

Social Science Computer Review

<http://ssc.sagepub.com/>

Web Surveys : Perceptions of Burden

Scott D. Crawford, Mick P. Couper and Mark J. Lamias

Social Science Computer Review 2001 19: 146

DOI: 10.1177/089443930101900202

The online version of this article can be found at:

<http://ssc.sagepub.com/content/19/2/146>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Social Science Computer Review* can be found at:

Email Alerts: <http://ssc.sagepub.com/cgi/alerts>

Subscriptions: <http://ssc.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ssc.sagepub.com/content/19/2/146.refs.html>

>> [Version of Record](#) - May 1, 2001

[What is This?](#)

Web Surveys

Perceptions of Burden

SCOTT D. CRAWFORD

MSInteractive

MICK P. COUPER

MARK J. LAMIAS

University of Michigan

Web surveys appear to be attaining lower response rates than equivalent mail surveys. One reason may be that there is currently little information on effective strategies for increasing response to Internet-based surveys. Web users are becoming more impatient with high-burden Web interactions. The authors examined the decision to respond to a Web survey by embedding a series of experiments in a survey of students at the University of Michigan. A sample of over 4,500 students was sent an e-mail invitation to participate in a Web survey on affirmative action policies. Methodological experiments included using a progress indicator, automating password entry, varying the timing of reminder notices to nonrespondents, and using a prenotification report of the anticipated survey length. Each of these experiments was designed to vary the burden (perceived or real) of the survey request. Results of these experiments are presented.

Keywords: Web surveys, survey methodology, Internet, affirmative action, online surveys

Web surveys have burst onto the survey research scene in recent years and have proliferated at a rapid pace. Despite the great promise of this method, there remain a number of concerns. Chief among these is the issue of coverage, given that not all persons in the United States can be reached using the Internet and given that no frame of all Web users exists. But among certain groups with very high penetration of Web and Internet usage, coverage is less of a concern. One such group is college students, and Web surveys of this population appear to be particularly popular. Although coverage is of less concern for surveys of student populations, nonresponse remains an issue.

Several studies comparing Web surveys with mail methods find lower response rates for the former. For example, Kwak and Radler (2000) obtained a response rate of 42% for mail and 27% for Web on comparable samples of college students. In a similar study, Guterbock, Meekins, Weaver, and Fries (2000) obtained response rates of 48% for mail and 37% for Web. In a survey of computer software companies in Australia, Medlin, Roy, and Ham Chai (1999) obtained a response rate of 47% for mail and 28% for Web. Similar differences have been found for e-mail surveys (for a review, see Couper, Blair, & Triplett, 1999; Schaefer &

AUTHORS' NOTE: The authors would like to thank Richard Bagozzi, Dwight F. Benton Professor of Behavior Science in Management at the University of Michigan School of Business Administration, and Richard Allen, Associate Professor at the University of Michigan Department of Communication Studies, for an excellent collaboration and assistance with funding this data collection. An earlier version of this article was presented at the Annual Meeting of the American Association for Public Opinion Research Conference, May 2000, Portland, OR.

Social Science Computer Review, Vol. 19 No. 2, Summer 2001 146-162
© 2001 Sage Publications, Inc.

Dillman, 1998). Given the differences between Web surveys and traditional mail surveys, the knowledge that has been gained over the years on maximizing mail survey response rates may not apply directly to Web surveys. Not only is the medium very different but also, those who use it may have different expectations and may respond to different inducements to participate in the survey. We thus need to find ways to increase rates of response to Web surveys. This article focuses on several approaches deemed likely to affect the likelihood of response to a Web survey.

There are two related nonresponse concerns in Web surveys. The first is overall nonresponse, or complete failure to react to the invitation to participate. Other forms of nonresponse prevalent in Web surveys are break-offs or abandonments (see Jeavons, 1998). In mail surveys, one cannot distinguish the two forms of nonresponse from each other, yet they may have different causes and implications for survey design. Web surveys may be designed interactively, whereby data are transmitted to the server as each question or page is completed, or passively, whereby data are only transmitted to the server after all data have been entered and the *submit* button is pressed. One advantage of interactive Web surveys over passive Web surveys or traditional mail surveys is that the former method yields data from all persons who began the survey, even if they failed to finish it.

In this study we chose to examine several potential response-enhancing features in an interactive Web survey. Part of the reason for this focus is that the content of the survey was out of our control. The survey was conducted for Richard Bagozzi and Richard Allen of the School of Business Administration and Department of Communication Studies, respectively, using an instrument that had been used previously in mail and telephone surveys. We judged this to be a burdensome survey, and hence, ideal for examining nonresponse issues. The principal investigators were amenable to experimentation on nonresponse but did not want the survey content altered.

DESIGN OF THE EXPERIMENTS

We included four experimental treatments in a $2 \times 2 \times 2 \times 2$ full factorial between-subjects design. Three focused on reducing overall nonresponse, whereas the fourth focused on the reduction of abandonment. The four experiments were as follows:

1. Length of survey mentioned in invitation e-mail (TIME)

In one version, sample persons were informed that the survey would only take about 8 to 10 minutes to complete, whereas in the other they were told it would only take about 20 minutes to complete. Our assessment was that the true length of the survey lay somewhere between the two estimates (a great deal depended on how many answers respondents would give in the early part of the survey). We hypothesized that sample persons in the 8 to 10 minute treatment would be more likely to begin the survey than those in the 20 minute condition. On the other hand, if the duration of the survey significantly exceeded the lower estimate, abandonment may be higher for this group. In other words, misrepresenting the length of the survey may get more people to start it but lead to higher break-offs once the length of the survey is clear to respondents.

2. Timing of reminders (REMIND)

In Dillman's (1978) total design method for mail surveys, reminders are timed to arrive approximately 1 week after the initial mailing. Several studies have found that Web survey

Welcome to the Affirmative Action Opinion Survey!!

This interview is voluntary. If you should come to any question you don't want to answer, just click "Next Question" and you will go on to the next question. The answers that you give will be kept confidential as allowed under federal and state law. There has never been a breach of confidentiality in the 50 years we have been conducting surveys.

Please enter your username and password below, then click on "START SURVEY" to begin.

Username:

Password:

Figure 1: Password Entry Screen

returns begin almost immediately but fall off at a rapid rate (e.g., Weible & Wallace, 1998), suggesting that if people are going to complete a Web survey, they do so within the first few hours or days. Given the fast-paced world of the Internet, we argued that a shorter lag between initial invitation and reminders, so that the request is still fresh in the sample person's mind when the reminder is received, would increase the likelihood of responding. We thus varied the timing of follow-up reminders. One group received a first reminder 2 days after the initial invitation was e-mailed and a second reminder 2 days after that. The second group received a single reminder 5 days after receiving the initial invitation. Both treatment groups received two additional reminders at the same time in the later stages of the survey.

3. Form of password access (PASSWORD)

It is important that access to Web surveys be controlled to prevent multiple completions by the same person or other forms of "spamming" or "ballot stuffing." On the other hand, it must be easy for respondents to suspend and resume completion of a survey at a later stage (especially for long surveys or those that may require the respondent to consult records or other persons). At the same time, the survey must reassure respondents that their answers are kept confidential. The typical method of accomplishing these goals in a Web survey is to supply a unique password to each sample person in the solicitation e-mail. This can be done in several ways. We tested two approaches: (a) providing a password that the respondent must enter in the opening screen of the Web survey to gain access, and (b) embedding a unique identifier (ID) into the Uniform Resource Locator (URL) provided to each respondent. Figure 1 shows how the introduction screen appeared with the manual password entry treatment. The automated entry screen appeared without the two form entry fields and instructions to enter the respondent username and password.

The latter approach is likely to minimize effort on the respondent's part and reduce the likelihood of access failure through incorrect password entry (see Couper, Traugott, & Lamias, 2000). On the other hand, providing an explicit password may provide the respondent with an added sense of assurance regarding the confidentiality of his or her responses. We thus expected that the automatic password (ID embedded in the URL) would result in lower nonresponse but that data quality might be higher in the manual password version.

4. Progress indicator (PROGRESS)

Our final hypothesis was focused more on those who began the survey, in the hope of motivating them to complete it. In an earlier study on a similar topic and similar sample of

Question #3	7% of Survey Completed
Please read the following alternatives with respect to affirmative action in the University of Michigan admission process and select the one that best fits your own personal view. Select one by clicking on the circle next to your choice. ☐ Admission policies <u>should not</u> take into account race or ethnicity ☐ Admission policies <u>should</u> take into account race or ethnicity.	
Next Question	

Figure 2: Example of Progress Indicator

students (Couper et al., 2000), a graphical progress indicator (in the form of a pie chart showing percentage complete) was found to increase completions by 3.4 percentage points. Although this difference did not reach traditional levels of significance, the authors argued that the additional download time associated with the progress indicator graphic might have dampened the potentially positive effect of the progress indicator.

Given their finding, we tested a text-based progress indicator in this study. To eliminate any possible effect of differential download times, the no-progress-indicator version downloaded exactly the same information (percentage of the instrument completed) but in the same color as the screen background, rendering the information invisible to respondents in this treatment. In the other condition, the progress indicator (“x% of survey completed”) appeared in the top right of the header bar (containing the question number in the upper left corner). (See Figure 2.) Progress was calculated and displayed at each item, based on the maximum number of questions remaining in the survey given the answers already provided.

Unlike the other three experimental manipulations, we expected the progress indicator to have no effect on the proportion of sample persons beginning the survey (accepting the invitation). However, consistent with Couper et al. (2000), we hypothesized that the presence of the progress indicator would reduce abandonment.

DESIGN AND IMPLEMENTATION OF THE SURVEY

The survey was designed to measure student attitudes toward affirmative action and related issues. The instrument contained a total of 46 questions (screens), some with several parts. A key feature of the survey (and one of the reasons for the concerns about burden) was an open-ended question early in the instrument (Question [Q]6) that asked respondents to provide five reasons for the policy they had chosen in an earlier question (either support for or opposition to affirmative action in college admissions at the university). For each of the reasons provided in Q6, respondents were then asked (in Q7) to provide a rationale why the reason was important to them personally, and again (in Q8) asked why the rationale was important. In other words, if five reasons were provided in Q6 (as requested), the three questions together required a total of 15 open-ended answers. Later questions again displayed these reasons and rationales to respondents in a closed format, asking them to rate the likelihood that the reason given “leads to or causes” the rationale they provided.

The sample was selected randomly by the University of Michigan registrar’s office from their current undergraduate student database. Due to the topic of the survey, the sample was stratified by race (White, Black, Asian, and Hispanic). The resulting combined sample is shown by race in Table 1. Two separate sample replicates were drawn, for reasons to be discussed in detail later.

TABLE 1
Sample Race Stratification

	<i>Replicate 1</i>		<i>Replicate 2</i>		<i>Overall</i>	
	n	%	n	%	n	%
White	999	25.5	665	45.5	1,664	30.9
Black	990	25.2	317	21.7	1,307	24.3
Asian	993	25.3	454	31.1	1,447	26.9
Hispanic	939	24.0	8	0.5	947	17.6
Unknown	0	0.0	17	1.2	17	0.3
Total	3,921	100.0	1,461	100.0	5,382	100.0

TABLE 2
Production Schedule by Reminder Treatment

	<i>2-Day Treatment</i>	<i>5-Day Treatment</i>
Original survey invitation	11/29/1999	11/29/1999
Server problem message	11/30/1999	11/30/1999
Restart/Replicate 2 original message	12/12/1999	12/12/1999
Reminder #1	12/14/1999	12/17/1999
Reminder #2	12/16/1999	NA
Final reminder	01/04/2000	01/04/2000
Survey closing notice	01/11/2000	01/11/2000

NOTE: NA = not applicable.

The survey instrument itself was created using Cold Fusion to link a series of Hypertext Markup Language (HTML) Web pages to a Microsoft Access database. This system allowed for response category order randomization, skip patterns, open- and closed-ended questions, and data fills required by the survey. Such features illustrate the benefits of an interactive Web survey design over a single, static HTML form. In addition, a variety of system information such as time stamps, Internet protocol (IP) addresses and browser types was recorded.

The data collection began in November 1999 and continued through January 2000. Table 2 details the progress of the survey as well as the reminder e-mails as specified in the experimental design. The second reminder for the 5-day treatment was not sent, as this would have fallen on December 24th.

The server used to maintain the survey was a Web server that has been in use at the Institute for Social Research and had served as an adequate platform for many static Web sites. Initial testing with this 486 Windows NT-based machine indicated that it could perform the tasks required of the Cold Fusion-based Web survey. Almost immediately after implementation of the survey, when large numbers of respondents signed on to complete the survey at the same time it was found to be inadequate and caused extremely slow processing of the survey as respondents navigated from question to question. These problems were amplified when the server began to crash repeatedly only hours after the survey invitations had been distributed. The problems caused the need to discontinue the survey while the system hardware was upgraded to a new machine. As noted in Table 2, an e-mail message was delivered to all respondents apologizing for the system problem and advising them to discontinue attempts to complete the survey until a system upgrade could be completed.

Dear xxx

Would you like to have a chance to provide your input on a very important issue here at the University of Michigan? This is your chance! We invite you to participate in a survey about affirmative action.

This survey will only take about 20 minutes to complete.

In appreciation of your participation, upon completion of the survey we will enter you in a drawing for one of three \$40 gift certificates and one \$100 gift certificate to Borders Bookstore.

To participate, please click on this link (or cut and paste it into a Web browser):
<http://projects.isr.umich.edu/aaos/start.cfm>

Username: 9954
 Password: 61256

Please enter the username and password when prompted at the beginning of the survey. This password ensures that you and only you can enter this survey.

If you have any questions about this survey or any problems with completing the survey, please e-mail Websurvey@isr.umich.edu.

Thank you,
 Dr. Richard P. Bagozzi
 School of Business Administration, University of Michigan

Dr. Richard L. Allen
 Department of Communication Studies, University of Michigan

Figure 3: Example E-Mail Solicitation (explicit user identifier and password)

As mentioned previously, two sample replicates were selected for this project. Because of the server problems and subsequent survey downtime, we decided to add “fresh” sample cases to the survey. The primary purpose for this new sample was to help measure any effect that the server problems may have had on the experimental and substantive data.

The replacement server, a 500 MHz Pentium III system running Microsoft NT, was found to be more than adequate to maintain a quick, interactive survey environment. There were no noted or reported problems during the remaining data collection on the new server.

Reminder e-mails were distributed as per the schedule dictated by the experimental treatments. All e-mail messages to respondents were individualized to the specific respondent using Cold Fusion and Microsoft Access. Reminder messages were only sent to those who had not yet completed the survey. The REMIND variable for the experimental treatment was used to send reminders only to the respondents who had been selected to participate in each reminder schedule. The PASSWORD and TIME variables were used to fill the appropriate text into the invitation and reminder e-mails. Figure 3 shows a sample of the invitation message sent with the manual password entry version. Figure 4 contains an extract of the embedded URL version to illustrate the difference between the two approaches to access control.

RESULTS

We began with an overall sample size of 5,430 (both replicates combined). Of these, 48 cases were positively identified as having an invalid e-mail address (the invitation was returned as undelivered). We removed these cases from the denominator for all subsequent analyses, so the effective sample size is the remaining 5,382 cases.

...

To participate, please click on this link (or cut and paste it into a Web browser):

<http://projects.isr.umich.edu/aaos/start.cfm?er=9954&sw=61256>

The address is encoded with a unique identifier to insure that you and only you can enter this survey.

...

Figure 4: Extract of E-Mail Solicitation With Embedded Identifier and Password

TABLE 3
Response Rate Results

	<i>Replicate 1</i>		<i>Replicate 2</i>		<i>Overall</i>	
	n	%	n	%	n	%
Complete	912	23.3	399	27.3	1,311	24.4
Partial	416	10.6	129	8.8	545	10.1
Nonresponse	2,593	66.1	933	63.9	3,526	65.5
Total	3,921	100.0	1,461	100.0	5,382	100.0

The final disposition of the sample cases from the two replicates separately and combined are presented in Table 3. The response rate differences between the two replicates are statistically significant ($\chi^2 = 11.3, p < .01$). However, the effects of the experimental treatments are similar across replicates, so we combine them for the remaining analyses (unless otherwise indicated). The partial group includes 21 persons who signed in to the survey but completed no items. For the initial analyses we treat these as partials, as they at least responded to the e-mail invitation. For the progress indicator analysis, we treat these as nonrespondents, as they were not exposed to the experimental treatment.

Overall, less than 1 in 4 sample persons completed the survey. Even if we combine the completes and partials, we still only get a response rate of 34.5%. This is lower than the 41.5% obtained for the earlier survey on affirmative action conducted on behalf of the student newspaper at the University of Michigan (Couper et al., 2000). A response rate close to 50% was obtained for another Web survey of Michigan students done at approximately the same time by Market Strategies, Inc. (Reginald Baker, personal communication, 2000). Given that the burdensome nature of the survey was not known by respondents at the time of the invitation, we attribute this lower response rate to the timing of the survey. The survey had to compete with other seasonal distractions such as finals and the holiday break.

As with any e-mail based sample, there are always questions on the coverage of the sampling frame (Are those who have e-mail addresses at the University of Michigan actually checking the system for messages?). To get some idea of how problematic this may have been for our sample, we completed a bulk process of "fingering" each e-mail address provided in the sample. This process provided us with the last date that e-mail was checked for that address. It is known that this method is not perfect, as automatically forwarded messages may falsely identify a respondent as a nonuser of e-mail when, in fact, he or she may be a frequent user. The result would be a number that overestimates the proportion of the sample that never checked their e-mail during the production period. We found that only 10.8% of the addresses in the sample had not checked their e-mail during the survey field period. Because this is a high estimate, we felt comfortable that the great majority of the nonresponders had

indeed received the invitation and had chosen not to respond, rather than being sample persons who had never seen the survey invitation.

Among those who completed the survey, respondents in the first replicate took significantly longer to complete the survey than those in the second replicate (21.6 vs. 17.8 minutes, respectively, $p < .01$). Given the reported server problems experienced in the first replicate, this finding verifies that the respondents were significantly slowed in their progress through the questionnaire.

Several key demographic variables on all sample cases (respondents and nonrespondents) were obtained from the registrar's office, permitting an analysis of possible nonresponse effects. We found no significant gender differences in the rate of nonresponse ($p = .09$). Race differences were also not significant ($p = .1$), which is reassuring given the topic of the survey. The effect of year in school on nonresponse is marginally significant ($p = .023$), with freshmen having the lowest nonresponse rate (62.9%) and sophomores the highest (68.3%), juniors and seniors occupying a middle position (65.7% and 65.5%, respectively). We have no explanation for these differences. In the remaining sections we focus on the effects of each of the experimental manipulations in turn.

Disclosure of Length of Survey

As hypothesized, those who were informed that the survey would take 8 to 10 minutes to complete had a lower nonresponse rate than those who were told it would take 20 minutes (63.4% vs. 67.5%, $\chi^2 = 11.6$, $p < .01$). However, the 20-minute group had a lower rate of break off once they started the survey (9% vs. 11.3% of the total sample), resulting in completion rates that are similar across treatment (23.5% for the 20-minute condition, 25.2% for the 8-10 minutes condition, $\chi^2 = 2.1$, $p = .15$). Indicating a lower burden increases the likelihood of sample persons starting the survey, but this strategy can potentially backfire when the reported time underestimates the actual time taken to complete the survey.

The average survey length for all completes was 20.43 minutes (median 19.51 minutes) after imputing for outliers.¹ For those who dropped out of the survey without completing the entire questionnaire, the average duration of the survey was 9.12 minutes (median 5.46 minutes). However, there was no association between the burden indicator (8 to 10 minutes vs. 20 minutes) and time before drop out. Those who were told the survey would take 8 to 10 minutes completed an average of 8.43 minutes before dropping out compared to an average of 9.67 minutes for those who were told the survey would take about 20 minutes ($t = 1.54$, ns). So the burden indicator does not appear to affect the time at which abandonment occurs. The overall times for completed cases are also not statistically significant between high- and low-burden groups.

Timing of Reminders

As expected, the 2-day reminder had a positive effect, but the effect is relatively modest (see Table 4). The effect of the reminder on survey completion appears to be stronger for the second replicate (30.3% for 2-day vs. 24.3% for 5-day reminder) than for first replicate (23.8% vs. 22.7%). The server crash and related problems experienced in Replicate 1 is a likely explanation for the dampening of the effect of the reminder experiment.

Reminders were found to immediately affect the speed of return for completed responses. Figure 5 shows the effect of each reminder mailing for the two treatment groups (2-day reminder and 5-day reminder) for the relatively uncontaminated second replicate. The 2-day treatment shows a constant increase in responses during the first 4 days of the data collection

TABLE 4
Effect of Reminder Interval on Completion

	2-Day Reminder (n = 2,696)	5-Day Reminder (n = 2,686)
Complete	25.6	23.1
Partial	10.6	9.6
Nonresponse	63.8	67.3
$\chi^2 = 7.35$, $df = 2$, $p = .025$		

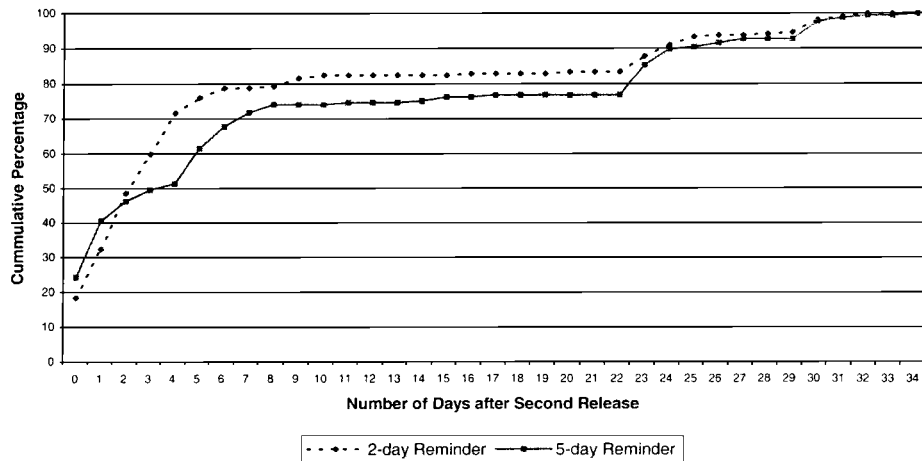


Figure 5: Cumulative Percentage of Surveys Completed by Day for Release 2

before leveling off. The 5-day treatment begins to slow after the 1st day but then shows an increase on the 5th day as a result of the reminder. As this figure suggests, the 2-day reminder not only had a marginal impact on the overall response rate but also affected the speed of returns.

Figure 6 presents the same data in a different format that shows more clearly the effect of the reminders on returns. The two spikes toward the end of the survey were a result of the final reminder (at Day 23) and the closing notice (at Day 31). These large spikes suggest that the effect of the holidays on nonresponse was indeed large (the final reminder was sent on January 4th and the closing notice on January 11th). Clearly, these results suggest that choices about the timing of surveys for student populations should be carefully considered.

Password Access

Our hypotheses relating to the method of password access were that (a) the automatic ID and password would increase the likelihood of signing on to the survey, but (b) requiring the sample person to key in an explicit ID and password would increase the honesty of reporting. As expected, the automatic password version resulted in significantly lower nonresponse than the manual entry version (63.2% vs. 67.8%, $\chi^2 = 12.9$, $p < .01$). This effect was found to

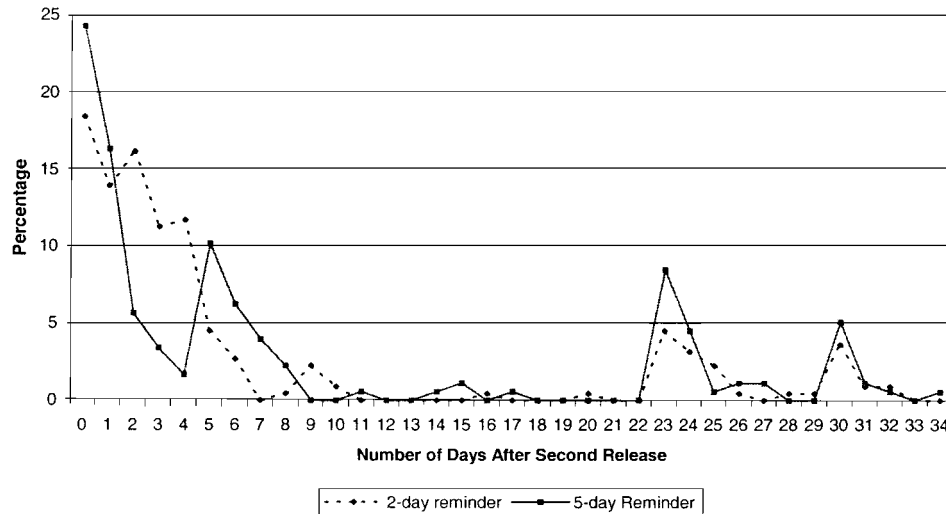


Figure 6: Percentage of Surveys Completed by Day for Release 2

be stronger in the first sample replicate (63.4% vs. 68.9%, a 5.4% difference) than in the second replicate (62.7% vs. 65.0%, a 2.3% difference). This difference could be due to a slightly more server-intensive process involved in the manual password entry version. This extra step could have created a longer wait for those in the first replicate and thus, additional burden to the respondent, leading to a higher level of nonresponse (fewer respondents continuing past the initial welcome screen).

We also hypothesized that data quality would be negatively affected by the automated password entry. Highly motivated respondents would be expected to respond to more of the initial open-ended questions, with greater quality and with more thought put into their answers. However, there was no significant difference in the number of open-ended questions answered at the beginning of the survey whether or not the automated or manual password entry was used ($p = .12$). Controlling for the replicate and the password treatment in a regression model, we found that the automatic password group completed the survey in about 1.5 minutes less time than the manual password group. Given this time difference, it is possible that the automatic password group was proceeding quickly through the survey without giving much thought to the responses, whereas the manual password group was taking more time to consider each question before answering. Perhaps the burden of entering a password filters out less-motivated individuals (as reflected in the difference in the response rate between the two treatment groups).

Additional experimentation on the higher level of data quality for those with the manual password treatment is warranted given these results.

Combined Effects of Nonresponse Reduction Measures

Given that the first three experimental manipulations were all aimed at reducing nonresponse, we ran a series of logistic regression analyses to examine the marginal effect of each factor on the likelihood of a response. The odds ratios from these analyses are presented

TABLE 5
Odds Ratios of Response (vs. nonresponse) From Logistic Regression

	<i>Complete + Partial vs. Nonresponse</i>	<i>Complete vs. Partial + Nonresponse</i>
Time (1 = 8-10 minutes)	1.19**	1.10
Password (1 = automatic)	1.23**	1.14*
Remind (1 = 2-day interval)	1.16**	1.15*

* $p < .05$. ** $p < .01$.

in Table 5. It is clear that each factor has an independent effect on the likelihood of beginning the survey (accepting the invitation), controlling for the other factors. The overall effect of TIME and PASSWORD appears to be stronger on those starting the survey (first column) but not as strong as those finishing the survey (second column). In other words, these strategies serve to get people to visit the URL of the survey site but other strategies are needed to keep them there once they begin the survey. The positive effect of the shorter reminder interval (2 days) appears to hold both for those starting the survey and for those completing it.

Progress Indicator

Whereas the other three treatments were expected to affect the likelihood of starting the survey (or at least visiting the survey site), the use of a progress indicator only affects those who actually began the survey. We thus dropped all nonrespondents (and visitors—those who logged in but did not complete any questions) for these analyses and compared those who completed the survey with those who started the survey but did not complete it.

Contrary to expectation, the progress indicator appears to have a negative effect. Among those who started the survey, 74.4% completed it when no progress indicator was present compared with 68.5% of those who received the progress indicator version ($\chi^2 = 7.7$, $df = 1$, $p < .01$). The same effect holds in both replicates when examined separately. In fact, the negative effect is even stronger in the second replicate, in which there were no server problems to slow down the completion of the survey.

One alternative hypothesis we explored was whether the effect of the progress indicator interacted with the burden (TIME) information in affecting completion. In other words, if respondents were told that the survey would only take 8 to 10 minutes, the slow progress of the indicator may make it clear that the survey would take longer, leading to higher levels of break off.

If anything, the data appear to suggest the opposite. The negative effect of the progress indicator appears to be even stronger in the 20-minute condition than in the 8 to 10 minute condition, as seen in Table 6. Overall, the high-burden (20-minute) group completed the survey at a slightly higher rate once started compared with the low-burden (8 to 10 minute) group (73.1% vs. 70.0%). We also tested the interaction of TIME and PROGRESS explicitly in a logistic regression model. The progress indicator continued to have a significant negative effect on completion, controlling for time. However, the interaction term (when added to the model) was not statistically significant).

Similarly, we counterhypothesized that if the progress indicator inaccurately reported the progress through the survey, it could have had a negative effect on the completion rates for respondents who felt that the survey was more burdensome than anticipated. That is, when the progress indicator read that 10% of the survey was completed, the actual time elapsed

TABLE 6
Effect of Progress Indicator on Completion, Controlling for Time Treatment

	Percent Completed	
	Progress	No Progress
8-10 minute condition $\chi^2 = 2.3, p = .13, n = 973$	67.8	72.2
20 minute condition $\chi^2 = 6.2, p = .013, n = 862$	69.4	76.9

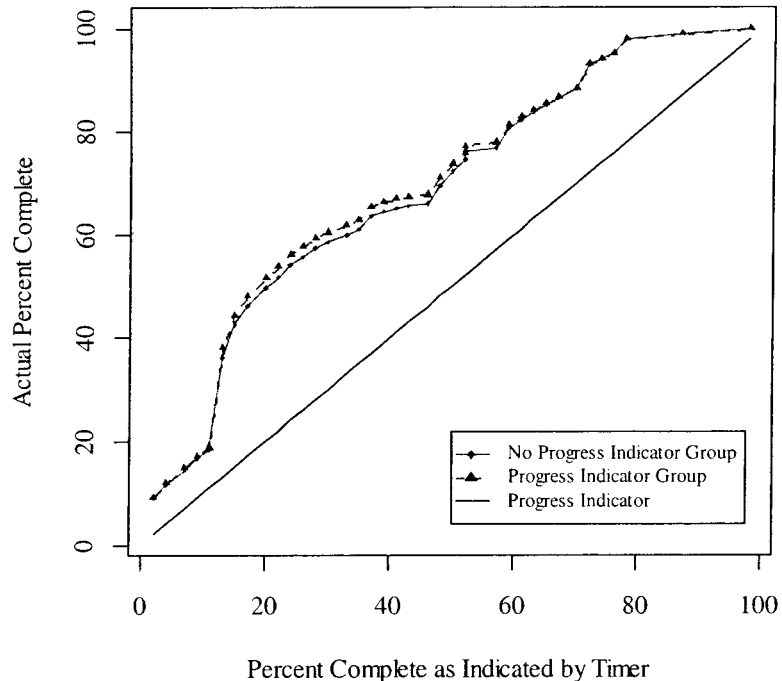


Figure 7: Percentage of Survey Actually Completed by Percentage Completed as Indicated by Timer

was potentially more than that in terms of duration of the survey. This is due to the manner in which the progress indicator was calculated: number of questions completed over total number of questions in the survey. It is clear from Figure 7 that the progress indicator underestimated the amount of time (or effort) that had been completed of the survey on average. For example, when the progress indicator showed the survey to be 20% complete, those who completed the entire survey had completed an average of 51% of the instrument measured in terms of time rather than items completed. Because most of the difference between the actual time remaining and the amount remaining according to the progress indicator is concentrated in the first quarter of the survey (and mostly at the series of open-ended items), it is

TABLE 7
Point of Abandonment by Presence or Absence of Progress Indicator

<i>Point of Abandonment</i>	<i>Progress (%)</i>	<i>No Progress (%)</i>
Q1-Q5 before open series	3.8	2.7
Q6-Q16 during open series	21.0	16.0
Q17-Q45 after open series	6.6	6.9
Q46 complete	68.5	74.4
Total	100	100

NOTE: Q = question.

conceivable that this created a false sense of higher burden to complete the remainder of the survey. In other words, rather than serving a motivating function, it may have served to exacerbate the perceived length of the survey.

A large proportion of the abandonments appears to occur during the series of open-ended question near the beginning of the survey, as evidenced in Table 7. Overall, 31% of all abandonments occurred on Q6 (the first of the open-ended questions), and 56.5% of all abandonments occurred during the set of open-ended items, (Q6-Q16). Those who saw the progress indicator are significantly more likely to abandon the survey during this set of open-ended items. As seen in Table 7, of those who started the survey, 21.0% of the progress-indicator group abandoned at this point compared with 16.0% of the no-progress-indicator group. This accounts for almost all of the difference between the two treatments (5.0% of the 5.9% total difference).

These findings suggest there is something about these open-ended items that produces a negative effect of the progress indicator. We considered the possibility that the progress indicator motivated respondents more and led them to give more complete answers at Q6, with the consequence that they received a larger number of follow-up questions at Q7 and Q8. The data suggest this is not the case. We find no significant differences in the number of answers provided to Q6 (an average of 3.15 for the progress and 3.13 for the no-progress groups) or the length of the answers between the progress-indicator and no-progress-indicator groups (an average of 231 and 232 characters, respectively).

As an aside, it is interesting to note that 44.2% of the sample who started the survey provided five reasons for or against affirmative action, whereas only 20% provided no reasons at this question (this latter number included those who abandoned the survey at this point). Those who completed the survey provided a surprisingly rich amount of information in response to this question. On the other hand, this could be due to self-selection; only the few who were particularly interested or motivated remained at this point, given the low response rate.

Another explanation we considered was that those who expressed greater interest in the topic (elicited in the first few items of the survey, before significant levels of abandonment occurred), would be less likely to abandon the survey, as they would be less susceptible to the effects of a progress indicator. We found no significant differences in rates of abandonment by several interest and salience measures (Q1, Q2, Q3, Q4, Q5). Furthermore, there is no interaction effect of these items and the presence or absence of a progress indicator on completion. The effect of the progress indicator remains negative regardless of the direction or strength of the respondents' reported views on affirmative action.

We also examined the effect of the progress indicator on completion by several demographic variables measured on the full sample. The effect of the progress indicator appears to vary systematically by both race and gender, as shown in Table 8.

TABLE 8
Effect of Progress Indicator on Completion, by Race and Age

	<i>Progress</i>		<i>No Progress</i>		<i>Test of Difference</i>
	%	n	%	n	
Race					
White	66.9	290	77.7	318	$\chi^2 = 8.8, p < .01$
Black	67.2	207	77.0	204	$\chi^2 = 4.91, p < .05$
Asian	70.4	250	69.9	239	$\chi^2 = 0.016, ns$
Hispanic	70.2	171	71.0	155	$\chi^2 = 0.025, ns$
Gender					
Male	67.1	450	71.3	453	$\chi^2 = 1.86, ns$
Female	69.9	468	77.3	463	$\chi^2 = 6.6, p = .01$

Among those who completed the survey, those who received the progress-indicator version expressed marginally more positive attitudes toward the survey than those in the no-progress condition (47.4% vs. 42.8% reporting the experience as *excellent* or *good*, $\chi^2 = 2.78, p = .095$). But this could again be due to the differential break-off rates in the two versions. Incidentally, none of the other experimental treatments appears to affect the subjective experience of the survey among those who completed it. Furthermore, there were no significant differences found in item missing data rates for 40 closed-ended questions with regard to the progress and no-progress treatments ($p = .36$).

Despite these findings, we remain convinced that progress indicators generally serve a useful purpose in Web surveys, and that the negative effect of the progress indicator we found here is an artifact of the particular survey instrument that was used. We are currently examining data from a second wave of data collection from responders to this first survey on affirmative action. The second survey contained many similar questions but did not contain any open-ended items. The preliminary results with an identically created progress indicator are showing a 4.4% increase in completion rate for the progress-indicator treatment.

The progress indicator should be used with caution, as it may sometimes backfire (as we have found here). Clearly, more research is needed to understand the conditions under which a progress indicator has a beneficial effect on Web survey completion.

SUMMARY AND CONCLUSIONS

The overall response rate to this survey suggests that our earlier concerns about high nonresponse rates and high abandonment rates that motivated this research were not unfounded. Even if we discount the server problems experienced in Replicate 1 and the unfortunate subsequent timing of the survey, we could not expect dramatic improvements in the response rate. These results suggest, in line with Bradburn (1978) and others (e.g., Bogen, 1996), that burden is not simply a function of survey length, but potentially involves a large number of other factors only a few of which we studied here.

In mail surveys, sample persons can peruse the entire questionnaire and make informed judgments about participation based on the content, length, design, and other elements of the instrument. In an access-controlled Web survey, such information is not available to sample persons until they decide to visit the survey site and begin the survey. In this respect, Web surveys are similar to interviewer-administered surveys in that the initial decision to participate is based on the relatively limited information conveyed in the advance letter or communi-

cated by the interviewer (see Groves & Couper, 1998). Information about the content of the survey is typically not a key part of the initial decision calculus.

In other words, the e-mail invitation plays a disproportionately important role in eliciting responses to a Web survey. At the same time, many of the standard tools available to the mail survey researcher (e.g., type of postage, type of envelope, personalized signature, use of legitimizing letterhead, inclusion of incentives) do not have direct electronic equivalents. If we assume complete coverage and access to the Web (i.e., every person in our sample did receive the invitation and was able to complete the survey if they wished to), we can see that only 34.5% of the sample (both completes and partials) accepted the invitation and at least signed on to the Web survey site.

As our findings show, factors that manipulate the perceived burden of the task had an effect on the likelihood of accepting the survey invitation. Those who were told the survey would take less time, those who received an automated (embedded) password, and those who received more frequent reminders were all more likely to accept the invitation. Although these factors certainly do help, they do not have large effects on the likelihood of signing on to the survey. Much remains to be learned about features of the e-mail invitation that help or hurt response rates in Web surveys.

Signing on to the Web site is a necessary but not sufficient condition for completing the survey. In a scrollable or static Web design, where all the questions are displayed on a single HTML page, the respondent can make more informed decisions about participation based on the content of the survey. In an interactive Web survey, as ours was, such information must come to respondents gradually as they complete successive items. High levels of abandonment may inform us about the perceived burden of the survey in progress and may even be used to identify particular items that may be responsible, as Jeavons (1998) has shown.

It is clear that the set of open-ended items in this survey played a key role in the relatively high abandonment rates we observed. Our conclusion is that the extra effort, both cognitive (in formulating a response) and psychomotor (in keying the response into the computer), was enough to dissuade relatively large proportions of respondents from continuing beyond this point in the survey. There is no interviewer present to motivate respondents to continue with the survey, and new ways must be found to maintain motivation and interest.

The failure of the progress indicator to have the desired effect on completions is, we believe, closely tied to the effect of these open-ended items. The fact that the progress indicator (based on the number of items completed) underestimated the duration of the survey, largely due to the open-ended items, may have produced the backfire effect we observed. Preliminary data from a similar experiment in the second wave of this panel study, which contains no open-ended items, indicate a positive effect of the progress indicator on survey completion, consistent with the finding of Couper et al. (2000).

Although we believe our finding on the effect of the progress indicator to be an anomaly, these results suggest caution in the use of such tools. Calibration of the progress indicator, for example by allocating a greater weight to open-ended, complex, or other time-consuming items, could result in a more accurate reflection of typical progress through the instrument. More accurate indicators of burden or effort might be needed. Clearly, more research is needed on the utility, format, and design of progress indicators in interactive Web surveys.

Although the content and design of the Web survey instrument may have little effect on initial willingness to participate, they are likely to affect the likelihood of seeing the survey through to completion. Perceived burden, and especially length of the survey, becomes an increasingly important issue for several possible reasons. One reason is based on speculation that the Web permits (and often encourages, e.g., with banner ads) easy navigation from one site to another. Given both the content and multimedia nature of the Web, attention spans

(and consequently tolerance for long surveys) may be lower than on paper. This could be exacerbated in a student population facile with the Web and influenced by the Web culture. Another reason may be that Web surveys are less portable than their paper equivalents. One can set a paper questionnaire aside for completion later or take it with one for completion elsewhere. Although Web surveys do not require single-sitting completion, the burden of resuming a partially completed survey may be greater, as the respondent has to recover the e-mail invitation, URL, and password to reenter the site. Furthermore, the survey must of necessity be completed at a computer connected to the Internet and equipped with a Web browser. These factors restrain the movement or flexibility of Web survey respondents in both time and space relative to their paper survey counterparts.

The challenge for Web survey designers is to find ways to get respondents to sign on to the survey site and to keep respondents motivated to complete the survey once begun. Furthermore, this must be done without unduly increasing the download time of the survey or distracting the respondent from the task at hand. We examined only one feature (progress indicator) aimed explicitly at lowering break-offs. But the multimedia capabilities of the Web offer many more opportunities and challenges in achieving the goals of maximizing survey completion and data quality.

Our findings suggest that Web surveys are different from mail surveys and that procedures which may work for the latter may not translate well to the former. It is as yet unclear, however, whether the underlying mechanisms are the same. Much work remains to be done to understand the factors that affect perceptions of burden in Web surveys and to explore the issue of nonresponse in such surveys.

NOTE

1. Outliers could occur as a result of respondents leaving the survey open to turn to another task. Outliers were deemed to be any observations falling above (or below) the 75th percentile (or 25th percentile) of the time spent on any given screen plus (minus) 1.5 times the interquartile range of the time spent on the screen. Bounds defining outliers were calculated separately for observations within each sample replicate. Observations falling outside these bounds on any particular screen were replaced by the mean value of all the observations for that screen after removing outliers. Time was imputed for 8.1% of all screens (9.4% for Replicate 1 and 4.8% for Replicate 2).

REFERENCES

- Bogen, K. (1996). The effect of questionnaire length on response rates—A review of the literature. *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 1020-1025). Alexandria, VA: American Statistical Association.
- Bradburn, N. M. (1978). Respondent burden. *Proceedings of the American Statistical Association, survey research methods section* (pp. 35-40). Alexandria, VA: American Statistical Association.
- Couper, M. P., Blair, J., and Triplett, T. (1999). A comparison of mail and e-mail for a survey of employees in federal statistical agencies. *Journal of Official Statistics*, 15(1), 39-56.
- Couper, M. P., Traugott, M., & Lamias, M. (2000). *Web survey design and administration*. University of Michigan, Survey Research Center, Ann Arbor. Manuscript submitted for publication.
- Dillman, D. A. (1978). *Mail and telephone surveys; the total design method*. New York: Wiley.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.
- Guterbock, T. M., Meekins, B. J., Weaver, A. C., & Fries, J. C. (2000, May). *Web versus paper: A mode experiment in a survey of university computing*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Portland, OR.
- Jeavons, A. (1998). Ethology and the Web: Observing respondent behaviour in Web surveys. *Proceedings of the Worldwide Internet Conference*, Amsterdam: ESOMAR.
- Kwak, N., & Radler, B. T. (2000, May). *Using the Web for public opinion research: A comparative analysis between data collected via mail and the Web*. Paper presented at the annual meeting of the American Association for Public Opinion Research, Portland, OR.

- Medlin, C., Roy, S., & Ham Chai, T. (1999, November/December). *World Wide Web versus mail surveys: A comparison and report*. Paper presented at ANZMAC99 Conference: Marketing in the Third Millennium, Sydney, Australia.
- Schaefer, D. R., and Dillman, D. A. (1998). Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly*, 62(3), 378-397.
- Weible, R., and Wallace, J. (1998). Cyber research: The impact of the Internet on data collection. *Marketing Research*, 10(3), 19-24, 31.

Scott D. Crawford is the Operations Manager at MSInteractive, L.L.C., a Market Strategies, Inc., company. Address correspondence to him at MSInteractive, L.L.C, 20255 Victor Parkway, Suite 400, Livonia, MI 48152; e-mail: scott.crawford@ms-interactive.com.

Mick P. Couper is Senior Associate Research Scientist at the Survey Research Center. He is also Adjunct Associate Professor of Sociology and Research Associate Professor in the Joint Program in Survey Methodology at the University of Michigan.

Mark J. Lamias is a statistician at the Survey Research Center at the University of Michigan. He holds degrees in statistics and political science.