A Call for Item-ordering Transparency in Online IS Survey Administration

Completed Research Paper

E. Vance Wilson
Worcester Polytechnic Institute
vancewilson@gmail.com

Mark Srite
University of Wisconsin-Milwaukee
msrite @uwm.edu

Eleanor Loiacono
Worcester Polytechnic Institute
eloiacon@wpi.edu

Abstract

Online survey applications typically offer the capability to individually randomize the order in which survey items are presented to subjects, a method that structurally eliminates several sources of method bias inherent to static surveys. IS researchers who use online surveys have a strong interest in knowing how prior surveys were administered in published research, however, we find this information is rarely available in current practice. This paper presents a call for increased transparency in reporting item-ordering methodology in future online IS survey research. This call is based on 1) a literature review of online survey research published in the AIS Senior Scholars' Basket of Journals, 2) results of new research comparing reliability and construct validity characteristics produced by individually-randomized vs. static survey administration methods, and 3) results of hypothetical structural equation modeling (SEM) analyses contrasting structural models following purification of the individually-randomized and static datasets.

Keywords

Openness, context effects, survey design, systematic bias, replicability.

Introduction

Survey item-ordering once was an area of methodological interest to IS researchers, yet in recent times this topic has been largely ignored. Two factors motivate a renewed focus on item-ordering. First, evidence has mounted that item-ordering is a key source of method bias that can introduce systematic errors into survey results (Edwards, 2008; Podsakoff et al., 2003; Podsakoff et al., 2012; Straub et al., 2004). Ignoring method bias can lead to acceptance of poor theories and development of flawed practical guidelines (Burton-Jones, 2009). Second, technological advances in online surveys have introduced new capabilities that overcome intractable obstacles associated with item-ordering in traditional pencil-and-paper surveys. These advances provide opportunities for IS researchers to use online capabilities not only to reduce costs, extend reach, and improve timeliness of survey research (Fricker and Schonlau, 2002), but also to enhance the rigor of survey results.

Our research comprehensively compares the effects of two distinct item-ordering approaches in the context of online IS survey research (see Figure 1). In one approach—made possible by the emergence of online survey applications—item order is individually-randomized for each subject.

In the alternate approach, survey items are presented to multiple subjects in the same static ordering. Researchers can apply a static approach in two distinct ways: (1) items that measure the same construct can be grouped so that they are presented adjacent to one another or (2) items can be intermixed through one-time randomization or other scheme so that those representing the same construct are spread throughout the survey (hereafter these are described respectively as grouped-static and intermixed-static approaches). Both static approaches were developed in the era of pencil-and-paper surveys, when individual randomization was difficult to operationalize. We focus in this study on grouped-static designs.
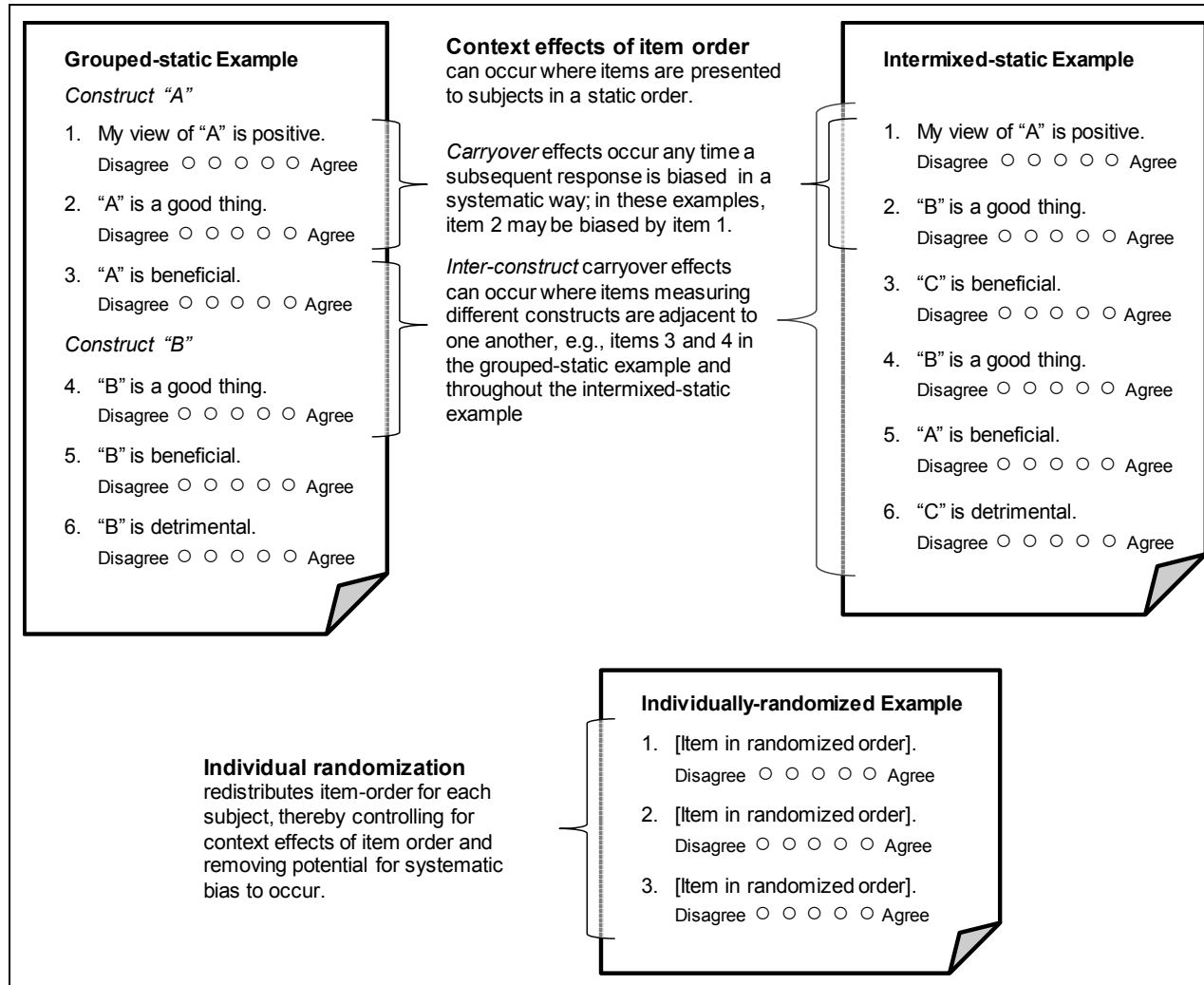
Figure 1. Sources of Context Effects of Item Order

Current Status of Item-order Reporting in Online IS Surveys

In order to gain more information on the status of online surveys and item-ordering practice in IS research, we reviewed the AIS Senior Scholars' Basket of Journals during the period 2010-2014. Our review identified 158 articles in which survey research was conducted using constructs that are represented by multiple survey items and scaled response measures, such as Likert scales. These are key attributes of studies that are designed for analysis via factor analysis or SEM methods. We found that 88 (56%) of the articles used online surveys for all or part of the reported research.

We did not identify obstacles in any of these 88 articles that would have prevented implementation of individually-randomized item-ordering in their survey designs. Yet we found only three (3.4%) in which any form of randomization or scrambling of item-ordering was reported (D'Arcy et al., 2014; Hoffman et al., 2014; Mathew and Chen, 2013). Further, none of these three articles reported whether randomization was individually performed or was part of an intermixed-static design.

We also found no articles that reported using static item-ordering, although this approach was implied in four articles—all describing grouped-static designs—where the same survey was administered in both paper and online forms (4.5%). In the remaining articles no information regarding item-ordering was available to readers, despite ubiquitous reporting of other methodological details. For example, all articles

reported the number of subjects that were surveyed, and 81 (92%) of the articles explicitly reported the scale format used for measurement (e.g., seven-position Likert scale).

The lack of transparency represented by this current situation presents a serious obstacle to IS researchers seeking to accurately replicate study findings, an issue that has become prominent in IS as well as other disciplines (Dennis and Valacich, 2014; Pashler and Wagenmakers, 2012). Having identified that item-ordering methodology is rarely reported in online IS survey research, we turn our attention to what we perceive as problems underlying use of static item-ordering.

Background

Virtually the entire early literature on item-ordering focused on static surveys administered using pencil-and-paper methods (Tourangeau et al., 2000). Researchers were concerned about the potential for subjects to provide biased responses to items based on other items they viewed earlier in the same survey, a phenomenon described as *context effects* that arise through semantic mechanisms (Krosnick and Presser, 2010; Schwarz and Sudman, 2012). Tourangeau et al. (2000, p. 197) explain that

> "the judgments called for by attitude questions are rarely absolute but are typically made in relation to some standard, generally an implicit one. It is hardly surprising, then, that attitude judgements turn out to be quite context-dependent. As survey researchers have demonstrated repeatedly, the same question often produces quite different answers, depending on the context.

It was reported decades ago that significant trade-offs exist in choosing among static ordering schemes in the attempt to avoid systematic bias arising from context effects (Harrison and McLaughlin, 1988; Kraut et al., 1975; Schuman and Presser, 1981). Yet identifying the most appropriate trade-off has proved contentious. Historically, two camps arose among researchers regarding the preferred form for item-ordering in the static survey designs of the era, one arguing for grouped-static ordering and the other arguing for intermixed-static ordering (Budd, 1987; Davis and Venkatesh 1996; Goodhue and Loiacono, 2002; Schriesheim et al., 1989).

More recently, important structural changes have occurred in the survey research landscape. Where Davis and Venkatesh (1996) and Goodhue and Loiacono (2002) studied pencil-and-paper designs, surveys in IS research are now increasingly administered online where, instead of intermixing items within one or a few versions, item-ordering can be programmatically randomized for each participant without adding to workload for survey administrators or data analysts.

The capability to individually randomize administration item-ordering is found in major commercial online survey applications including Qualtrics, SnapSurveys, SurveyGizmo, and SurveyMonkey. SurveyGizmo introduces its RandomizeQuestions capability in this way:

> "Survey researchers frequently use randomization as a tool to combat survey bias. Randomizing the order of questions, pages, and/or answer options in your survey prevents bias introduced by order and/or survey fatigue." (https://help.surveygizmo.com/help/randomize-questions)

Despite the logical appeal of this explanation, it is principally based on anecdotal evidence. Although a substantial literature studies the effects of item randomization leading to intermixed-static survey designs, we find little empirical research that compares individually-randomized designs to static designs, despite several papers that promote benefits of using individual randomization (Bradlow and Fitzsimons, 2001; Raghunathan and Grizzle, 1995; Weinberger et al, 2006).

Our approach in the present study is grounded in research by Wilson and Lankton (2012), who conducted an exploratory item-ordering study using an online IS survey. In addition to providing insights regarding empirical effects of individually-randomized item-ordering, their study identifies several important issues that have not received much prior research, centering on an unexpectedly high level of item response anomalies they encountered in the grouped-static condition of their study.

Item Response Anomalies

An anomaly describes *something that deviates from what is standard, normal, or expected.* In the Wilson and Lankton (2012) study, anomalies of several distinct types appeared with disproportionate frequency when subjects' responses were compared between individually-randomized and grouped-static

conditions. Their conditions were implemented as alternate versions of an online survey that assessed multi-item measures of IS constructs related to technology adoption.

In the absence of bias, we can expect that the mean numeric response to a given survey item will vary between conditions as dictated by the central limit theorem; i.e., in large samples approximately five percent of mean scores will vary significantly between conditions, given a 0.05 probability level. Yet Wilson and Lankton (2012) reported that 30% of item responses were significantly different between individually-randomized and grouped-static conditions in their study. Their interpretation was that these anomalies arose from context effects of item order that occurred in the response process. Because individually-randomized item-ordering eliminates structural conditions for such context effects to occur, their interpretation was that such an excessive rate of item response anomalies is evidence of systematic bias in the grouped-static survey condition.

Wilson and Lankton (2012) reported two major context effects. They found numerous *carryover effects* that take the form of increased consistency or contrast between responses when people are asked to form a judgment about some item following previous stimuli (Sudman et al., 1996). Some of these were *inter-scale effects*, a form of carryover that occurs as subjects transition between items representing different scales in static surveys. Because inter-scale effects arise between conceptually unrelated items, they represent a form of carryover effects that may exist in intermixed-static surveys as well as grouped-static surveys (see Figure 1).

Research Questions

The findings of Wilson and Lankton (2012) raise a troubling specter, yet we recognize that results from their single study are not sufficient to drive wholesale changes in online survey practice among IS researchers. This recognition leads to an overarching question:

Can the findings of Wilson and Lankton (2012) regarding item response anomalies be corroborated by research that employs a different set of items in a new study population?

Assuming the answer is affirmative, we propose it will be important to quantify not only the incidence of item response anomalies but to also identify what effects, if any, these anomalies have on consequential research outcomes. This proposition leads to our research questions.

What are the effects of individually-randomized vs. grouped-static survey administration on:

- *RQ1: Prevalence of item response anomalies?*
- *RQ2: Scale reliability?*
- *RQ3: Convergent and discriminant validity?*
- *RQ4: Structural models produced through purification of the resulting datasets?*

We describe how we addressed these research questions in the following sections.

Research Method

The research was conducted among business students at a large Midwest US university. Our survey addressed students' perceptions of trust relating to their university's Learning Management System (LMS), which provided them access to course materials, assignments, recorded lectures, grading information, and online communication services. Measures used in this study were a close adaptation to the LMS context of the comprehensive trust instrument developed by McKnight et al. (2002).

Table 1 shows the order in which grouped-static items were administered, as well as an example item from each construct. We collected responses using five-position Likert scales marked as "Strongly Agree", "Agree", "Neutral", "Disagree", and "Strongly Disagree", coded respectively as 1 through 5. In addition, subjects could select a "Not Applicable" response option.

A Qualtrics online survey was used to implement this research. Subjects were randomly assigned to a research condition, notified of their rights and responsibilities regarding participation in the research, and then presented with several survey items on each screen. Their responses to these items were stored prior to advancing to the next screen.

All subjects responded to the same items, the only difference being the order in which items were presented to them, i.e., grouped-static or individually-randomized. No construct labels were used to identify items in any of the conditions. When subjects failed to complete an item, they were prompted to do so before continuing to the next screen. If subjects disconnected during the process of completing either survey, the application would return them on re-entry to the position following the last point in the survey that they had previously completed.

In total, 171 subjects (74 male, 97 female, 22 years mean age) completed the individually-randomized condition and 174 (80 male, 94 female, 21 years mean age) completed the grouped-static condition. For analysis via one-way ANOVA, these sample sizes have statistical power greater than the 0.80 level recommended by Cohen (1992) to identify medium-size effects.

Results

Our results generally corroborate reports by Wilson and Lankton (2012). Results are organized in the following sections by order of the research questions we raised previously in the paper.

RQ1: Item Response Anomalies

SPSS one-way ANOVA was performed to compare item means between individually-randomized and grouped-static conditions[1]. Differences in these values may be interpreted as evidence of item-ordering effects, given that only item order varied between the conditions (Bickart, 1993; Weijters et al., 2009).

We found eight instances of carryover (13% of all items) and three instances of inter-scale effects (18% of transitions between construct scales). When item scores were summated to form simple construct measures, we found these varied significantly between the survey versions for Institution-based Benevolence and Personal Innovativeness constructs (12% of constructs). These differences occur more frequently than expected between items ($z = 2.60$, $p = 0.005$) and construct measures ($z = 2.27$, $p = 0.012$). Our findings regarding item response anomalies corroborate Wilson and Lankton (2012).

RQ2: Scale Reliability

Cronbach's alpha was calculated separately for each construct using SPSS. Composite reliability (CR) was calculated using AMOS confirmatory factor analysis (CFA) in which all constructs within the target study were simultaneously evaluated. Alpha and CR were numerically higher for every grouped-static scale when compared to the corresponding individually-randomized scale, the difference averaging approximately 23% for each reliability measure. Significance testing of reliability statistics between conditions (Feldt, 1969) showed differences were significant in 14 of the 17 measured constructs using alpha statistics and in 12 of 17 using CR statistics. Wilson and Lankton (2012) did not report CR statistics; however, our findings regarding alpha reliability do corroborate their findings.

RQ3: Convergent and Discriminant Validity

Assessment of convergent validity and discriminant validity was conducted through CFA. Grouped-static administration increased loadings on theorized constructs by an overall average of 31%, with increases ranging from 18% (Disposition to Trust) to 41% (Trusting Beliefs).

---

[1] The complete survey instrument and detailed versions of analyses summarized in this paper are available from the lead author on request.

Trust Instrument Area	Construct	Items	Example Item*
Disposition to Trust	Benevolence (DB)	3	In general, people really do care about the well-being of others.
	Integrity (DI)	3	In general, most folks keep their promises.
	Competence (DC)	3	I believe that most professional people do a very good job at their work.
	Trusting Stance (TS)	3	I usually trust people until they give me a reason not to trust them.
Institution-based Trust	General (IG)	2	I am comfortable using the University's network.
	Benevolence (IB)	3	I feel that most University faculty/staff would act in a student's best interest.
	Integrity (II)	3	I am comfortable relying on University faculty/staff to meet their obligations.
	Competence (IC)	3	I feel that most University faculty/staff are good at what they do.
	Structural Assurance (ISA)	4	The University's network has enough safeguards to make me feel comfortable using it.
Trusting Beliefs	Benevolence (TBB)	3	I believe that the [LMS] system would act in my best interest.
	Integrity (TBI)	4	I would characterize the [LMS] system as honest.
	Competence (TBC)	4	The [LMS] system is competent and effective in its work.
Trusting Intentions	General (GN)	4	I can always rely on [LMS] in a tough situation.
	Follow Advice (FA)	5	I would feel comfortable acting on information given to me by [LMS].
Personal Innovativeness (PII)		5	I like to explore new Web sites.
General Web Experience (GWE)		4	On average, how much time per week to you spend on reading newspapers on the Web?
Perceived Site Quality (PSQ)		5	Overall, [LMS] works very well technically.

* The term "[LMS]" was replaced with the name of the university LMS during actual survey administration.

Table 1. Overview of the Trust Instrument

Summarized CFA validation statistics are presented in Table 2 for the four trust instrument areas that comprise multiple constructs. In each area, discriminant validity violations were identified in the individually-randomized condition that do not appear in the grouped-static condition, and this was also the case for convergent validity violations in three of four areas. Only 18% of the validity violations identified in the individually-randomized condition emerged in the grouped-static condition (11 vs. 2 violations). These findings suggest that grouped-static item-ordering inflates measures of convergent and discriminant validities, thereby masking validity issues for individual items and, potentially, for entire constructs. Wilson and Lankton (2012) did not address construct validation in their study.

RQ4: A Hypothetical SEM Evaluation of the Purified Datasets

In our final analysis, we conducted purification of the two datasets produced in this study and then followed with AMOS SEM analysis. The objective of this exercise is to answer the question, *"What would the difference be in study outcomes if researchers chose to apply one item-ordering approach vs. the other?"* Although conducted as a hypothetical exercise, we propose that answering this question is important as it completes the research cycle by putting survey data to its intended end use.

In developing their trust instrument, McKnight et al. (2002) theorized that Personal Innovativeness is predicted by the constructs comprising Disposition to Trust, which can be modeled as a set of first-order relationships among latent factors. We evaluated this nomological model separately for both the

individual-randomized and grouped-static datasets. Our first step in this process was to purify each dataset (Churchill, 1979; MacKenzie, Podsakoff, and Podsakoff, 2011). We applied these commonly used scale purification criteria:

- Items with low loadings are removed to achieve convergent validity and reliability of constructs, i.e., removal occurs where CR < 0.700 and/or AVE < 0.500;

- items are removed to achieve discriminant validity among constructs, i.e., removal occurs where the square root of AVE is less than the intercorrelation with any other construct; and

- where these prior steps do not resolve convergent and/or discriminant validity issues or only one item remains to represent the construct, the construct is removed from subsequent analysis.

The purification process utilized CFA, beginning with data summarized in Table 2 and continuing iteratively until the above criteria were met. At that point, separate structural models were created and run for the individually-randomized and grouped-static datasets, producing results shown in Table 3.

Six items were removed from the individually-randomized model, completely eliminating the Competence and Integrity constructs of the Disposition to Trust area. No items were removed from grouped-static models. Following purification, explained variance was not significantly different between models, but all model fit statistics for the individually-randomized dataset were superior to statistics from the grouped-static dataset. Results from this hypothetical exercise show that artificial gains in reliability and validation do not necessarily translate into improved SEM results for grouped-static surveys.

Discussion

We find strong support for the contention by Wilson and Lankton (2012) that item-ordering is an important methodological issue in our field, yet our literature review found that there is virtually no reporting of item-ordering methodology by online IS survey researchers. Perhaps the majority of current IS researchers are choosing to individually randomize item-order in their online surveys. Yet even if this is the case, we argue that it is essential for researchers to report their methodological choice so this information is explicitly available for replicative studies. Certainly, this information is at least as important to subsequent researchers as the scale used to collect responses, which was reported by over 90% of researchers in our literature review.

Item Response Anomalies and the Inflation of Reliability and Validity Statistics

Reliability statistics were numerically higher for every grouped-static scale in our study. Similar to prior authors (Goodhue and Loiacono, 2002; Wilson and Lankton, 2012), we interpret inflation of reliability statistics to stem from the artificial similarity created by proximity of items rather than from subjects' careful evaluation of item content. Although some researchers may consider the heightened reliability statistics produced by grouped-static surveys (averaging 23% higher in our study) to be innocuous and reassuring, we observed a significant cost in the form of false positives that were produced using this approach. In six of the 17 constructs we studied (35%), either Cronbach's alpha, CR, or both statistics were lower than .70 (poor-to-questionable) in the individually-randomized condition. Yet the inflated reliability statistics in the corresponding grouped-static condition gave these scales the appearance of meeting acceptable or good reliability criteria.

As previously reported by Wilson and Lankton (2012) we also found excessive presence of inter-construct context effects, suggesting that adjacent items do not have to be conceptually related in order for item anomalies to occur. Future research should explicitly study effects related to intermixed designs, especially where construct items are intentionally separated, as recommended by Hui (2012) and Weijters et al. (2009).

Trust Instrument Area		Constructs in this Area	Average Item Loading in CFA	Validity Violations		
				Convergent Validity	Discriminant Validity	Constructs with Violations*
Disposition to Trust	Randomized	4	.657	3	1	3
	Grouped-Static	4	.762	0	0	0
Institution-Based Trust	Randomized	5	.643	4	4	5
	Grouped-Static	5	.825	0	0	0
Trusting Beliefs	Randomized	3	.654	3	3	3
	Grouped-Static	3	.897	0	0	0
Trusting Intentions	Randomized	2	.611	2	2	2
	Grouped-Static	2	.810	2	0	2

* Overall number of constructs found to have convergent and/or discriminant validity violations

Table 2. CFA Validation Summary Statistics

Model: Disposition to Trust constructs predict Personal Innovativeness (PII)	Items Remaining Following Purification	R^2 of PII Prediction in Final Model	Significance of Difference in R^2	Final Model Fit Statistics*	
Individually-Randomized	DB1 TS1 DB2 TS2 DB3 TS3	.081	F = 1.12, p = 0.205, Effect size (f^2) = 0.007	X^2/df GFI AGFI NFI CFI RMSEA	1.403 0.944 0.910 0.892 0.965 0.049
Grouped-Static	DB1 DI1 DB2 DI2 DB3 DI3 DC1 TS1 DC2 TS2 DC3 TS3	.087		X^2/df GFI AGFI NFI CFI RMSEA	1.725 0.888 0.843 0.855 0.932 0.065

Table 3. SEM Analysis of Structural Models Developed through Purification of Datasets

Few prior item-ordering studies have addressed effects on construct validity, and none of these tested individually-randomized surveys. The CFAs we conducted show grouped-static factor loadings to be increased by more than 25% over the individually-randomized loadings. As with reliability measures, we interpret this finding to result from effects of item proximity in the grouped-static condition rather than item content. By this interpretation, the summarized results in Table 2 illustrate how grouped-static item-ordering can effectively mask violations of convergent and discriminant validity. We propose that the IS research community would be better served if marginal scales were refined or discarded rather than being retained in research designs on the basis of inflated reliability and construct validity statistics produced by grouped-static item-ordering.

Effects on SEM Analysis

We did not find any prior research that investigated effects of item-ordering on results of subsequent SEM analysis, however, we propose such research can provide useful insights in weighing the costs and benefits of the alternative methods. In creating our hypothetical SEM evaluation, we developed two separate purified datasets as if we were planning typical analyses. Our findings suggest that the advantages of inflated reliability and construct validity that researchers may envision for grouped-static item-ordering can dissipate during the overall process of SEM analysis. In the end, we found the individually-randomized structural model was similarly predictive to the grouped-static model. Furthermore, the individually-randomized model exhibited better fit characteristics and was more parsimonious.

Conclusion

Advances in technology often necessitate rethinking of established practices. In this paper, we focused on technological support for individual randomization of item-order in online surveys. Although this issue may seem inconsequential and mundane to some, we found that use of grouped-static item-ordering methods can undermine rigor of IS research in ways that are difficult to detect or to control for.

In sum, we propose there is now sufficient evidence to call on IS journal and conference editors to promote transparency by encouraging authors to report their item-ordering methodology in all future manuscripts that use online survey methods.

References

Bickart, B. A. 1993. "Carryover and Backfire Effects in Marketing Research," *Journal of Marketing Research* (30:1), pp. 52-62.

Bradlow, E. T., and Fitzsimons, G. J. 2001. "Subscale Distance and Item Clustering Effects in Self-Administered Surveys: A New Metric," J*ournal of Marketing Research* (38:2), pp. 234-261.

Budd, R. J. 1987. "Response Bias and the Theory of Reasoned Action," *Social Cognition* (5), pp. 95-107.

Burton-Jones, A. 2009. "Minimizing Method Bias through Programmatic Research," *MIS Quarterly* (33:3), pp. 445-471.

Churchill Jr, G. A. 1979. "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* (16), pp. 64-73.

Cohen, J. 1992. "A Power Primer," *Psychological Bulletin* (112:1), pp. 155-159.

D'Arcy, J., Herath, T., and Shoss, M. K. 2014. "Understanding Employee Responses to Stressful Information Security Requirements: A Coping Perspective," *JMIS* (31:2), pp. 285-318.

Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319–340.

Davis, F. D., and Venkatesh, V. 1996. "A Critical Assessment of Potential Measurement Biases in the Technology Acceptance Model: Three Experiments," *International Journal of Human-Computer Studies* (45), pp. 19-45.

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1992. "Extrinsic and Intrinsic Motivation to Use Computers in the Workplace," *Journal of Applied Social Psychology* (22), pp. 1111–32.

Dennis, A. R., and Valacich, J. S. 2014. "A replication manifesto," *AIS Transactions on Replication Research,* (1:1), pp. 1-4.

Edwards, J. R. 2008. "To Prosper, Organizational Psychology Should... Overcome Methodological Barriers to Progress," *Journal of Organizational Behavior* (29:4), pp. 469-491.

Feldt, L. S. 1969. "A Test of the Hypothesis that Cronbach's Alpha or Kuder-Richardson Coefficient Twenty Is the Same for Two Tests," *Psychometrika* (34), pp. 363-373.

Fricker, R. D., and Schonlau, M. 2002. "Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature," *Field Methods* (14:4), pp. 347-367.

Goodhue, D. L., and Loiacono, E. T., 2002. "Randomizing Survey Question Order vs. Grouping Questions by Construct: An Empirical Test of the Impact on Apparent Reliabilities and Links to Related Constructs," in *Proceedings of the 35th Hawaii International Conference on System Sciences*.

Harrison, D. A., and McLaughlin, M. E. 1993. "Cognitive Processes in Self-report Responses: Tests of Item Context Effects in Work Attitude Measures," *Journal of Applied Psychology* (78:1), pp. 129-140.

Hoffmann, C. P., Lutz, C., and Meckel, M. 2014. "Digital Natives or Digital Immigrants? The Impact of User Characteristics on Online Trust." *Journal of Management Information Systems* (31:3), pp. 138-171.

Hui, W. 2012. "Item Context Effects in IS Survey Research," *Pacific Asia Journal of the Association for Information Systems* (4:4).

Kraut, A. I., Wolfson, A. D., and Rothenberg, A. 1975. "Some Effects of Position on Opinion Survey Items," *Journal of Applied Psychology* (60:6), pp. 774-776.

Krosnick, J. A., and Presser, S. 2010. "Question and Questionnaire Design," Handbook of Survey Research (2), pp. 263-314.

Mathew, S. K., and Chen, Y. 2013. "Achieving Offshore Software Development Success: An Empirical Analysis of Risk Mitigation through Relational Norms," *Journal of Strategic Information Systems* (22:4), pp. 298-314.

MacKenzie, S. B., Podsakoff, P. M., and Podsakoff, N. P. 2011. "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly* (35:2), pp. 293-334.

McKnight, D. H., Choudhury, V., and Kacmar, C. 2002. "Developing and Validating Trust Measures for E-commerce: An Integrative Typology," *Information Systems Research* (13:3), pp. 334-359.

Pashler, H., and Wagenmakers, E. J. 2012. "Editors' Introduction to the Special Section on Replicability in Psychological Science A Crisis of Confidence?" *Perspectives on Psychological Science* (7:6), pp. 528-530.

Podsakoff, P. M., MacKenzie, S. B., and Podsakoff, N. P. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It," *Annual Review of Psychology* (63), pp. 539-569.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology* (88:5), pp. 879-903.

Raghunathan, T. E., and Grizzle, J. E. 1995. "A Split Questionnaire Survey Design," *Journal of the American Statistical Association* (90:429), pp. 54-63.

Schriesheim, C. A., Solomon, E., and Kopelman, R. E. 1989. "Grouped Versus Randomized Format: An Investigation of Scale Convergent and Discriminant Validity using LISREL Confirmatory Factor Analysis," *Applied Psychological Measurement* (13), pp. 19-31.

Schuman, H., and Presser, S. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Working, and Content.* New York: Academic Press.

Schwarz, N., and Sudman, S. 2012. *Context Effects in Social and Psychological Research,* New York: Springer-Verlag.

Smith, P. C. 1969. *The Measurement of Satisfaction in Work and Retirement: A Strategy for the Study of Attitudes,* Chicago, IL: Rand McNally and Company.

Straub, D., Boudreau, M-C., and Gefen, D. 2004. "Validation Guidelines for IS Positivist Research," *Communications of the Association for Information Systems* (13:24).

Sudman, S., Bradburn, N. M., and Schwarz, N. 1996. Thinking about Answers: The Application of Cognitive Processes to Survey Methodology, San Francisco: Jossey-Bass.

Tourangeau, R., and Rasinski, K. A. 1988. "Cognitive "Processes Underlying Context Effects in Attitude Measurement," *Psychological Bulletin* (103), pp. 299-314.

Tourangeau, R., Rasinski, K. A., Bradburn, N., and D'Andrade R. 1989. "Carryover Effects in Attitude Surveys," *Public Opinion Quarterly* (53), pp. 495-524.

Tourangeau, R., Rips, J., and Rasinski, K. A., 2000. *The Psychology of Survey Response,* Cambridge, UK: Cambridge University Press.

Tourangeau, R., Singer, E., and Presser, S. 2003. "Context Effects in Attitude Surveys Effects on Remote Items and Impact on Predictive Validity," *Sociological Methods & Research* (31:4), pp. 486-513.

Weijters, B., Geuens, M., and Schillewaert, N. 2009. "The Proximity Effect: The Role of Inter-item Distance on Reverse-item bias," *International Journal of Research in Marketing* (26), pp. 2-12.

Weinberger, A. H., Darkes, J., Del Boca, F. K., Greenbaum, P. E., and Goldman, M. S. 2006. "Items as Context: Effects of Item Order and Ambiguity on Factor Structure," *Basic and Applied Social Psychology* (28:1), pp. 17–26

Wilson, E. V. and Lankton, N. K. 2012. "Some Unfortunate Consequences of Non-randomized, Grouped-item Survey Administration in IS Research," In *Proceedings of the 2012 International Conference on Information Systems* (ICIS), Orlando, FL.