380-427.
- Taylor, S., P. A. Todd. 1995. Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research* **6**(2) 144-176.
- Tourangeau, R., K. A. Rasinski. 1988. Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin* **103**(3) 299-314.
- Tourangeau, R., K. A. Rasinski, N. Bradburn, R. D'andrade. 1989. Carryover Effects In Attitude Surveys. *Public Opinion Quarterly* **53**(4) 495-524.
- Tourangeau, R., L. J. Rips, K. Rasinski. 2000. *The Psychology of Survey Response*, Cambridge University Press, New York.
- Venkatesh, V., F. D. Davis. 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science* **45**(2) 186-204.
- Venkatesh, V., M. G. Morris, G. B. Davis, F. D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* **27**(3) 425-478.
- Williams, L. R., B. K. Brown. 1994. Method Variance in Organizational Behavior and Human Resources Research: Effects on Correlations, Path Coefficients, and Hypothesis Testing. *Organizational Behavior and Human Decision Processes* **57**(2) 185-209.
- Williams, L. R., J. A. Cote, M. R. Buckley. 1989. Lack of Method Variance in Self-reported Affect and Perceptions at Work: Reality or Artifact? *Journal of Applied Psychology* **74**(3) 462-468.