STRATEGIES FOR IMPROVING DATA RELIABILITY FOR ONLINE SURVEYS: A CASE STUDY

Tung-Zong (Donald) Chang
Metropolitan State University of Denver
Campus Box 79, Denver, CO 80217 USA
changd@msudenver.edu

Nicole Vowles
Metropolitan State University of Denver
Campus Box 79, Denver, CO 80217 USA
nvowles@msudenver.edu

ABSTRACT

Due to the wide availability of the Internet, online surveys have become a mainstream data collection method and offer new-found opportunities in market research. However, there are inherent weaknesses associated with online surveys. A case study is used to illustrate how a carefully devised strategy in sample specification and selection, data processing, screening, and editing can boost the quality of online survey data. Reliability measures are compared between the paper and online versions of a newly developed scale of consumer attitudes toward environmental friendly household cleaners and an established scale on materialism. The results show that with a carefully developed strategy, planning, and execution, online survey data can be equal or superior to that of equivalent paper survey data. This provides important applications and implications to researchers and practitioners who are interested in taking advantage of the new data collection medium.

Keyword: Online Survey, Self-Administered Survey, Reliability

1. INTRODUCTION

The advent of the Internet has transformed many aspects of our daily lives. One area that has seen dramatic changes is the increasing acceptance and popularity of online surveys at the expense of other primary survey

research methods, such as telephone surveys, mail surveys, and personal interviews. With broader coverage and wider accessibility of the Internet, online surveys have become ubiquitous as evidenced by its presence in numerous websites and pop-up windows.

The early motivation to collect via the Internet was due to its convenience, speed, and cost. As more data are collected online, pros and cons have been identified in the literature. In addition to common issues related to survey methods, such as validity, reliability, respondent confidentiality and consent, among others, online surveys also have unique advantages and disadvantages related to the nature of the Internet. We first examine some of the widely recognized advantages and disadvantages. A strategy is developed and reported to counter one of the key issues often associated with online surveys: data reliability. Finally, we report the results of a comparison study, online vs. paper, based on reliability measures of one newly developed scale and one established scale.

1.1 Advantages of Online Survey Method

Time Efficient and Flexible: Without a doubt, online surveys can be conducted very quickly and efficiently, often in a matter of hours or days, assuming sufficient server capacity and Internet traffic. Research has shown that the majority of responses can be returned within two days[1]. This allows greater flexibility, because respondents can complete the survey at their own pace and at a time convenient to them.

Interactivity without Interviewer Bias: The interactivity nature of the Internet allows greater flexibility in questionnaire design without interviewers involved, thus effectively eliminating the concerns of interviewer bias[2].

Personalized Messages and Questions: The interactive nature of the Internet provides invaluable opportunities for personal marketing, customized marketing, and one-on-one marketing. The same advantages also apply to survey research. Personal and customized messages can be built into an online survey. Questions can be personalized and customized depending on research needs.

Quick or Immediate Follow-Up: Some of the main advantages associated with online surveys are the ability to conduct research extremely fast and to immediately follow-up. As soon as an online survey is completed and submitted, it is possible to deliver an instant follow-up, such as a thank-you message or a verification message. With proper planning, programming, and execution, it is also possible to effectively screen out potentially invalid data via instant follow-ups.

Cost Effectiveness: Online surveys are often cheaper, as most costs including hosting, programming, software, and other Internet- and computer-related expenses associated with web-based surveys are less expensive compared with other primary survey methods. While paper surveys are usually considered the cheapest, online surveys can further eliminate waste in survey production and distribution[3].

Targeted Sample Selection: Similar to how marketers have used the Internet to better choose their target customers, Internet technology allows researchers to more accurately select respondents, e.g., with a pop-up invitation window when certain responses are recorded by the system. The online survey method is often the logical choice when the targeted sample populations are part of virtual communities. Internet surveys are often better received by younger populations who are active participants of virtual communities[4].

Unrestricted Geographic Coverage: One key advantage of the Internet is its ability to effectively eliminate geographic boundaries. As a result, distance and geographic barriers are no longer concerns for survey research. Respondents can be reached in any corner of the world as long as the Internet is accessible.

Desensitize Sensitive Subjects: Being able to complete an online survey with privacy and convenience might reduce respondents' sensitivity toward certain controversial or personal subjects. Respondents may find it less embarrassing to answer certain private questions or feel more comfortable discussing controversial topics. Online surveys can be very useful in collecting certain information that is difficult to be shared via other survey methods, such as personal interviews or phone interviews. Online surveys can also be used to collect information related to unpopular beliefs or attitudes. In addition, online surveys provide stronger anonymity than other primary survey methods.

Less Processing Errors: Responses are recorded instantly online. With careful design, respondents may not be allowed to input invalid responses. As a result, typical data input and processing errors can be avoided. The remaining potential data errors are typically editing errors as well as data processing and transformation errors.

1.2 Disadvantages of Online Survey Method

Constrained by Internet Accessibility: The first shortcoming of online surveys involves Internet inaccessibility. Certain populations still lack adequate access, physically or mentally, to the Internet. Even with reliable access, some potential respondents do not feel secure enough to

provide information via the Internet. Since their inception in the 1990s, online surveys have been best designed for respondents who have easy access to the Internet and feel comfortable sharing information online or are early adopters and innovators in society[5].

Truthfulness Concerns: An often cited issue with online surveys is respondent truthfulness. While such a concern is also true for other survey methods, it is even more applicable for online surveys. Underage users, who are often unintended research target subjects, may complete online surveys and provide false information for various reasons. It is usually more difficult to detect such impropriety, because it is difficult or impossible to carry out compatible verification methods.

Ethical and Legal Concerns: Privacy is an inherent issue associated with the Internet. Such an issue is more evident as online surveys conducted via commercial, for-profit online survey services, such as SurveyMonkey, Zoomerang, SurveyGizmo, and Qualtrics. Data collected via commercial services are stored on the service provider's servers and thus are not under complete control of the researcher. When survey data are stored online and managed by third party providers, the researcher cannot provide complete confidentiality, anonymity, privacy, and security. This could present a certain degree of ethical and legal dilemmas. Greater care must be taken with online surveys regarding informed consent and potential deception, even if unintentional.

Selection/Sampling Bias: Frequent Internet users are likely very different from infrequent users or non-users. As a result, population bias is another inherent issue with online survey methods. Some potential respondents might not be reachable via the Internet. Therefore, online surveys might be more appropriate for experienced Internet users or users who are more comfortable with online idea exchange. Overall, Internet studies tend to capture younger populations[4].

Inadequate Response Rates: There is a general, natural resistance to new technology. As many people are concerned about various issues associated with Internet use, some have reported lower than usual response rates due to the high number of non-responses[1, 6, 7].

Inaccurate Populations: Compounding with the issue of inadequate response rate, some researchers might inappropriately expand the target distribution lists to secure a sufficient number of respondents in order to meet the client's needs or requirements for statistical adequacy. Furthermore, many online surveys are made available online, open to any visitor. As a result, unexpected respondents might complete an online survey without the

researcher's knowledge. When an online survey is distributed beyond its original target population, it invalidates the original intent of the survey.

1.3 The Need for a Quality Enhancement Strategy

Despite the disadvantages of online surveys, the pros as a whole often outweigh the cons. Their use is expected to continue to increase as the Internet is now considered a necessity and no longer a selective service in developed economies. With the recent push to make the Internet available to all households in the U.S., low-cost Internet service is now provided to low-income populations. For survey researchers, the issue may no longer be whether the online survey method is appropriate but rather how to improve the quality of online surveys and conduct them correctly. Survey researchers have proposed a number of approaches including the use of commercial research companies. The offer of incentives to respondents, commonly used by commercial researchers to increase response rates, has also been suggested. Others have proposed proper netiquette when conducting surveys online.

In response to the needs cited above, this paper uses a case study to present a strategy with specific steps for improving online survey data quality with the following objectives:

- eliminating unwanted responses,
- maintaining data consistency,
- retaining valid data, and
- enhancing overall data reliability.

2. CASE STUDY: HOUSEHOLD CLEANER USAGE SURVEY

A survey on consumer attitudes, product usage, environmental concerns, materialism, and other related factors was conducted in spring 2011. Data were collected from various on-campus marketing classes and online marketing classes. For on-campus classes, paper self-administered questionnaires were distributed and collected during class. For online classes, students were directed to the online version of the survey via email. The on-campus survey provides an invaluable benchmark for comparison purposes. The online version was prepared and available on Qualtrics, a popular online survey service for researchers. In addition to online questionnaire design, Qualtrics also provides online assistance with data tabulations, analysis, tables, charts, etc.

To overcome a number of weaknesses often associated with online surveys, a strategy was developed to effectively deal with specific issues in the data collection, screening, processing, and editing process. This strategy has the fundamental objective to keep the online respondents as similar to the on-campus respondents as possible. The strategy is outlined below:

1. Sample Population Specification: The choice of sample population includes students enrolled in marketing classes in a metropolitan university serving a well-defined geographic area. This choice effectively eliminates concerns over Internet accessibility, as all students have access to a computer and the Internet. Conducting a survey for academic purposes removes some of the application and generalization issues related to commercial research and allows us to define the sample population in such a way that a homogeneous population is acceptable. Defining sampling population in a specific manner helps minimize potential sampling bias or selection bias.

2. Sample Screening: Once the sample population was defined, a screening mechanism was built into the online survey to sort and screen out unwanted responses, thus removing the concerns of an inaccurate population. Respondents were asked how they found out about the online survey and to provide a four-digit ID number for verification purposes. Respondents who failed to answer the source question correctly or failed to provide an adequate ID number were removed.

3. IP Address Verification: IP addresses could be traced back to original geographic proximities. As a second layer of filtering, the online survey was limited to a defined geographic coverage, consistent with typical geographic coverage of on-campus students to maintain equivalent geographic coverage.

4. Removal of Repetitive or Similar IP Addresses: While it is possible that multiple qualified students might reside in the same residence using the same computer and Internet service, repetitive or similar IP addresses might indicate other concerns. This is a trade-off between reduced sample size and data validity; the latter is considered more important. Therefore, to avoid double-counting and other potential issues, only first original responses were retained when there were multiple repetitive or similar IP addresses.

5. Removal of Invalid Responses: All data were first screened individually to remove questionable responses such as same-across-the-board, excessive missing values, and other inadequacies that could be spotted visually.

6. Removal of Inconsistent Responses: The final checks were made by comparing responses for selected pairs of questions, which were supposed to be similar or dissimilar. When a respondent failed a number of repeated tests, it was removed.

3. RESULTS

We compare the on-campus version and the online version based on their reliability measures (Cronbach's alpha) on two main constructs: consumer attitudes toward environmental friendly (green) household cleaners and materialism. Each construct contains components and sub-components. The consumer attitude scale is developed based on a literature review and input from in-depth interviews; it is an untested scale. The materialism is an existing, established short-form scale[8]. The consumer attitude construct consists of two main components: product-environmental belief and consumer readiness. The product-environmental belief component consists of three sub-components: general belief, product-related perceptions, and significance of concerns. Consumer readiness contains sub-components of trade-off and purchase intention. Materialism has three components: success, centrality, and happiness[9].

Cronbach's alpha is reported for each main construct and its components and subcomponents. The results are shown in the table. Most alpha values are greater than 0.70, a common benchmark used in social science or marketing literature. Overall, the reliability of the consumer attitude scale and materialism scale is satisfactory. The consumer attitude scale, in general, outperforms the established materialism scale. A comparison between the paper and online versions reveals that the reliability measures for the online version consistently outperform those of the paper version. All but one alpha value for the online version are greater than those of the paper version. As a result, we conclude that the online survey data is of higher quality than the paper version data.

Table 1. Reliability analysis of constructs, components, and sub-components

Constructs, Components, and Sub-Components	Paper α	Online α	Combined α
Overall Consumer Attitude Toward Green Household Cleaners (18 items)	0.93	0.95	0.95
Product-Environmental Belief: Does product usage have a negative effect on the natural environment? (12 items)	0.93	0.96	0.95
1. General Environmental Beliefs: is it real? (3 items)	0.81	0.86	0.84
2. Product-Related Environmental Perceptions: Is product usage a concern? (5 items)	0.83	0.92	0.89
3. Significance of the Product-Associated Environmental Concerns: Is the concern significant? (4 items)	0.94	0.96	0.95
Consumer Readiness: Is the consumer ready for green household cleaners? (6 items)	0.86	0.87	0.88
4. Price (trade-offs) for Addressing the Environmental Concerns: Is it expensive to switch? (3 items)	0.78	0.79	0.79
5. Green Household Cleaner Purchase Intention (3 items):	0.89	0.91	0.90
Overall Materialism: (12 items)	0.76	0.83	0.82
1. Success (3 items)	0.63	0.70	0.67
2. Centrality (3 items)	0.71	0.67	0.71
3. Happiness (3 items)	0.68	0.80	0.77

4. CONCLUSION

We attribute the superior quality of online survey data to the strategy that was implemented for data collection, screening, processing, and editing. In fact, we believe that we have effectively addressed many concerns that are associated with online surveys and show that the online survey method can be superior to other survey methods, in this case, self-administered surveys. The finding provides invaluable insight on how to utilize modern technology to advance data collection methodology.

5. REFERENCES

[1] D. Bachmann, J. Elfrink, and G. Vazzana, Tracking the progress of e-mail vs. snail-mail. *Marketing Research*, 8(2), p31-35, 1996.

[2] L. Sproull, and S. Kiesler, Reducing social context cues: Electronic mail in organizational communication. *Management Science*, 32(1), p1492-1512, 1986. http://dx.doi.org/10.1287/mnsc.32.11.1492.

[3] S. Fricker, M. Galesic, R. Tourangeau, and T. Yan, An experimental comparison of web and telephone surveys. *Public Opinion Quarterly,* 69(3), p370-392, 2005. http://dx.doi.org/10.1093/poq/nfi027.

[4] M. Schonlau, A. van Soest, A. Kapteyn, and M. Couper, Selection bias in web surveys and the use of propensity scores. *Sociological Methods*

&   *Research*,   37(3),   p291-318,   2009.
http://dx.doi.org/10.1177/0049124108327128.

[5] J.H. Ellsworth, and M.V. Ellsworth, *Marketing on the Internet*. New York: Wiley, 1997.

[6] D.A. Dillman, G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B.L. Messer, Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, 38(1), p1-18, 2009. http://dx.doi.org/10.1016/j.ssresearch.2008.03.007.

[7] T.H. Shih, and X.T. Fan, Comparing response rates from web and mail surveys: a meta-analysis. *Field Methods*, 20(3), p249-271, 2008. http://dx.doi.org/10.1177/1525822X08317085.

[8] M. Richins, The material values scale: Measurement properties and development of a short form, *Journal of Consumer Research*, 31(1), p209-219, 2004. http://dx.doi.org/10.1086/383436.

[9] M. Richins, and S. Dawson, A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, 19(3), p303-316, 1992. http://dx.doi.org/10.1086/209304.