

Association for Information Systems

## AIS Electronic Library (AISeL)

---

Proceedings of the 2018 AIS SIGED  
International Conference on Information  
Systems Education and Research

SIGED: IAIM Conference

---

2018

### ON THE IMPORTANCE OF DATA QUALITY IN INFORMATION SYSTEMS RESEARCH AND PH.D. CURRICULA

James Burleson

*Department of Management, HR and Information Systems Orfalea College of Business California  
Polytechnic State University, jburleso@calpoly.edu*

Michelle Carter

*Department of Management, Information Systems, and Entrepreneurship Carson College of Business  
Washington State University*

Jalal Sarabadani

*Department of Management, Information Systems, and Entrepreneurship Carson College of Business  
Washington State University*

Follow this and additional works at: <https://aisel.aisnet.org/siged2018>

---

#### Recommended Citation

Burleson, James; Carter, Michelle; and Sarabadani, Jalal, "ON THE IMPORTANCE OF DATA QUALITY IN INFORMATION SYSTEMS RESEARCH AND PH.D. CURRICULA" (2018). *Proceedings of the 2018 AIS SIGED International Conference on Information Systems Education and Research*. 19.  
<https://aisel.aisnet.org/siged2018/19>

This material is brought to you by the SIGED: IAIM Conference at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2018 AIS SIGED International Conference on Information Systems Education and Research by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# ON THE IMPORTANCE OF DATA QUALITY IN INFORMATION SYSTEMS RESEARCH AND PH.D. CURRICULA

*Research-in-Progress*

James Burleson  
Department of Management, HR and Information Systems  
Orfalea College of Business  
California Polytechnic State University  
jburleso@calpoly.edu

Michelle Carter  
Department of Management, Information Systems, and Entrepreneurship  
Carson College of Business  
Washington State University

Jalal Sarabadani  
Department of Management, Information Systems, and Entrepreneurship  
Carson College of Business  
Washington State University

## Abstract:

Data quality procedures are vital in conducting survey research, yet they are under-emphasized in information systems (IS) Ph.D. curricula and published journal articles. In this research-in-progress, we offer the “5-C Framework” to evaluate the current state of IS survey research, as it pertains to data quality, provide insight into where IS Ph.D. curricula may be lacking, and offer a basis for developing new curricula that address those gaps. In pursuit of our objectives, we present preliminary findings from our analysis of IS survey research between 2008 and 2017. This work should interest those developing IS Ph.D. curricula. Establishing guidelines, based on the 5-C framework, can aid educators in teaching Ph.D. students how to enact and communicate data quality procedures effectively.

Keywords: **data quality, Ph.D. curricula, survey research**

## I. INTRODUCTION

*Jane is a doctoral candidate in her fourth year of the Information Systems Ph.D. program. She collected data for her dissertation via an online survey to 500 full-time employees. Although Jane successfully completed all her statistics and methods courses, opening the results of her survey gives her pause. Jane knows all about the importance of data validity and reliability, but she has no idea what to do with the messy data set that confronts her. With the pressure to complete her analysis in time to graduate before the end of the academic year, Jane is filled with anxiety. She is supposed to be competent, and she has been well-trained; how is it possible that she does not know how to proceed?*

The preceding scenario is not uncommon for doctoral students in today's academic environment. While students recognize the need for data quality, the procedures for ensuring such quality are often elusive or shrouded in mystery. Lack of experience or knowledge of how to proceed fosters anxiety in students and can lead to unnecessary errors during analysis. Unfortunately, our field provides little guidance in the way of published discussions regarding how best to enact data quality procedures in survey research.

Surveying is a prevalent methodology in Information Systems (IS) research [Compeau et al., 2012]. Despite the methodology's numerous benefits, including the relatively short time spent in data analysis and ease of access to subjects, data gathered from surveys often requires cleanup work before further analyses can be performed. Put simply, survey data is often messy. The use of human subjects in survey research poses potential threats to data quality [DeSimone and Harms,

2017; DeSimone et al., 2015]. Survey data sets are often prone to challenges of validity due to missing responses, inattentive subjects, skewed results, etc.

As a means of highlighting and alleviating the problems associated with data quality in our field, this study seeks to answer two key questions. First, ***what is the current state of data quality procedures in survey research?*** Second, ***what considerations need to be addressed with regard to data quality to ensure the highest validity of results in survey research?***

For the first question, we aim to draw attention to the areas where our field inadequately enacts and/or communicates data quality procedures. Doing so will demonstrate where our curricula are potentially lacking and accentuate the difficulty students face in learning from current research. The second question, we aim to provide guidance regarding best practices in enacting data quality procedures and in communicating those procedures in published research. The importance of both enacting and communicating data quality procedures is clear.

To ensure the highest degree of confidence in survey research, researchers conducting survey methodologies must enact data quality procedures. Not only is enacting of data quality procedures important, so too is the selection of *well-founded* procedures. Research has been conducted for decades regarding how best to ensure the highest quality data sets from surveys [Groves, 1987]. Researchers bear the responsibility for incorporating the best-established procedures into their methodologies.

Communicating data quality procedures used in research is also essential. Whereas enacting data quality procedures impacts the *actual* quality of data used in hypothesis testing, communicating those procedures impacts *perceived* data quality, thereby influencing the confidence in a study. Data quality procedures that are not communicated may be assumed to have been ignored, thus reducing the perceived validity of analysis.

Communicating data quality procedures also offers guidance for others to follow. Lacking such guidance, students must seek their own means of ensuring data quality, which can lead to confusion and errors. Additionally, communicating data quality, especially in a relatively standardized format, eases the process of assessing studies, especially for journals and conferences. A standardized procedure for communicating data quality procedures not only increases the perceived validity of the study, it also reduces the effort required in determining whether necessary procedures were followed.

This study should provide insight into where our current IS PhD curricula may be lacking and provide a framework for developing new curricula to fill in those gaps. As a first step in addressing our research questions, we delineate the data quality procedures that ought to be followed in survey research.

## II. DATA QUALITY PROCEDURES – THE “5-C FRAMEWORK”

One of the greatest challenges in teaching data quality procedures to Ph.D. students is identifying the entire domain of data quality in survey research. Thus, before we can provide guidance regarding *how* to ensure that students learn best practices, we must first establish the purview of what we mean when we discuss data quality.

In this section, we categorize and explain the core procedures that researchers should consider to ensure the highest quality of data in survey methodologies. For each category, we describe the domain of procedures and the importance of their enactment. The naming convention offered was developed to provide a simple mnemonic device to aid memory of the categories and associated actions. Using this framework, we seek to provide guidance regarding the core aspects of data quality and to offer standardized procedures that PhD students will be able to refer to and follow when conducting survey research.

## Categories of Data Quality Procedures

Table 1 contains the five categories which comprise the framework. For each, a general overview of procedures is provided, along with a discussion of the category's importance.

Table 1. Categories of Data Quality Procedures

Category	Definition
Correspondence	the degree to which all data in a data set fall within the intended sample frame for the study
Completeness	the degree to which all survey questions are fully answered by all respondents
Carelessness	the degree to which all respondents were attentive and engaged while taking the survey
Composition	the dispersion of the data after the survey is completed, specifically as relates to the normality of the data and the presence of any outliers
Credibility	the degree to which the results of a survey are unnecessarily biased

### Correspondence

The correspondence of a data set pertains to the degree to which all the data it contains fall within the intended sample frame for the study. This category of procedures includes properly defining the intended sample frame of the study, taking steps to prevent respondents outside the sample frame from taking the survey, and removing any responses that do not adhere to the sample frame during analysis.

Including respondents that do not adhere to a study's sample frame hinders researchers' ability to draw proper conclusions from their data [Groves, 1987]. For example, suppose a team of researchers wishes to investigate the impact of burnout on Chief Information Officer (CIO) retention. As part of their study, the researchers seek out 100 CIOs to complete a survey containing questions about their workload, stress level, job tenure, etc. Any conclusions that they reach from this data set would be held in question should any of the 100 respondents hold a position in an organization other than CIO.

The conclusions of a study are only applicable if the data used to make those conclusions are also applicable. No meteorologist would identify the temperature in Boston by using a thermometer located in San Francisco. Thus, one important aspect of data quality in survey research involves ensuring that data used in a study properly correspond to the intended target. When the sample frame of a study does not correspond to the intended population, the study is prone to "coverage error" [Couper, 2000]. Researchers must act to both ensure this correspondence and communicate it to their audience. Failing to do so could considerably diminish the confidence in knowledge generation.

### Completeness

The completeness of a data set pertains to the degree to which all necessary survey questions are fully answered by all respondents. Like the correspondence category, this category of procedures includes the steps taken to prevent respondents from actively skipping or accidentally overlooking any questions on a survey. This category also involves the steps taken after a survey to account for any missing data, which might include the removal of responses or the sophisticated recalculation of missing responses to allow for accurate analysis.

Missing data has both theoretical, practical, and statistical implications in knowledge generation [Schafer and Graham, 2002; Tsiriktsis, 2005]. Theoretically, missing data could reveal an

underlying bias in the sample, if individuals with certain characteristics or preferences uniquely failed to complete all or part of the survey [Roth, 1994]. Practically, missing data could indicate issues related to the reliability of the survey instrument, if the absent data can be traced to an unnecessary difficulty in taking the survey or a faulty line of questioning that rendered some or all the survey unclear or difficult to understand. Statistically, missing data alters the way a data set is analyzed [Tsiriktsis, 2005]. Missing responses are often removed from the data set, thereby reducing the size of the sample. Even under conditions where missing data can be replaced, the replaced data is merely an estimate of the original response, thus opening the analysis to questions of authenticity. Thus, it is vital that researchers not only act to safeguard against missing data where possible, but also communicate the actions taken. Knowledge generated from an incomplete or statistically estimated data set is less likely to be trusted and codified for the future.

### **Carelessness**

The carelessness of a data set pertains to the degree to which all respondents were attentive and engaged while taking the survey. Many surveys, especially those delivered remotely through an online medium, are prone to issues related to inattention [Huang et al., 2015]. In these cases, respondents either do not sufficiently read questions or provide adequate care in their responses, leading to overly (often impossibly) fast response times, straight-lining, patterned responses, etc. [Revilla, 2016]. This category of procedures includes steps that can be taken to increase respondents' attention, verify that proper attention was paid during survey taking, and remove responses that indicate a lack of attention.

Inattention in survey taking reduces the validity of the resulting data set. Data gathered from inattentive sources is unlikely to accurately represent the true values of those sources [Maniaci and Rogge, 2014]. It is vital in survey analysis that data gathered from respondents be their authentic responses. Otherwise, any conclusions drawn from those responses are subject to skepticism. For example, suppose a researcher utilizes a survey to examine the relationship between information overload and perceived ease of use. If enough respondents fail to attentively read the questions and simply mark "Strongly Agree" for every question, the data might indicate (falsely) that overloading users with information makes a system easier to use! Thus, to ensure the highest confidence in knowledge generation through survey research, scholars must consider the extent to which carelessness features in a data set.

### **Composition**

The composition of a data set pertains to the dispersion of the data after the survey is completed, specifically as it relates to the normality of the data and the presence of outliers. This category of procedures is almost entirely statistical, as very little can be done during survey composition or administration to prevent skewed results or extreme responses. Nonetheless, during data analysis, it is important that researchers examine their data to determine if actions must be taken to prevent any unnecessary influences on survey results.

The composition of a data set influences the conclusions which are drawn from that data. When assessing composition, researchers must consider both the dispersion of the entire data as well as the positioning of individual responses. Regarding the entire data set, many statistical analysis techniques are affected by a non-normal distribution. For example, Goodhue, Lewis, and Thompson [2012] found that data that is highly skewed and kurtotic reduces power and can influence results of structural equation modeling (both CB-SEM and PLS-SEM) and regression. Regarding individual responses, researchers should examine the data for outliers, both univariate and multivariate [Osborne and Overbay, 2004]. Like many of the other categories of procedures, accounting for the composition of data is important because if it is unaccounted for, any confidence in the knowledge generated based off that data can be eroded.

### **Credibility**

The credibility of a data set pertains to the degree to which the results of a survey are unnecessarily biased. Biases introduce theoretical threats to the validity of a study by inserting additional

influences on the survey results. Some of the most common biases that researchers often consider are non-response bias (biased results stemming from a meaningful difference in respondents and non-respondents) [Armstrong and Overton, 1977] and common method bias (in which respondents' answers are influenced by the method selected) [MacKenzie and Podsakoff, 2012]. This category of procedures includes the actions taken before and during survey administration to eliminate the threat of such biases, as well as the actions taken during data analysis to dispel any concerns that such biases may have been present.

Data that are biased lead to results that are biased. In the case of non-response bias, when present, the conclusions drawn from analysis would only pertain to those who share the same characteristics as those who responded to the survey. All others would not accurately be represented in the study. In the case of common method bias, when present, the conclusions drawn from analysis would be flawed, as the responses were unduly influenced by the method selected for the study. The goals of survey research must include the aim to extract the purest information from the respondents, without any influences unaccounted for in the investigation. Should any biases be left unexamined, the confidence in the results could be hindered.

## Implications

Across the five categories of the framework, numerous implications abound if researchers do not enact and communicate data quality procedures. In this section, we highlight some of the broader implications of ignoring data quality, and how such ignorance might inhibit knowledge creation in IS survey research.

### Implication #1: Data Misalignment

One negative consequence that could emerge from a failure to enact and communicate data quality procedures is the actual or perceived mismatch between the data used to confirm a set of hypotheses and the population in which that data should be drawn from. This consequence is especially tied to the correspondence and credibility categories of the framework, from different perspectives. If researchers fail to properly consider correspondence, they risk the allowance of respondents into a sample that are not a part of the population for the study. This is an issue of errant *inclusion*, otherwise referred to as sampling error [Groves, 1987]. If researchers fail to properly consider credibility (especially non-response bias), they risk limiting the sample to a biased subsection of the population. This is an issue of errant *exclusion*, otherwise referred to as non-response error [Groves, 1987].

Misalignment between the sample utilized in a survey and the population from which the sample should have been drawn has severe consequences for knowledge generation. It is assumed in statistical sampling that the conclusions drawn from a sample should be generalizable to the larger population. However, if the sample does not fully represent the population, the generalizability of the survey is altered, if not negated entirely.

Researchers must be cognizant of the potential problems associated with errant inclusion and exclusion of respondents. Proper consideration of the procedures included within our framework should help to alleviate these concerns, both in terms of actual data quality concerns and the perceptions of those concerns, should the procedures enacted not be communicated publicly.

### Implication #2: False Conclusions

Another negative consequence that could emerge from a failure to adhere to this framework is inaccurate conclusions from the survey data set. Careless, incomplete, or improperly composed data increases the potential to make a false conclusion. For example, earlier we described the implications of straight-lining on data analysis. Without considering the carelessness of that data set, the researchers might have observed a correlation between two variables that did not exist, or perhaps worse, a correlation that should have been reversed.

Survey data is often utilized in knowledge generation as a means of confirming one or more hypotheses derived from theory [Pinsonneault and Kraemer, 1993]. The conclusions from that data provide the evidence needed to support prior claims. If the conclusions were reached under false assumptions, knowingly or unknowingly, then the validity of the entire work is diminished.

Data cleansing is a vital activity in the process of knowledge generation. Just as improper interpretations of theory lead to inaccurate hypotheses, so does errant data lead to inaccurate conclusions. The procedures described in this framework should aid researchers in avoiding the errors which could lead to false conclusions in survey research.

### **Implication #3: Statistical Error**

Finally, while failure to enact and communicate data quality procedures has theoretical consequences, it also has statistical consequences that must be considered. Many of the statistical methods that are utilized in survey research rely on assumptions which may inadvertently be violated if data quality procedures are not followed. For example, regression analysis assumes the normality of a data set [Jarque and Bera, 1980]. If this assumption is not met, the analysis may not be feasible or may lead to a false conclusion.

Statistical considerations also include potential limitations to the power of a study. An overabundance of missing data could inhibit the ability to detect a potential relationship even if it were present. In the case of outliers, extreme values could influence a study's conclusions unnecessarily. This implication relates to the second, in terms of leading to potential false conclusions, but from a different perspective. While low-quality data could lead to false conclusions reached through accurate statistical methods, such low-quality data could also inhibit the ability to properly complete those statistical methods. Thus, low-quality data has implications both for the end of the methodological procedural chain as well as the steps which lead to that end.

## **III. REVIEW OF CURRENT RESEARCH**

To identify areas that are most deficient in our field, with an eye toward how we can best augment IS Ph.D. curricula, we are in the process of evaluating how and to what extent the 5-C framework is enacted in published research. To that end, this section details our methodology for reviewing the current state of IS research as pertains to data quality and offers preliminary results<sup>1</sup>.

### **Procedure**

To properly understand the current state of IS research, we are conducting a thorough review of survey research in our field. The first step in this process was identifying all articles that used a self-report survey methodology in the Information Systems Basket of 8 Journals (the *European Journal of Information Systems*, *Information Systems Journal*, *Information Systems Research*, *Journal of the Association for Information Systems*, *Journal of Information Technology*, *Journal of Management Information Systems*, *Journal of Strategic Information Systems*, and *Management Information Systems Quarterly*) [AIS, 2011] over a ten-year period. In all, we began with 2,692 articles.

Our coding procedure is as follows: First, we conducted a keyword search (using the terms, "survey," "questionnaire," and "panel") of all articles published in these journals from 2008-2017. We included all papers with surveying as the principal method for data collection and primary data used for analysis (i.e., data obtained and managed by the authors). We excluded papers with surveying as subordinate to some other method (e.g., an experimental study that uses a follow-up

---

<sup>1</sup> Note: as data analysis is still ongoing, please recognize that many of the statistics presented in this research-in-progress submission may not be finalized. Nonetheless, the figures reported here should be close to the final results, barring additional cleanup work.

survey) or data gathered from an archived source (e.g., papers that use data from Gallup surveys). This provided a final sample of 419 articles for coding. Some articles used more than one sample, thus the total number of surveys coded was 430.

All articles were coded using the 5-C framework described earlier. Our primary concern was whether the data quality procedures were discussed in each paper and what procedures were enacted. Specifically, we examined each article's method section and coded the following items: sample size (actual responses received); final sample size (used in analysis); procedural remedies (e.g., screening questions, reverse-coded items, attention-check questions), invalid or impossible responses, inattention (e.g. straight-lining, patterned responses, overly fast responses), missing data, outliers, normality, non-response bias, and common method bias. We coded whether each of these items was discussed in the paper, how many responses were removed because of each item, and the criteria/procedure utilized for each item.

## Preliminary Results

One priority in our investigation is understanding how many, and why, cases were removed from survey results. In coding the surveys, we sought to identify the initial sample size, the final sample size, and how many cases were removed due to data quality procedures. The initial results are provided in Tables 2 and 3.

Table 2. Average Number of Cases Removed in Survey Research

Description	Value
Initial Sample Size	434.25
Final Sample Size	390.47
Difference	43.78

Of the 430 surveys analyzed thus far, 201 (46.7%) reported removing cases as a part of data quality procedures. The remaining 229 surveys either reported no cases removed or did not discuss in the text of the paper whether or how many cases were removed. Notably, we discovered articles that merely listed the number of "usable" or "valid" cases, without mentioning the initial sample size, thereby preventing the reader from knowing how many cases were removed.

Table 3. Data Quality Procedure Mentions

Data Quality Procedure	Mentioned	Not Mentioned	Percentage
Sample Frame	105	325	24.4
Inattention	22	408	5.1
Impossible Responses	5	425	1.2
Missing Data	159	271	37.0
Univariate Outliers	21	409	4.9
Multivariate Outliers	21	409	4.9
Normality	150	280	34.9
Common Method Bias	308	122	71.6
Non-Response Bias	217	213	50.5

Regarding procedures, we looked for any mention of the data quality procedures mentioned in Table 3, whether or not they resulted in any cases removed. For example, a survey was coded affirmatively if the article discusses inattention even if no inattention was found. Surveys were coded negatively if there was no mention of the procedure in the entire text of the paper.



Immediately evident from this analysis is that our field has placed an emphasis on checking for Common Method Bias (71.6%) and Non-Response Bias (50.5%). Of the procedures analyzed, these were the only two discussed in more than half of the surveys analyzed. Common Method Bias seems increasingly emphasized, as its discussion in articles has grown over time, displayed in Figure 1. Non-Response Bias is somewhat consistent over time in its frequency of discussion.

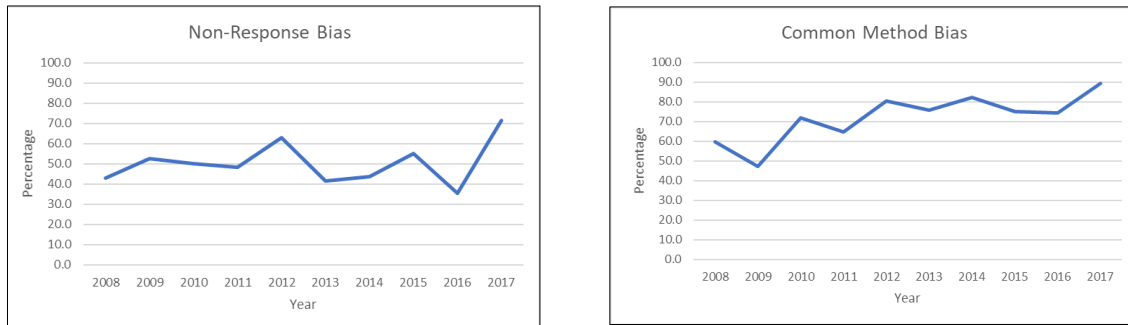


Figure 1. Mentions of Method Bias Over Time

Further analysis from Table 3 reveals some areas that are discussed far less often in published articles. For example, only 159 (37.0%) of the surveys analyzed mentioned missing data in the paper. Even lower frequencies were noted for inattention, outliers, and normality. Because these data quality procedures are not mentioned in many of the papers, it is unclear whether checks were even made. Furthermore, concerning Ph.D. students, if data quality procedures are not discussed in articles, as is evident for many of these procedures, students cannot learn best practices and norms cannot be established.

Going forward, our plan is to finalize our coding and dive deeper into data analysis. We will analyze not only the mentioning of data quality procedures, but also the methods used for each procedure. For example, early indications reveal that a large majority of papers that found missing data utilized casewise deletion, where the entire case is removed from analysis. This runs counter to the recommendations of some experts, who offer more advanced techniques that can prevent entire cases from being removed (e.g. Newman, 2014).

#### IV. NEXT STEPS

Utilizing the 5-C Framework, we will examine current research on data quality procedures to develop a set of prescriptive guidelines for researchers. The guidelines will address how best to enact data quality procedures, as well as how to communicate those procedures in written form. For example, regarding the “Completeness” category of the framework, we will provide guidance on how researchers can improve the completeness of their data during data collection, how they can account for missing data during analysis, and how such procedures should be communicated once the study is documented.

The guidelines will serve three aims. First, we aim to provide future researchers with a set of best practices on how to address each category of data quality procedures and improve data quality to the extent possible. This should ease the burden on researchers in deciding how to tackle this difficult issue and ensure that best practices become codified in the IS knowledge base. Second, we aim to provide a standard method for communicating data quality procedures in written form. It is our hope that we can remove some of the mystery surrounding when, how, and why to communicate the procedures enacted to improve data quality. Third, for Ph.D. students, we aim to identify opportunities for incorporating these guidelines into the Ph.D. curricula. This should alleviate the anxiety experienced by Jane, and students like her, described in our opening vignette. More importantly, it will help the field in its endeavors to train a new generation of IS scholars to produce knowledge that we can all have confidence in.

## V. IMPACT FOR PH.D. CURRICULA

In the opening vignette, we described a common problem facing Ph.D. students conducting survey research. Early results from our review indicate that little guidance is being provided through published research regarding how to enact data quality procedures. It is imperative that the curricula in our Ph.D. programs fill in these gaps. Through a finalized review of current research and the completion of our data quality guidelines, we aim to offer the following implications for improving data quality education in our field:

- First, the presentation of the 5-C framework should provide a foundation on which to build Ph.D. curricula. The current ambiguous nature of data quality procedures necessitates a guiding framework. Through our categorization of correspondence, completeness, carelessness, composition, and credibility, professors will be better equipped to discuss data quality procedures and students will have a way of easily recalling the primary facets of data quality.
- Second, through our review of data quality procedures in current research, we will highlight areas in which our field may be deficient. Improvements in curricula cannot be made until we fully understand where the problems lie.
- Third, using the 5-C framework, we will offer guidance regarding some best practices for each of the categories of data quality procedures. While an exhaustive discussion of each category would be overly cumbersome, our aim is to summarize contemporary insights and provide a roadmap for Ph.D. students eager to dive deeper into proper methods.
- Finally, more generally, we hope that this study serves to shine a light on some of the seemingly neglected aspects of data quality in survey research. It is imperative that we work together to establish norms surrounding how best to both enact and communicate data quality procedures. This has implications not only for Ph.D. students, but for all researchers in our field.

In sum, while this study has implications that extend beyond Ph.D. curricula, there are many ways in which it can be helpful for those who teach Ph.D. students. We hope to help our fictionalized Jane with the uncertainty she faces over the proper treatment of her data set.

## VI. REFERENCES

- AIS (2011) Senior Scholars' Basket of Journals. Retrieved May 1, 2018 from <https://aisnet.org/general/custom.asp?page=SeniorScholarBasket>.
- Armstrong, J. S., and Overton, T. S (1977) "Estimating nonresponse bias in mail surveys", *Journal of marketing research*, 396-402.
- Compeau, D., Marcolin, B., Kelley, H., & Higgins, C (2012) "Research commentary—Generalizability of information systems research using student subjects—A reflection on our practices and recommendations for future research", *Information Systems Research*, 23(4), 1093-1109.
- Couper, M. P (2000) "Web surveys: A review of issues and approaches", *The Public Opinion Quarterly*, 64(4), 464-494.
- DeSimone, J. A., and Harms, P. D (2017) "Dirty Data: The Effects of Screening Respondents Who Provide Low-Quality Data in Survey Research", *Journal of Business and Psychology*, 1-19.
- DeSimone, J. A., Harms, P. D., and DeSimone, A. J (2015) "Best practice recommendations for data screening", *Journal of Organizational Behavior*, 36(2), 171-181.
- Goodhue, D. L., Lewis, W., and Thompson, R (2012) "Does PLS have advantages for small sample size or non-normal data?", *MIS Quarterly*, 981-1001.
- Groves, R. M (1987) "Research on survey data quality", *The Public Opinion Quarterly*, 51, S156-S172.

- Huang, J. L., Liu, M., and Bowling, N. A (2015) "Insufficient effort responding: Examining an insidious confound in survey data", *Journal of Applied Psychology*, 100(3), 828.
- Jarque, C. M., and Bera, A. K (1980) "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics letters*, 6(3), 255-259.
- MacKenzie, S. B., and Podsakoff, P. M (2012) "Common method bias in marketing: causes, mechanisms, and procedural remedies", *Journal of Retailing*, 88(4), 542-555.
- Maniaci, M. R., and Rogge, R. D (2014) "Caring about carelessness: Participant inattention and its effects on research", *Journal of Research in Personality*, 48, 61-83.
- Newman, D. A (2014) "Missing data: Five practical guidelines", *Organizational Research Methods*, 17(4), 372-411.
- Osborne, J. W., and Overbay, A (2004) "The power of outliers (and why researchers should always check for them)", *Practical assessment, research & evaluation*, 9(6), 1-12.
- Pinsonneault, A., and Kraemer, K (1993) "Survey research methodology in management information systems: an assessment", *Journal of management information systems*, 10(2), 75-105.
- Revilla, M (2016) "Impact of raising awareness of respondents on the measurement quality in a web survey", *Quality & Quantity*, 50(4), 1469-1486.
- Roth, P. L (1994) "Missing data: A conceptual review for applied psychologists", *Personnel psychology*, 47(3), 537-560.
- Schafer, J. L., and Graham, J. W (2002) "Missing data: our view of the state of the art", *Psychological methods*, 7[2], 147
- Tsikriktsis, N (2005) "A review of techniques for treating missing data in OM survey research", *Journal of operations management*, 24(1), 53-62.