MODE EFFECTS IN AN INTERNET/PAPER SURVEY OF EMPLOYEES

Carl Ramirez, U.S. General Accounting Office, Kevin Sharp, Schulman Ronca & Bucuvalas, Inc.,
Luis Foster, Schulman Ronca & Bucuvalas, Inc.
Carl Ramirez, U.S. GAO, 441 G St. NW, Rm. 2921, Washington DC 20548, ramirezc.ggd@gao.gov

Key Words: web survey, response effects

Abstract

At least into the near future, many self-administered web questionnaires are more likely to be used in multimodal surveys with paper and other modes, rather than as the sole method of data collection (Dillman, Tortora & Bowker, 1998). Therefore, a more thorough understanding of web-versus-mail and other mode effects is critical. In a 1999 opinion survey of 3,200 employees at a federal agency, both paper and web response options were offered. While mode was not an experimental treatment, and more respondents chose to complete the web version of the questionnaire, differences in response patterns are apparent. Four such modal differences are discussed. We compared the percentage of open-ended item response opportunities taken, and the average word count per response to those open-ended items across modes, and found that both were generally greater in the web version. Item nonresponse rates were compared and were found to be significantly lower in more web items than in corresponding paper items. Finally, average scores on 5 point attitude scale items were compared. Marginally higher ratings were observed for more web items than corresponding paper items. The paper reviews possible reasons for these differences.

Survey Background

Open-ended response volume and item response rate data was obtained from a 1999 survey of approximately 3,200 full-time employees of the U.S. General Accounting Office (GAO) administered by Schulman, Ronca & Bucuvalas, Inc. (SRBI). A similar paper version of this agency-wide employee survey had been conducted annually from 1989 through 1992, but not since.

All employees in the population were precontacted by email by with instructions and a password to access the survey website. Employees were asked to complete the survey either at work or at home. All GAO personnel have working email accounts and access to networked PCs. However, while most employees have their own PCs in their offices or workstations, some share access to PCs in group work areas. While most GAO employees are white-collar professionals accustomed to working with computers and the internet, some support staff are less familiar with computer technology.

After accessing the website, respondents could either complete the web version or request that a paper copy be mailed to their household. Several weeks after the start of fieldwork, employees not yet accessing the website to either begin the electronic version or request a paper questionnaire were automatically sent paper questionnaires. Later telephone followup with nonrespondents to encourage paper or web participation was also conducted.

Because mode of administration was not experimentally controlled, respondents could choose either, although they were encouraged through official communications and the way fieldwork was structured to use the web version. For example, several email precontacts and agency announcements encouraged them to use the web because it would save time and money by not requiring mailings. Also, the web version was made available at the start of the field period, while respondents wishing to answer on paper had to explicitly request this through the website, or had to take no action until a paper questionnaire was sent to them as part of nonresponse followup. Not surprisingly, 87% of responses came from the web version.

With 38 questions, some comprising many individual items, the average pretested length of the survey was approximately 45 minutes. The paper version was 17 pages long. While many of the items in the survey were straightforward attitude and behavior items about worklife, some were considered challenging, difficult, or sensitive by prestested respondents.

The web version of the survey was implemented using Quancept-Web and hosted on SRBI's website. While GAO management conducted some of the precontact and other communication with respondents, SRBI administered all other aspects of the survey.

Fieldwork was conducted in September and October 1999. Of the 3,243 eligible employees in the population at the time of the survey, 2,817 or 87% provided usable responses. 2,442 (87%) of those responses came from the web version, while only 375 (13%) came in the form of paper responses.

Modal Differences

Mode effects, or variations in survey response patterns (response effects) attributable primarily to the unique characteristics of a particular mode of administration, have been documented in many experiments comparing a variety of modes, more recently including the web (e.g., Robb et al., 1999). We looked for evidence of mode effects in terms of the following behaviors: 1) proportion of open-ended answer opportunities taken by respondents, 2) subsequent volume of open-ended narrative provided in those open-ended items, 3) item nonresponse rates on fixed-choice questions, and 4) distribution of responses over 5-point scale items.

Before testing our hypotheses on the expected differences on these behaviors, we examined how respondents who chose to answer through each mode differed, in terms of their demographic characteristics.

Demographic Differences Between Web and Paper Respondents

There were significant differences in the demographic characteristics of respondents who chose to respond by paper instead of the web. The 375 paper respondents tended to be older, more likely to have plans to retire in the near future, and were somewhat more likely to be in the lower pay grades. However, there was no significant difference in the proportion of managers versus nonmanagers who chose the paper option. Females tended to respond by paper at a slightly higher rate than males, as did nonwhites.

Mode Effect #1: Proportion of open-ended response opportunities taken

We measured how often respondents eligible to answer a particular open-ended item actually did so by providing at least one word of narrative. Since the web survey required a response on each question before moving to the next question, respondents were instructed to type "none" if they did not wish to provide a comment. Therefore, only positive responses other than "none" were counted in the analysis.

Hypothesis:

One possible effect of the two different modes could be the percentage of open-ended responses given by respondents, regardless of the comparative length of the response for web versus paper. We can assume, at least for those who can type, that typing is easier than writing for most people. We should therefore expect respondents completing the web survey to produce a greater proportion of open-ended responses, regardless of the length of response, than those completing the survey by paper.

Results:

On three of the five items that asked respondents to provide detailed narrative responses, the differences in response rates were in the expected direction: those using the web version were more likely to provide at least some narrative response. This difference was statistically significant for two of these items (Table 1).

The final two questions, which did not follow this pattern, were 1) an item which asked why employees chose to work at GAO that prompted answers tending to be short descriptors (e.g., "more money" or "job security") rather than extended narrative, and 2) an "other-specify" response blank which followed 13 short sentence pre-coded answers to a question asking why a respondent might be leaving the agency. The type of response elicited by the "other-specify" question is a hypothetical one which many employees have little need to consider. Roughly one-third (37 percent) of all respondents provided any answer to the entire question, while only 4 percent filled something in the other-specify blank that was different from what was already covered by the pre-listed response options.

This data also suggests that the *type* of open-ended question asked may encourage a greater proportion of web respondents than paper respondents -- questions designed to produce "negative" responses showed a significant difference in the proportion of respondents providing a response between survey modes. The two "negative" open-ended questions ("elaborate on any of your responses concerning any of the management teams" and "what things about GAO would you change and why") encourage employees to criticize the performance of their managers and of the agency itself. For these two questions, there were a significantly greater percentage of web responses attempted than there were for the paper version of the survey. The other three questions are more "neutral" ("why did you choose to work at GAO," "if you might leave GAO what are the reasons") or "positive" ("what things are working well"). The differences in the proportions completing a response between web and paper on these three "non-negative" questions were not found to be statistically significant. This suggests that the web tends to increase item response when the question asks for potentially detailed or involved responses, often associated with questions asking for negative criticism. Survey mode does not appear to affect item response when the question requires shorter or less detailed responses.

Mode Effect #2: Volume of Narrative Provided in Open-End Responses

We counted the number of words in narrative responses of at least one word (other than non-substantive replies such as "none" or "N/A") to the same five open-ended items.

Hypothesis:

Another potential mode effect is the relative length of the responses, or level of detail, provided by respondents to open-ended questions. We hypothesize that the web mode should produce a higher volume of narrative responses in those items that are answered.

Besides the previously mentioned reason that typing is easier than writing for most people, the web version also did not limit the number of words that could be entered. In effect, it allowed unlimited space for respondents to provide as much detail to their answers as they wanted. The paper survey, at least in theory, provided unlimited space as well provided that the respondent attached additional sheets of paper to the questionnaire. While some respondents did in fact attach an additional sheet, most restricted their comments to spaces provided on the questionnaire which ranged from 7 lines ("why did you choose to work at GAO") to 8 lines ("elaborate on any of your responses concerning any of the management teams," "what things about GAO would you change and why," "what things are working well") which translates to 2.0 to 2.5 inches of space per question. The exception was the "other specify" question, which provided only 1 line on paper, but an unlimited field on the web version.

Results:

The length of open-ended responses was greater in answers given through the web version -- all five comparisons showed differences in the expected direction. Three out of five of these comparisons showed significant differences in the mean word count between the Internet and the paper surveys (Table 2).

As in the preceding response rate section, for both of the "negative" questions, fairly substantial differences were observed, while the "positive" item ("what things are working well") produced no such difference. The web version did, however, produce a significantly higher word count than did paper on one of the "neutral" questions ("why did you choose to work at GAO").

Similarly, the web version did not impact response volume if the open-ended question was an "other specify" question to a pre-listed group of response options. Even though the word capacity of the web version was unlimited, respondents tended to restrict their comments to a few words regardless of the mode they used. This suggests that the fewer words required by a response, the less will be the difference between the web and paper survey modes on either average length of response or item response rates, as shown in the previous section.

Overall, the data on both of our open-ended item hypotheses suggest that the web version tends to increase the percentage of those responding to some questions and the length of their answers, especially if those questions are designed to produce detailed responses. This is especially the case with questions designed to elicit critically negative responses. The implication is that for some questions, particularly those that evoke highly critical or intensely-felt opinions, both item response and the length of the comments may be increased by using a web survey rather than a traditional paper survey, while use of the web may have little or no effect compared to paper on gaining responses when the questions are for more factual or limited word responses.

Mode Effect #3: Item Nonresponse rates Between Web and Paper Versions

In the paper version of the questionnaire, if a respondent wished to leave a fixed-choice item unanswered, he or she simply passed by to the next item or turned the page, leaving that item blank. However, the program governing the web survey required an actual keystroke to record data before moving to the next item, even if the datum recorded was just a non-substantive "no answer" code that respondents were told to use when they couldn't or didn't want to answer a question. (See Fig. 1.) There may be other reasons for item nonresponse in addition to the respondent not having a valid opinion or knowledge of the item or feeling threatened by a sensitive question. If even a small percentage of item nonresponse in a survey is due to "satisficing" -- that is, the cognitive burden of reading a question, comprehending it, formulating a judgment, and reporting that judgment outweighs the perceived value of reporting such an opinion -- then we might expect item nonresponse to be lower in the web version: a respondent to the web survey may reason that because they have to take an action (click a box) to advance to the next question, they might as well choose a substantive answer category instead of the "no answer" category. That is, there may be a segment of respondents to an item that actually do have an opinion,

however weakly held, that becomes no more burdensome to report than "no answer" because they have to click a response regardless. However, in the paper version, not answering may present a real savings in effort and time to this kind of "marginal" respondent.

Hypothesis:

We expected item nonresponse to be higher in the paper version than in the web version.

Results:

Item nonresponse rates on a total of 71 attitude items were compared. For the 2,817 useable questionnaire returns, mode (web or paper) was cross-classified with item disposition (missing, nonmissing), where nonmissing values were any substantive (e.g., a rating of 1-5) or nonsubstantive (e.g., "don't know," "too early to tell," etc.) answers. A missing value on the web version was only the explicit "No Answer" code, and only a blank in the paper version. While most (64 of the 71) comparisons were in the expected direction (paper nonresponse higher), significance testing was complicated by the large number of total observations and -- because nonresponse rates were low for most items -- the small number of observations in the paper/missing cell. More conservative test statistics appropriate for such comparisons were computed, and of the 71 differences:

31 were in the expected direction (paper nonresponse higher than web) and significant
32 were in the expected direction but not significant at the p<.05 level
7 were in the wrong direction (web nonresponse higher) but not significant
1 was in the wrong direction and significant at the p<.05 level.

Because of the overwhelming (if not always significant) tendency for paper item nonresponse to be higher, we felt this data upheld the hypothesis. This may be mostly due to the "required response" artifact of this web implementation, however. Another experiment involving a federal employee survey (Treat, 1997) found opposite effects (lower item response rates) for internet (email) versus paper.

Mode Effect #4: Average Ratings on 5-point Attitude Items:

Because the web version had to have an explicit 7th "no answer" response category, usually positioned to the far right of the 5 anchored scale radio buttons and the one

"don't know"/"too early to judge"/"no basis for opinion" or other nonsubstantive answer radio button, we theorize that the visual metric of such a scale might appear different from one that a paper respondent would see, which would only have 6 checkboxes. (See Figures 1 and 2 for an example.) If, as some measurement specialists have stated, respondents "learn" about the likely distribution of response meaning from the numbering and positioning of response categories (Schwarz, et al., 1985; 1991), those respondents wishing to record an "average" opinion on some item without making the cognitive effort to read and comprehend the actual scale used might reasonably select an answer checkbox near the middle of the apparent range of response options. That is, such a respondent might pick the 4th response in an apparent 7-point scale as the "middle," while he or she might choose between the 3rd and 4th response if the response options totaled 6 checkboxes. A similar measurement artifact might be expected when respondents wanted to pick responses that were "more positive" or "more negative."

Hypothesis:

Web respondents, presented with 7 answer options, should on average record higher answers (where the leftmost item is scored 1 and the rightmost item is 7) than should the paper respondents.

Results:

A total of 64 5-point scale opinion items were selected from the survey. None were embedded in any skip patterns or other context that might affect response patters differentially between the modes (e.g., the web version of an item on dealing with external customers might have a preceding filter question to determine whether the respondent dealt with external customers, while the paper version had no such filter). Responses of 6 (the nonsubstantive "don't know" or similar answer) and 7 (the explicit "no answer" response in the web version) were set to missing values, and the mean 1-5 rating was calculated. These average ratings were compared across modes.

Of the 64 comparisons, 45 (70%) were in the expected direction -- the average web rating was numerically higher than the paper rating, while the other 19 (30%) showed paper ratings that were at least marginally higher than the corresponding web rating.

Of the 64 comparisons, only 4 were in the expected direction (web rating higher than paper) and significant (p<.05), while 41 were in the expected direction but not significant. 18 were in the wrong direction and not

significant, and 1 difference was in the wrong direction and significant. Because we are comparing means of an ordinal 5-point scale across groups, and these distributions are not normal, nonparametric tests were used to compare the two groups.

Although a majority of the differences were in the expected direction, the statistically small magnitude of most of these differences limits conclusions that should be drawn from this data.

Discussion

Four examples of response differences between modes were demonstrated, to varying degrees of strength.

The first two mode effects, both involving open-ended response, were fairly straightforward and have been observed before. They both stem from the relative ease of entering narrative data with a keyboard rather than with paper and pencil, and the absence of any limitation on the length of open-ended responses in the electronic version. These differences were most pronounced in questions where detailed and potentially sensitive responses were common.

The last two effects -- fixed choice item nonresponse and ratings distribution -- are particularly interesting. In this survey, neither is an example of a "mode effect" *per se*. Both were most likely due to a certain artifact of administration: the web program requirement of having the respondent provide some data to the web server on each and every item before advancing to another screen with the next questions. Implementing the survey in this way may have made possible the effects of 1) marginal respondents providing substantive answers that they otherwise would not have, and 2) subjecting respondents to a response scale that

made a higher value response appear to be the midpoint of the response distribution than in the paper version. This is an example of how web technology, as it is typically implemented, may effect survey results in unanticipated ways.

References

Dillman, Don A., Robert D. Tortora, and Dennis Bowker (1998) "Principles for Constructing Web Surveys." SESRC Technical Report 98-50, Pullman, Washington.

Robb, William, Tammy Ouellete, Larry Luskin, and Marc Scott (1999) "A Direct Comparison of Web-based and Telephone-based Data Collection Modes." Paper presented at the 1999 Annual Conference of the American Association for Public Opinion Research. St. Petersburg, Florida, May 1999.

Schwarz, Norbert, Hans-J. Hippler, B. Deutsch, and F. Strack (1985) "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments," *Public Opinion Quarterly* 49:388-395.

Schwarz, Norbert, Barbel Knaeuper, Hans-J. Hippler, Elisabeth Noelle-Neumann, and Leslie Clark (1991) "Rating Scales: Numeric Values May Change the Meaning of Scale Labels," *Public Opinion Quarterly* 55:570-582.

Treat, James B. (1997) "The Effects of Questionnaire Mode on Response in a Federal Employee Survey: Mail versus Electronic Mail," *Proceedings of the Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, pp. 600-604.

Table 1: Proportion of Open-Ended Response Opportunities Taken by Survey Mode

Question	Web: Percent of opportunities taken	Paper: Percent of opportunities taken
"If you would like, elaborate on any of your responses concerning any of the management teams..."	**39.9%***	31.2%
"What things about GAO would you change immediately if you could, and why?"	**83.7%***	78.4%
"What things are working well in GAO?"	**70.6%**	68.8%
"Why did you choose to work at GAO?"	79.0%	**80.0%**
"If you think you might leave GAO in the near future, what are the most likely reasons you might do so? -- Other (please specify)"	3.9%	**4.3%**
Unweighted N	2,442	375

*Difference significant at p<.05

Table 2: Mean Word Count by Survey Mode

Question	Web: Mean word count per response (if >0 words) (unweighted n)	Paper: Mean word count per response (if >0 words) (unweighted n)
"If you would like, elaborate on any of your responses concerning any of the management teams..."	131.9* (975)	96.3 (117)
"What things about GAO would you change immediately if you could, and why?"	91.5** (2045)	60.5 (294)
"What things are working well in GAO?"	30.9 (1724)	27.0 (258)
"Why did you choose to work at GAO?"	31.6** (1928)	22.8 (300)
"If you think you might leave GAO in the near future, what are the most likely reasons you might do so? -- Other (please specify)"	14.3 (96)	9.9 (16)

* Difference significant at p< .05   ** p<.001

Figure 1: Typical Likert Item in **Web** Version

Figure 2: Typical Likert Item in **Paper** Version