

## INTERNET VERSUS MAIL AS A DATA COLLECTION METHODOLOGY FROM A HIGH COVERAGE POPULATION

Nancy Bates, U.S. Census Bureau, Demographic Surveys Division  
Washington, DC 20233

### 1. Introduction

In 1998-1999, various estimates of the adult population with internet access ranged from a low of 27 percent to a high of 53 percent (see summary of estimates in Couper, 2000). Access is still far from universal yet continues to grow rapidly since the introduction of the World Wide Web. Consequently, survey methodologists are turning to this mode as a primary or supplemental method for self-administered survey data collection. To date, the survey methods literature contains few quantitative studies that document both advantages and disadvantages of this new mode.

When used as a primary mode of data collection, most studies show response rates to e-mail or internet surveys below those obtained by the more traditional mail method (Schuldt and Totten, 1994; Kittleson, 1995; Tse et al 1995; Couper, Blair and Triplett, 1999; Kwak and Radler, 2000). However, when used thoughtfully within a multimode implementation strategy, Schaefer and Dillman (1998) were able to achieve an e-mail response rate equal to that of mail among a population of university faculty.

Information about the speed of internet survey response is more definitive. Studies unanimously report that internet surveys have a quicker turnaround time than mail surveys. Kwak and Radler (2000) reported an average of 9 days for mail compared to around 2 for internet. Schaefer and Dillman (1998) reported 14 days on average for their mail panel compared to 9 days for internet. Finally, Schuldt and Totten (1994) reported a much higher internet to mail ratio of returns (9:1) in the early days of data collection.

In addition to response rate differences, internet/e-mail surveys introduce questions about data quality. A review of the literature yields mixed reports. Some experimental studies report higher item nonresponse rates for e-mail (Sproull, 1986; Bachmann, Elfrink, and Vazzana, 1996), while others report very little or no difference (Couper, et al. 1999; Mehta and Sivadas, 1995; Tse et al., 1995) and still others report lower item nonresponse and higher quality response to open-ended questions in e-mail/internet surveys (Shafer and Dillman, 1998; Kwak and Radler, 2000).

Response by internet may also affect data content. For example, if the demographic profile for internet respondents is unique from that of mail respondents, we might see substantive differences to survey answers by mode. This is important to know if

data users plan to study construct associations, that is, the inter-relationships between survey variables. In their organizational climate survey of government statistical agencies, Couper et al.(1999) found that e-mail respondents rated their agency more positive than mail respondents once differences in the sample compositions were controlled. Likewise, Kwak and Radler (2000) found significant differences in answers to technology-related questions from mail versus Web respondents. A recent study by the Pew Research Center (Flemming and Sonner, 1999) compared opinions gathered from a Web survey to those from a Random Digit Dial telephone survey. They reported finding large deviations in some response distributions by mode yet failed to pinpoint consistent patterns to explain the differences.

In our study, we examine the issues of internet response rate, response speed, respondent characteristics, data quality, and data content by analyzing a mode experiment. The study population is employees of a government statistical agency, the vast majority of which had job-related internet access at the time of the survey (approximately 88 percent).

### 2. Experiment Methodology

In October-November of 2000, the U.S. Census Bureau conducted an organizational assessment survey of agency employees. The survey instrument was a questionnaire developed and maintained by the Office of Personnel Management (OPM). It asked employees to rank the agency on topics such as rewards and recognition, training and career development, work environment, and job security. The assessment was closer to a census than a survey because it included all employees at the headquarters office, all regional offices, the telephone interview facilities, and the data processing center. Field interviewers and Census 2000 temporary hires were the only employees purposely excluded from the assessment. As part of the survey, the agency embedded a small mode experiment within the roster of headquarters employees with internet access. We did so to test response rate, data quality, and respondent characteristic differences between a mail paper/pencil questionnaire and an internet version.

As described in Couper (2000), our type of internet survey is best labeled a probability based list sample from a high coverage population. Prior to drawing the sample for the mode experiment, we sorted all headquarters employees into two groups: those with

internet access and those without. Those without internet access (about 12 percent) were excluded from the experiment and assigned a paper/pencil questionnaire designated as non-experimental. By restricting the sample frame to those with known internet access, we controlled for internet coverage bias and hopefully removed this confounding factor common to many internet surveys.

On the down side, however, restricting the frame meant our experiment could only be generalized to populations known to have internet access, specifically, internet access at the workplace. The internet access pool of employees were sorted and randomly sampled into one of three panels. The first was assigned a paper and pencil questionnaire, the second an internet questionnaire and the third group was assigned to a shorter paper/pencil survey that had been used previously at the agency. The third questionnaire was used for historical comparisons and was not part of the mode experiment (and thus not reported here). The experimental portion of the sample was comprised of 1,645 paper-assigned cases and 1,644 internet designated cases.

The paper questionnaire was 18 pages long and consisted of 176 questions/statements divided into 21 topical sections. It concluded with a 15 question section on background and employment information. With the exception of the demographic questions, employees could answer all questions by checking a box along a 5 point Likert scale. All scales were bi-polar with end points like 'very dissatisfied' and 'very satisfied' or 'strongly disagree' and 'strongly agree'. The midpoints were labeled and neutral, for example, 'neither dissatisfied nor satisfied'. For the majority of items, the questionnaire also offered a 'don't know' category.

The internet version was an HTML form that required a web browser of 4.51 or above. The contractor designed the form for Netscape browsers since this is the agency standard. The internet version contained the same questions as the paper version but respondents used radio buttons to select response options. The questionnaire was divided into 22 'pages' with each page containing a battery of questions. Respondents navigated the form by using the scroll bar to move up and down within a page and NEXT and PREVIOUS buttons at the bottom of the form to move between pages. Respondents used a SUBMIT button at the end of the form to send in their completed questionnaires. The form also provided a QUIT option for respondents wanting to quit part way through, save their answers, and complete the form at a later time. Respondents choosing to do so were required to provide an e-mail address so a user code could be supplied for re-entry.

Prior to the survey, the human resources division (HRD) raised awareness with posters, broadcast e-mail

messages, and drop-in articles in employee newsletters. On October 18<sup>1</sup>, the Census Bureau delivered a packet of survey materials to each employee's workstation. The sample of employees in the paper panel received a letter from the Census Bureau director explaining the purpose of the survey, the mode experiment, and a message indicating they had been selected to complete the enclosed paper questionnaire. The packet also contained a postage-paid return envelope addressed to the contractor. Employees selected for the internet sample received a similar letter but were told they had been selected to complete their survey via the internet. The letter supplied the internet address, a survey user identification and survey password. Survey user identifications and passwords were unique at the division level.

Two weeks after the initial mailout, the Bureau mailed a blanket reminder card (paper). Employees in the paper panel were reminded to complete and mail back their surveys and instructed that replacement surveys could be obtained from their administrative offices. For the internet sample, the card supplied the internet address but also gave the option of completing a paper questionnaire obtained from the administrative office. For both panels, the surveys were completely anonymous and lacked any type of unique identification for tracking purposes. Instead, each questionnaire type was captured during data processing according to a form-type identifier (original paper questionnaire, internet questionnaire, or replacement paper questionnaire).

While the anonymous design may have helped encouraged response, it limited the nonresponse follow-up options and response rate analysis. Without identifiers, we could not target reminders nor track individual responses by panel assignment. Consequently, employees who lost or misplaced their original paper questionnaire and completed the paper replacement had to be dropped from the mode analysis (because their original panel assignment could not be determined). Likewise, employees in the internet panel who opted to pick up and complete a paper replacement also had to be excluded. Finally, the lack of identifiers made it possible for a respondent to mail back more than one paper questionnaire (or submit multiple internet responses).

---

<sup>1</sup> The contractor delivered a complete set of survey packages to Census prior to October 18. However, after printing errors and technical problems with the internet version were discovered, a decision was made to halt distribution, reprint the job, correct the internet instrument, and redistribute a new set on October 18. The recall was successful for all but two divisions who inadvertently delivered the original packets. Thus, a number of employees (127) received duplicate mailings.

Given the extreme length of the questionnaire, however, we believe this is unlikely.

Throughout the data collection period, employees received periodic reminders by way of broadcast e-mail announcements, division-level e-mails, reminders on pay-stubs, and announcements that the daily division-level response rates could be viewed on the agency intranet. The HRD also set up a special room for two days where employees could complete their surveys without interruption. A division-level contact was also designated to answer questions and encourage response.

### 3. Results

#### 3.1. Response rates and speed

We calculated response rates by dividing the number of usable questionnaires received or postmarked by 11/22 by the original sample number minus the number of ineligible cases. Ineligible cases consisted of mail packages that were returned unopened to HRD because the employee had resigned, was intermittent and not at work, a summer employee or intern, or on extended leave at the time of the survey. Such cases reduced the number of eligibles to 1,569 for the mail panel and 1,571 for the internet panel.

Following closeout, the contractor examined the internet database and removed several duplicate and blank records. Seventy-nine records were completely blank and another 81 were exact duplicates of previous submissions. None of the duplicates came from either of the two divisions that received the mailing packages prematurely, therefore, we hypothesize they resulted from technical problems, not the duplicate mailing. Finally, a total of 17 internet records were determined to be 'insufficient partials' and categorized as nonrespondents. These records contained 121 or more blank answers out of the 161 organizational assessment questions (75 percent or greater had missing data). We conducted a similar review of paper insufficient partials and treated these as nonrespondents in the response rate calculations.

Table 1 contains the survey response rates and response speed by mode. Overall, the internet had a significantly higher rate of return compared to the mail panel (66.6 percent versus 62.8 percent, respectively). Prior to the reminder card, the response rate for the internet panel was also higher than the mail questionnaire. Before the reminder went out, around 84 percent of the internet submissions had already been received compared to 70 percent of the mail returns. Thus the internet had a faster rate of return than mail (on average, it took 9 days for mail returns to arrive prior to the reminder compared to only 3 days for internet submissions). After the reminder card, the response rates reversed and mail had the higher rate. The mail panel experienced a significant response spike a few days after the card arrived with

returns slowly decreasing thereafter while the internet panel exhibited a much more modest spike at the reminder card and a quick decline thereafter. These results are likely an artifact of both the methodology and the types of people left in the respective samples at the time of the reminder card.

#### 3.2 Respondent characteristics

We next explored whether certain groups might be under-represented among internet respondents. Couper et al. (1999) reported that males, non-minorities, and those in higher pay grades were more likely to respond to an e-mail survey than mail in a survey of government statistical agencies. Likewise, Kwak and Radler (2000) report that younger males tended to be relatively over-represented in their web survey respondents from a university student population. To determine whether respondent characteristics differ by mode of response in our study, we examined selected demographics (sex, age, race, education, supervisory status and grade level).

Of the six characteristics examined, absolute differences between modes were small. The internet appeared to represent slightly more males, persons aged 30-39 and persons in higher grades than mail responses. However, none of these differences were large enough to be statistically significant and we conclude that while more people responded to the internet version, the types of people who responded were basically the same as those responding by mail.

#### 3.3 Data quality

To every extent possible, the internet questionnaire mirrored the layout and format of the paper version. All questions were in-scope for all respondents therefore neither survey contained skip instructions and the internet version contained no automatic edits. Provided the similarities in questionnaire design, the random assignment of mode, and the similarity of respondent characteristics by mode, we hypothesized that the item nonresponse rates would be similar between the internet and paper returns.

We examined two indicators of data quality measured by level of missing data. First, we calculated a distribution of organizational assessment item completion rates by mode (table 2). The 1-24% category reflect returns that contained fewer than 25 percent answers to the organizational assessments questions (organizational items + personal experience items). These were subsequently defined as "insufficient partials" and excluded from the numerator in response rate calculations. The internet had a higher occurrence of these cases (eight times the percent of the paper version). This could be indicative of respondents who started the

survey, decided to exit before finishing, but then never returned to complete it. Alternatively, it could reflect persons who paged through the electronic version in one sitting, answering a few questions here and there before finally submitting a less-than-complete record. It is plausible that the internet survey is more likely to be left unfinished if not completed at one sitting since it required respondents to exit the survey, wait for a user code to be returned, re-access the internet at a later time, enter the new user code to log back into the survey, and finally resume and complete the partially completed questionnaire.

Table 3 contains the average item missing rates for the three sections of the questionnaire (answers of 'don't know' were considered non-missing entries). These averages are based upon records that completed 25 percent or more of the organizational assessment items. Contrary to our original hypothesis, we found significant differences in item missing rates for both the organizational assessment section (questions that asked for overall impressions of the agency) and the section dealing with personal experiences (questions asking about the employee's own personal experiences within the agency). For both sections, the internet cases had a higher average percent of item nonresponse than the mail (3.0 percent versus 1.71 percent; 1.46 percent versus 1.03 percent). We found no differences in item nonresponse by mode for the employment and demographic questions<sup>2</sup>.

Our final examination of data quality focuses on the content of answers provided by the two modes of response. If, as Couper et al. (1999) suggest, persons holding a more negative view of agencies are less likely to answer organizational assessment surveys, we would expect the panel with the lower response rate to reflect a more positive assessment of the agency. In our case, this would translate into higher satisfaction ratings from the mail panel. However, the random assignment of mode coupled with the absence of respondent characteristic differences noted earlier might conversely support the null hypothesis -- that response distributions will not differ across modes. We examined the subtopic mean scale scores for the organizational and personal assessment portions of the questionnaire.

The majority of assessment subject mean scores did not significantly differ across modes (6 of the 22 were statistically significant). However, in all six cases where

---

<sup>2</sup>This section was last in questionnaire and asked about: length of government service, tenure at agency, agency division, pay grade, pay category, supervisory responsibility, age, gender, race, ethnic background, job category, education, disability status, child care responsibilities, and elder adult responsibilities.

differences did exist, the internet respondents consistently gave more favorable assessments compared to mail respondents. This is somewhat perplexing considering we found no differences between age, sex, race, pay grade, education or supervisory status of internet versus mail respondents and that the direction of the difference is contrary to the idea that the mode with lower nonresponse yields lower satisfaction scores. But, the fact that relatively few topic mean differences were found and that only two were significant beyond the .05 level may suggest a broader conclusion of little or no real response differences or patterns by mode.

#### 4. Discussion

Our study randomly assigned a sample of federal agency employees with confirmed internet access into one of two panels. The first panel received a paper version of an organizational assessment survey and were instructed to mail it back. Employees in the second panel were instructed to access an internet site and complete the questionnaire online. The internet response rate surpassed the mail response (66.6 percent compared to 62.8 percent). Several factors may have contributed toward the success of the internet panel. First, employees with internet privileges have access to high speed web connections conveniently available at their own desktop computers. Second, the survey was designed to be compatible with a Netscape browser, the Census Bureau standard. Finally, the sample had a very high education level and contained many people who use computers and the internet as a routine part of their jobs.

The internet exhibited a faster turnaround time with an average of 6 days overall compared to 13 for mail. Over half (53 percent) of the internet submissions were received within the first 3 days of the data collection. The reminder card had less of a booster effect for the internet panel compared to the paper -- presumably because the nonrespondent pool was smaller for the internet at the time the card went out. The smaller effect for internet could also have resulted from using a paper reminder (as opposed to e-mail with a hyperlink to the survey) as the delivery method for the follow-up.

While the internet panel garnered more responses, the types of people who responded were similar to those who responded by mail. However, we found significant differences in the rates of missing data between modes -- the internet had more insufficient partials and a higher rate of item nonresponse to the organizational assessment questions. We also discovered a few instances where content of response varied by response mode.

Findings from our experiment offer encouraging support for the internet as a self-administered data

collection mode among a high coverage population where internet use is part of the daily work routine. No efforts were made above and beyond the mail panel to encourage internet response. For example, the internet panel used a mailing piece similar to the mail panel to announce the survey and instruct respondents where to access the internet questionnaire – it did not make use of e-mails with embedded hyperlinks to encourage response. Nor did the internet panel require special follow-up procedures – it made use of a paper reminder card just as the mail panel did. Nonetheless, the internet panel achieved a higher response rate and required less time to administer. Both factors have desirable cost implications when conducting a survey.

Additionally, unlike previous studies, we failed to uncover glaring demographic differences in the mail versus internet respondent populations. This is good news when considering the addition of the internet in a multiple mode survey or when switching from mail to internet as the primary response vehicle. However, our study also suggests that data quality may be an area of caution when it comes to the internet. Although differences were not catastrophic, the internet had more missing data on the opinion items compared to mail responses. This may be a function of the questionnaire’s length, but should be noted regardless. Our study also suggested a slight bias in the content of answers provided by internet respondents – this too should be considering when debating the internet as a data collection method.

References

Bachmann, D., Elfrink, J. and Vazzana, G. (1996). Tracking the Progress of E-mail versus Snail-mail. *Marketing Research*, 8:31-35.

Couper, M. (2000). Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*, 64:464-494.

Couper, M., Blair, J., and Triplett, T. (1999). A Comparison of Mail and E-mail for a Survey of Employees in U.S. Statistical Agencies. *Journal of Official Statistics*, 15:39-56.

Flemming, G. and Sonner, M. (1999). Can Internet Polling Work? Strategies for Conducting Public Opinion Surveys Online. Paper presented at the annual meeting of the AAPOR, St. Petersburg, FL.

Kittleston, M. (1995). An Assessment of the Response Rate via the Postal Service and E-mail. *Health Values*, 18:27-29.

Kwak, N. and Radler, B. (2000). Comparison between mail and Web surveys: Response Pattern, Respondent Profile, Data Quality, and Construct Association. Paper presented at the annual meeting of the American Association for Public Opinion Research. Portland, Oregon.

Mehta, R., and Sivadas, E. (1995). Comparing Response Rates and Response Content in Mail versus Electronic Mail Surveys. *Journal of the Market Research Society* 37:429-39.

Schaefer, D. and Dillman, D. (1998). Development of a Standard E-Mail Methodology: Results of an Experiment. *Public Opinion Quarterly*, 62:378-397.

Schuldt, B. and Totten, J. (1994). Electronic Mail versus Mail Survey Response Rates. *Marketing Research*, 6:36-39.

Sproull, L., (1986). Using Electronic Mail for Data Collection in Organizational Research. *Academy of Management Journal*, 29:159-69.

Tse, A., Tse, K., Yin, C. Ting, C, Yi, K. Yee, K. and Hong, W. (1995). Comparing Two Methods of Sending Out Questionnaires: E-mail versus Mail. *Journal of Market Research Society*, 37:441-46.

Table 1.  
Comparison between Mail and Internet Surveys: Response Rates and Response Speed

	Rec'd. by Day:	Mail	Internet	Difference <sup>1</sup>
Response rates:				
Overall	36	62.8% <sup>2</sup> (986) <sup>3</sup>	66.6% (1046)	- 3.8 *
Before reminder	14	44.2% (694)	55.7% (875)	-11.5 ***
After reminder card	15-36	18.6% (292)	10.9% (171)	7.7 ***

Table 1 (con't.)

Response speed <sup>4</sup> (mean no. days):	Mail	Internet	Difference
Overall	13 days	6 days	7 days***
Before reminder card	9 days	3 days	6 days***
After reminder card	9 days	5 days	4 days***

\*  $p < .05$ , \*\*\*  $p < .001$

Notes.

1. Test statistics for response rate differences are  $X^2$  tests and those for response speed are t-tests.
2. When computing response rates, the final samples sizes are 1,569 for the mail survey and 1,571 for the internet survey.
3. Actual number of questionnaires returned at each point in time are in parentheses.
4. When response speed was computed, 11 mail survey cases with return dates prior to the official start date were excluded from the analysis. These are cases from the two divisions that inadvertently received survey packages early.

Table 2.  
Item Completion Rate by Survey Mode (Organizational Assessment Questions)

	Mode			
	Mail		Internet	
% Questionnaire Completed	%	N	%	N
0-24% (insufficient partials) <sup>a</sup>	0.2	2	1.6	17
25-49%	0.2	2	1.7	18
50-79%	0.7	7	1.2	13
80-89%	1.7	17	0.4	4
90-100%	97.2	960	95.1	1010

<sup>a</sup> Insufficient partials were defined as having too much missing data to be considered complete cases and were defined as eligible, non-interviews in response rate calculations.

Table 3. Mean item missing data rates by mode (percents)

	Mail		Internet	
	mean	(s.e.)	mean	(s.e.)
115 organizational items	1.71	(.08)	3.00	(.10)***
46 personal experience items	1.03	(.05)	1.46	(.06)***
15 demographic/employment items	2.92	(.14)	2.59	(.13)
(n)	(986)		(1,046)	

\*\*\*  $p < .001$