Do Mail and Internet Surveys Produce Different Item Nonresponse Rates? An Experiment Using Random Mode Assignment

Morgan Millar[1], Don Dillman[1]

[1] Washington State University

A significant limitation of most data quality comparisons between Web and mail survey responses is that the individuals who choose to respond to one survey mode have different characteristics than those who select the alternative mode. As suggested by other papers in this issue of *Survey Practice*, such differences in, for example, Internet access, education, and income, may contribute to mode differences in item nonresponse. This makes it difficult to isolate how mode itself affects this aspect of survey quality.

The surveys of undergraduate students analyzed herein differ from general public surveys in this regard, inasmuch as all students in this fairly homogenous (for education and age) population have Internet access and are accessible by both postal mail and e-mail. Since college students are relatively highly Internet-literate and course work requires nearly daily use of the Internet, it is possible to randomly assign students to Web-only and mail-only treatment groups. In this paper, we determine if mail and Web item nonresponse rates are similar when randomly assigning individuals to a response mode. Additionally, we examine patterns of item nonresponse for different types of question formats.

Data and Methods

We examined results from two surveys of randomly selected undergraduate students at Washington State University (WSU). The surveys were conducted in the spring and fall of 2009, respectively. The first study focused on students' general satisfaction with WSU and their educational experiences. The second study primarily asked about the effects of the economic downturn and university budget cuts. The spring questionnaire contained 36 numbered questions requesting as many as 100 responses (depending upon skips), and the fall questionnaire contained 33 numbered questions requesting up to 74 responses.

The paper and Web questionnaires were created using a unified mode design to standardize the visual appearance and reduce the potential for mode differences (Dillman, Smyth, and Christian 2009). Both studies used multiple

Table 1   Unit Response Rates and Number of Cases Used in these Analyses, by Treatment Group for Each Study.[a]

	Total unit response rate[c]	Cases used in this study[d]
Spring 2009 treatments (sample size)[b]		
Mail response (681)	53.2	351
Web response 1 (676)	50.2	281
Web response 2 (678)	64.5	387
Sum of web responses		668
Fall 2009 treatments (sample size)		
Mail response (683)	43.9	299
Web response 1 (487)	42.5	194
Web response 2 (589)	38.2	209
Web response 3 (586)	21.2	115
Web response 4 (669)	20.5	126
Sum of web responses		644

[a] Both surveys included additional treatments which offered a choice of mail or web response. These treatments are not used in this study because they did not randomly assign a response mode. Each survey included multiple web response treatment groups which varied in implementation strategies, such as mode of correspondence and inclusion of token cash incentives. This explains the variation in response rates across web treatments. When using similar implementation strategies, the mail and web treatments obtained similar response rates.

[b] Undeliverables are subtracted out of sample sizes.

[c] Response rate= (number completed + number partial completes/ sample size). This corresponds to AAPOR RR6.

[d] For the first study, in the final set of correspondences we offered nonrespondents the opportunity to respond via the alternate mode of that which they were originally offered. These "mode switch" respondents are not included in this study. Also, for purposes of these analyses, we excluded web "drop-off" respondents, or partial completes. However, both of these types of responses are included in the calculation of the unit response rates.

experimental treatment groups to compare the effects of mode and implementation strategies on response rates (the experimental designs were detailed in Millar and Dillman 2011). Table 1 displays the response rates for each treatment, as well as the number of responses used in the following analyses from each treatment.

We limited the following analyses to responses resulting from the random assignment to a response mode. Those respondents who were offered a choice of response mode were excluded. We also excluded Web break-off respondents (there were no comparable mail break-offs). We did not include responses such as "don't know" and "not sure" in the counts of item nonresponse. We calculated the percent of missing responses for each respondent, and then calculated an average, composite rate of missing responses by mode of response for each study. We used z tests for differences in proportions to determine if there are any statistically significant differences in the rates of item nonresponse across modes.

Results

Item Nonresponse Rates by Mode

Table 2 displays the average rate of item nonresponse for each study, by mode. The table shows that although the mail item nonresponse rates in these studies were slightly higher than those of the Web, there were no statistically significant

Table 2   Item Nonresponse Rates for Both Studies, by Mode of Response.[a]

	Web		Mail		p-value
	n	%	n	%	
Spring 2009 Study	668	2.14	351	2.47	0.742
Fall 2009 Study	644	1.18	299	1.89	0.390

[a] Percentages represent an average of the item nonresponse rates for all respondents to each mode, for all items in the questionnaire.

differences in the average rate of item nonresponse between the two modes. This suggests that mode differences are negligible when individuals are randomly assigned to a response mode.

Although the composite rates in Table 2 indicate there was no overall mode difference, several individual items within each survey exhibited significant mode differences in item nonresponse rates. In the first study, 19 percent of the items had significant mode differences in item nonresponse rates. Forty-nine percent of the questionnaire items in the second study showed statistically significant mode differences in item nonresponse rates, which is noticeably higher than the percentage of items in the first study. This high percentage is especially surprising because there was no statistically significant mode difference in the composite rates of item nonresponse for either study. This encouraged us to explore potential sources of mode differences in item nonresponse.

Effect of Question Format

We first examined whether any particular question formats were common among those items exhibiting significant mode differences in item nonresponse. In both studies, multi-item, battery list questions were a frequent source of mode differences in item nonresponse rates. In the first study, 37 percent of the items showing significant mode differences in item nonresponse rates were part of multi-item, battery list questions. In the second study, this percentage was 57 percent. In both surveys, mail respondents were more likely than Web respondents to skip items within a multi-item battery.

Another source of mode differences in the second survey was four qualitative, open-ended questions. In contrast to the multi-item trend, these questions produced higher item nonresponse rates in Web than in mail. This suggests there is no simple trend in the effects of question format. Mode effects on item nonresponse are more complicated than indicated by common perception, which is that mail consistently produces higher item nonresponse rates than Web.

Table 3 Item Nonresponse Rate by Question Type and Mode.[a]

	Total[b]		Web		Mail		p-value
	n	%	n	%	n	%	
Spring 2009							
Open-ended questions (8)[c]	1019	5.03	668	5.33	351	4.46	0.547
Multi-item questions (69)	1019	2.38	668	2.29	351	2.55	0.792
Other question types (23)	1019	0.99	668	0.69	351	1.56	0.182
Branched-to questions, of any format (5)	1019	1.49	668	1.26	351	1.94	0.395
Fall 2009							
Open-ended questions (9)	963	5.18	644	6.17	299	3.05	**0.044**
Multi-item questions (37)	963	1.13	644	0.76	299	1.93	0.117
Other question types (30)	963	0.56	644	0.20	299	1.35	**0.028**
Branched-to questions, of any format (9)	963	2.71	644	3.33	299	1.37	0.085

[a] Percentages represent an average of the item nonresponse rates for all respondents to each mode, for all items of that particular format in the questionnaire.

[b] For each study, we used z-tests to compare item nonresponse rates of open-ended, multi-item, and other question types to each other (p values not shown). For the first study, these tests indicate open-ended questions produced a significantly higher rate of item nonresponse than both multi-item and other question types. Additionally, multi-item questions' item nonresponse rate was significantly higher than that of other question types. In the second study, the item nonresponse rate for open-ended questions was also significantly higher than those of both other question formats. Statistical tests were adjusted for multiple comparisons using the Bonferroni-Holm method (Holm 1979).

[c] Numbers in parentheses indicate the number of questionnaire items of that format which were included in the survey.

We next calculated the average respondent item nonresponse rates for three question types across modes: multi-item lists, open-ended questions, and other question types (such as yes/no, ordinal scales, and nominal scales), as well as for questions of any format which were branched to in a skip pattern. These rates for each study are summarized in Table 3. The table indicates that, for the first study, the average rates of item nonresponse for each question type are not statistically different across modes. However, the second study exhibited significant mode differences in item nonresponse for open-ended questions as well as other question types. Surprisingly, the mode difference for multi-item questions was not statistically significant. Thus, although certain question types seemed to be associated with significant mode differences in item nonresponse rates, the composite rates for each question type do not consistently vary across modes in either study.

We also examined whether the composite rates of item nonresponse for open-ended, multi-item, and other question formats (shown in the first column of data in Table 3) are significantly different from each other, regardless of mode. These results indicate that open-ended questions have significantly higher rates of item nonresponse than both multi-item and other question formats for both studies. Additionally, in the first study, multi-item questions have a significantly higher rate of item nonresponse than other question formats.

Effect of Survey Topic

We also considered whether the change in survey topic could have affected mode differences in item nonresponse. The topic of economic troubles and budget cuts (which was more prevalent in study two) seemed to matter. Forty-nine percent of the items exhibiting significant mode differences in responses in the second study were items related to these topics. Since this topic appeared less in the first study, only five percent of the significant mode differences in this study were related to it. This could partially explain why we observed a higher percentage of individual items showing mode differences in item nonresponse in the second study. Mail respondents were less likely than Web respondents to answer questions related to economics/budget cuts. Perhaps these topics appeared more sensitive and respondents were more comfortable responding to such items via Web.

Conclusion

This paper shows that in studies in which sampled individuals are randomly assigned to a response mode, the composite rate of item nonresponse is the same for Web and mail. However, our deeper analysis suggested the story is more complex. We found that despite the overall similarity across modes, a good number of individual items, particularly in the second survey, exhibited significant mode differences in item nonresponse rates. Our data demonstrate there may be other factors which play a role in creating mode differences in item nonresponse rates. In particular, this analysis showed that question format and survey topic may impact item nonresponse depending upon the mode of study.

Further cognitive interview work could help illuminate why these factors may differentially affect response in mail and Web. Without such insights, we can only speculate as to why survey topic and question format may matter. For example, perhaps multi-item battery list items seem more burdensome to respondents completing paper questionnaires as opposed to online questionnaires. Also, responding over the Internet may be more comfortable for students when dealing with topics that can be more personally sensitive, such as financial well-being.

This paper suggests that when we use random assignment to response modes, there is not a great deal of difference between mail and Web in item nonresponse rates. However, this does not mean response mode is inconsequential. Other factors appear to create such differences. Therefore, surveyors should continue to consider how mixing modes may affect data quality, and invest in future research to disentangle the sources of mode differences.

Acknowledgements

REFERENCES

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.

Holm, S. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6: 65–70.

Millar, M.M., and D.A. Dillman. 2011. "Improving Response to Web and Mixed-Mode Mode Surveys." *Public Opinion Quarterly* 75 (2): 249–69.