Using an E-Mail Invitation to Screen Survey Respondents

(Timothy Triplett, Natalie Abi-Habib, Adam Safir)

Abstract

This article reports the results of an experiment in which approximately half of a randomly drawn split ballot sample was asked a key screener question in their initial e-mail invitation, while the second half of the sample was asked the same question at the beginning of the survey instrument. The sample consisted of registered users of the National Survey of America's Families (NSAF) online public use data who agreed to be contacted for survey purposes. The expectation was that placing a screener question in the e-mail invitation would increase the overall number of respondents, but that it could also lead to an increase in the number of eligible respondents incorrectly screening themselves out, whether intentionally or otherwise. However, the experiment showed that asking the screener in the e-mail invitation had no effect on the response rate, proportion of eligible respondents screening out, or survey estimates.

1. Introduction

Conducting mail surveys is less expensive then phone or in-person surveys (Hatry et al. 1998); hence most customer surveys have traditionally used mail as the primary mode of data collection. Customer surveys often use a screener question to determine respondent eligibility and reduce the number of incorrectly completed questionnaires. However, respondents who correctly screen out on the first question rarely return their questionnaire, perhaps believing that they provide too little information to be of use. Not receiving returns from those who screen themselves out does not have much impact on survey results, but does undermine the ability to

accurately calculate response rates, in addition to losing information that could improve the sample frame.

Customer satisfaction surveys are increasingly moving from traditional mail to web based data collection, since web based data collection can provide instantaneous results and save on the postage, printing, and labor costs associated with mailing and data entry (Salvucci, Parker, and Wenck 2002). Transitioning from the traditional mail survey mode to web-based surveys provides many more options for the designer (Couper, Traugott, and Lamias 2001). One new design option is the ability to include skip patterns that can help reduce the number of failed returns due to perceived respondent burden Information from respondents who fail to return their survey form after screening out at the initial questionnaire item can be captured in the web survey mode because of the medium's ability to seamlessly record transactions between the respondent and the survey database. The web survey mode largely reduces respondent burden and improves the ability of researchers to calculate response rates and gather information about their sample. However, whether administered via the web or in the traditional mailed paper format, customer surveys based on a frame of registered users can suffer non-response as a result of delivery to non end-users. For example, a registrant may not be the end-user, or may have completed the registration process for exploratory purposes only. Consequently, in the web survey mode there is some concern that the non end-user is unlikely to follow the hyperlink to the Internet survey, thus never seeing the screener question. An eligible respondent may conclude that the survey subject is not relevant to their situation, and that the perceived burden of reviewing the survey content is not worth their time. Researchers may be able to circumvent this problem by providing the screener question within the e-mail invitation. Including the screener question in the e-

mail invitation may improve response rates as well as provide researchers with additional information about respondents. However, a concern is that placing the screener question in the e-mail invitation can be perceived by potential eligible respondents as an easy way to 'opt-out' of the survey request.

2. Data Source and Methods

The Urban Institute's Assessing the New Federalism (ANF) project periodically invites its registered data users to complete a user satisfaction survey. The web survey is accessed and tracked using an e-mail invitation and web survey form. In the past, the screener question was not included in the e-mail invitation. However, the results of previous surveys indicate that a significant portion of registered data users do not actually use the data. For the February 2004 survey we conducted an experiment, providing a random half of our respondents the screener question in the e-mail invitation. In this paper we will refer to respondents who received the screener question in the e-mail invitation as the experimental group and those who received the screener question in the survey form as the control group.

The wording of the screener question was identical for both the control and experimental group:

> *"Have you ever used the National Survey of America's Families (NSAF) data files or the online analysis tools?"*

In the experimental group, this question was included within the e-mail invitation text, presenting respondents with a separate link to follow if they had not used the NSAF data files or online analysis tools. After clicking this link, respondents were taken to a web page where they were thanked and told that their e-mail address would be removed from the survey list. Respondents in the control group were not presented with the separate link, rather they were simply asked to

participate in the survey by clicking on the survey link provided. After proceeding to the web survey form, the above screener question was asked (see Figure 1).

3. Hypotheses

In this paper we tested the following hypotheses:

2.1 Including a data use question on the e-mail invitation would result in a higher response rate.

2.2 Providing respondents with a data use question in the e-mail invitation would result in a greater number of eligible respondents incorrectly opting-out, whether inadvertently or intentionally.

2.3 Placing a data use question on the e-mail invitation, instead of including it as the first survey question, would not affect the distribution of respondent characteristics or other aspects of sample composition.

Figure 1. Initial and Reminder E-mail Invitation Wording, by Experimental Group

	Experimental Group	Control Group
FIRST E-MAIL	We are conducting a survey with registered users of the National Survey of America's Families (NSAF) data resources website who agreed to be contacted for survey purposes. The Urban Institute would like to measure users' satisfaction with and the effectiveness of public use datasets. Our goal is to improve how the data are made available to you. The survey takes 3-5 minutes to complete and your participation is greatly appreciated. Please take a few moments to fill out the questionnaire by clicking on the link below: [SurveyLink] If you have not used NSAF datafiles or the online analysis tools please click on the link below: [Screen-OutLink] Best regards, Natalie Abi-Habib If you have any questions regarding the survey please contact me at: nabihabi@ui.urban.org	We are conducting a survey with registered users of the National Survey of America's Families (NSAF) data resources website who agreed to be contacted for survey purposes. The Urban Institute would like to measure users' satisfaction with and the effectiveness of public use datasets. Our goal is to improve how the data are made available to you. The survey takes 3-5 minutes to complete and your participation is greatly appreciated. Please take a few moments to fill out the questionnaire by clicking on the link below: [SurveyLink] Best regards, Natalie Abi-Habib If you have any questions regarding the survey please contact me at: nabihabi@ui.urban.org

REMINDER E-MAIL	Last week you should have received an e-mail asking for your participation in the ANF & NSAF Registered User Survey. We are still accepting responses and would greatly appreciate your feedback on the ANF data resources website. Your feedback will aid us in improving how public use data are made available in the upcoming release of 2002 data and documentation. The survey is very brief and should take 3-5 minutes to complete. To participate, please click on the link below. If you have already filled out your questionnaire, we thank you for your response and ask that you please disregard this message. [SurveyLink] If you have not used NSAF datafiles or the online analysis tools please click on the link below: [Screen-OutLink] Best regards, Natalie Abi-Habib If you have any questions regarding the survey please contact me at: nabihabi@ui.urban.org	Last week you should have received an e-mail asking for your participation in the ANF & NSAF Registered User Survey. We are still accepting responses and would greatly appreciate your feedback on the ANF data resources website. Your feedback will aid us in improving how public use data are made available in the upcoming release of 2002 data and documentation. The survey is very brief and should take 3-5 minutes to complete. To participate, please click on the link below. If you have already filled out your questionnaire, we thank you for your response and ask that you please disregard this message. [SurveyLink] Best regards, Natalie Abi-Habib If you have any questions regarding the survey please contact me at: nabihabi@ui.urban.org

We anticipated that including a data use screening question on the e-mail invitation would encourage and facilitate response from those who otherwise might not have responded at all, believing that as a non-data user, the survey was not intended for them.

With respect to the second hypothesis, we expected that some data users would perceive the screener question as an opt-out link and use it to avoid the burden associated with completing the questionnaire. Additionally, we expected that a small share of respondents would click the wrong link inadvertently, or might misunderstand the question.

Finally, we did not anticipate observing a difference in the characteristics of the two groups, expecting that the group of respondents who screened out on the e-mail invitation would be the same type or class of respondents proceeding with the survey and responding “No” to the data use question in the control group.

4. Experimental Study Design

In order to gain access to the NSAF data resources, users are required to complete a registration form. After registering to use the data, users receive a username via e-mail, which can be used to access downloadable public use data files and several online data analysis tools.

The registration form asks users to provide information about their research interests, the data resources they plan to use as well as basic contact information including e-mail address. There is also a question asking permission to contact the user:

> *In an effort to increase the ease of use of the data, we will periodically contact a random selection of users to learn about their unique data needs. May we contact you?*

The population for the NSAF Data User Survey consisted of 938 registered users who agreed to be contacted and who had accessed the NSAF public use web site between July 1, 2003 and January 31, 2004. The control group sample had a total of 467 e-mail addresses and the experimental group totaled 471 e-mail addresses. Both groups had a small percent of non-working addresses that the survey software removed from the sample prior to the mail out of invitations: 3% in the control group and 5% in the experimental group. There was also a number of non-working addresses that the survey software failed to identify. The invitations sent to these addresses were returned undeliverable. The total sample excluded a small number of eligible users who participated in a previous pilot survey, but included 22 e-mail addresses that were in the pilot sample but did not participate (see Table 1).

Table 1. Sample Composition

	Control Group	Experimental Group	Total
Sample Size	467	471	938
Undeliverables	16	26	42
Number of Addresses in Pilot	10	12	22
Response Rate (%)[+]	31.0	29.0	30.1

[+] $X^2=0.203$, $df=1$, $p>.05$; the relationship between experimental group and response rate is not significant

The survey was administered using a web form, accessed through an e-mail invitation that included an HTML link to the survey and a unique identifier for each respondent. The first e-mail invitation was sent the morning of Wednesday, February 4, 2004. A reminder e-mail was sent to all non-respondents the afternoon of Thursday, February 12, 2004. The initial and follow-

up invitations were sent out on different weekdays and at different times of day to facilitate contact with registered data users (Newcomer, Triplett 2004).

Control group e-mail invitations included a paragraph about the survey and its objectives, followed by a single survey-link to access the survey. Once the survey was accessed, these respondents were asked a screener question to determine if they had ever used the NSAF data files or online analysis tools. The experimental group received the same e-mail invitation with a survey-link to access the survey, although they were also provided with a screen-out link to be clicked if they had not used the data. Respondents who clicked the screen-out link were taken to a web page that told them they would not receive further requests to complete the survey.

5. Data Analysis

In order to test our three hypotheses we analyzed the data for differences between the control and experimental groups. First, we looked at the overall response rates achieved by each group, then at the distributions of the individual survey items by each group, and finally we compared the results of an item non-response analysis conducted for both groups. None of these analyses revealed any difference between the control and experimental groups.

The overall response rate for the survey was 30.1%. Differences between the sample groups were not much different, with the control group achieving an overall response rate of 31.0% and the experimental group achieving a response rate of 29.4%. These rates are calculated using a conservative method, which removes from the denominator only non-working e-mail addresses. The discovery of a lower response rate for the experimental group, though statistically there was no difference renders our first hypothesis null: Including a data use question on the e-mail invitation did *not* result in a higher response rate for the experimental group over the control group.

This leads us to the investigation of our second and third hypotheses, that certain eligible respondents in the experimental group may use the screener question hyperlink to effectively

"opt-out" of taking the survey, and that respondent characteristics as measured by the survey items would differ between the two groups. The lack of a difference in overall response rates between the two groups suggests that the former did not occur, however, without a follow-up survey to measure characteristics of these respondents it is impossible to know if some of them are actually NSAF data users. In a simple analysis of each groups' responses to every survey item, there is little to suggest a difference in the characteristics or habits of each groups' use and opinions of NSAF data resources. The distribution of responses to each item is relatively similar between the control and experimental groups. Where large differences occur they are not significant at the 95% confidence level. The lack of significant differences for survey items between the two groups may be a function of the limited sample size; some cell sizes are too small for meaningful analysis (e.g., 5 respondents or fewer).

Table 2. Various Estimates by Experimental Group (Percents)

	Control Group	Experimental Group	Total Sample
Heavy vs. Light Users of NSAF			
Heavy (20+ hrs in past year)	31.8	38.3	35.0
Light (<20 hrs in past year)	67.1	62.0	64.5
R skipped question	1.2	0.0	0.1
Satisfaction with NSAF Documentation			
Very Satisfied	31.3	28.2	30.0
Somewhat Satisfied	55.0	58.0	56.3
Not Too Satisfied	5.0	4.2	4.6
Not at All Satisfied	0.0	1.4	0.1
Never Used Documentation	6.3	4.2	5.3
R Skipped Question	2.5	4.2	3.3
Familiarity with Constructed NSAF Variables			
Yes	55.0	58.0	56.3
No	43.0	37.0	40.0
R Skipped Question	3.0	5.6	4.0
Organization Affiliation			
Federal/State/Local Government	7.5	7.0	7.3
College or University	52.5	57.7	55
Other Non-profit	20.0	14.1	17.2
Private Company	6.3	4.2	5.3
Other	3.8	0.0	2.0
R Skipped Question	10.0	16.9	13.2

In additional attempts to find differences between the control and experimental groups, an item non-response analysis was conducted in which the average number of missing items was calculated for each respondent based on the number of items he/she should have answered (a factor dependent on each respondent's skip pattern). For example, a respondent who intentionally or accidentally skipped five questions in a skip pattern where he or she should have answered 10 questions would have a item non-response rate of 0.5. Within each experimental group the average item non-response rate is taken across all respondents in that group and is reported here as the mean item non-response rate. We found that the experimental group had a slightly higher mean item non-response rate than did the control group (19.1% vs. 14.7%, respectively), but this difference was not significant at the 95% confidence level.

6. Issues and Limitations

Web surveys administered via an e-mail invitation face several obstacles and as a result are often used as a method for supplementing mail surveys in an effort to improve survey response rates (Keesling, 2002). Obstacles include the need to penetrate inboxes overrun with spam and junk-mail, as well as the need to assure potential respondents that the survey link is safe and free of viruses. In the week prior to the initial mail-out of the NSAF Public User Survey's e-mail invitation, a virus scare made headlines around the world. The virus called MyDoom or Novarg is downloaded to a PC when the recipient opens an e-mail attachment file. Some of our respondents expressed concern over virus risks; one respondent even directly replied to our e-mail invitation asking how and why she had been sent the invitation:

> *"Since I am not sure how I got registered and am not aware that I use your service, and because of virus concerns, please tell me more about why [I] should complete this survey. Thanks."*

Saliency of the NSAF was not strong for this particular respondent but through our e-mail correspondence she eventually recognized the subject and legitimacy of our survey and organization and did respond. It is noteworthy that this respondent was in the control group so did not receive the screener question in her e-mail invitation. The question arises that if this respondent had been in the experimental group, would she simply have used the screener-link in the invitation?

The respondent discussed above used the personalization of our e-mail invitation to question the survey's credibility directly. Other registered users in our sample are making use of anti-spam software to block e-mails from unrecognized senders (e.g., e-mail addresses which are not in the user's contact book). These software programs block e-mail from entering the user's inbox and send a reply to the sender asking the sender to follow a set of specific directions. By following these directions the sender legitimizes his or her identity and the e-mail is allowed into the user's inbox, otherwise it is deleted. Three NSAF public users in our sample employed such software. In each instance we followed the required steps to legitimize our e-mail address. However, two of the three registered data users used the software program to block our e-mail address permanently so that the reminder e-mail was also blocked.

7. Conclusions

As pointed out by Porter and Whitcomb (2003), very little of the web survey methodology has focused on how the process of contacting respondents affects the probability of response. In contrast, research on contacting respondents for telephone, personal and mail surveys has been well researched. Much of this discrepancy will change as the procedures for conducting web surveys become established. Until more research is completed, a reliance on what has been learned from other modes of data collection will influence expectations for web surveys. However, techniques that have been developed and tested over time to increase response

rates in mail surveys (Dillman 2000) do not always work for web surveys. We thought that by providing a screening option in the e-mail invitation response rates would improve and the number people incorrectly screening themselves would increase. These conclusions were based on past survey research experience in using an obvious filter question. The fact that we did not see a higher response rate or an increase in the number of screen-outs in the experimental group does not necessarily mean that a similar survey design could not produce this result, but it certainly suggests that more research in needed in understanding the effects that wording and method of contact has on the web survey respondent.

REFERENCES

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. Web Survey Design and Administration. Public Opinion Quarterly 65, no. 2, Summer: 230-53.

Dillman, Don. 2000. Mail and Internet Surveys: The Tailored Design Method. New York: Wiley.

Hatry, Harry P., John E. Marcotte, Thérèse Van Houten, and Carol H. Weiss. 1998. Customer Surveys for Agency Managers: What Managers Need To Know. Washington, DC.: Urban Institute.

Keesling, Randall. 2002. Use of Web-Based Data Collection to Supplement and Improve Mail Survey Response Rates. In Proceedings of the Section on Survey Research Methods. American Statistical Association, 1794-98.

Newcomer, Kathryn E., and Timothy Triplett. 2004. Using Surveys. In Handbook of Practical Program Evaluation (Second Edition), edited by Joseph S. Wholly, Harry P. Hatry, and Kathryn E. Newcomer, Chapter 9. San Francisco: Jossey-Bass.

Porter, Stephen R., and Michael E. Whitcomb. 2003. The Impact of Contact Type on Web Survey Response Rates. Public Opinion Quarterly 67: 579-88.

Salvucci, Sameena, Albert Parker, and Stephan Wenck. 2002. Using Internet-Based Survey to Measure Customer Satisfaction. In Proceedings of the Section on Survey Research Methods. American Statistical Association, 3055-60.