The Influence of Web-based Questionnaire Presentation Variations on Survey Cooperation and Perceptions of Survey Quality

Jill T. Walston[1], Robert W. Lissitz[2], and Lawrence M. Rudner[3]

This experiment compares cooperation rates across conditions of a web-based survey administered directly on an Internet site. Results indicate that, as in traditional survey modes, expected time burden, overall survey appearance, and official sponsorship can have an influence on survey response rates.

Key words: Internet surveys; nonresponse; questionnaire design.

1. Introduction

The Internet can be an excellent medium for many survey research applications. Web-based survey systems can administer and process large numbers of surveys typically for a substantially lower cost than traditional survey modes. This survey method is becoming increasingly popular, yet the empirical research to guide web survey designers is still young. Understanding what motivates potential respondents to cooperate with a request for online survey participation is one of the areas of research that will inform web-based questionnaire design principles and guide the practice of creating and administering these surveys. Research in this area is relatively new compared to the decades of research on gaining cooperation for more traditional survey modes. This study is intended to contribute to the emerging empirically-based theory about web-based survey respondents' behavior. We look at factors that have been found to influence response rates in traditional survey modes and examine their effect in the web-based mode. We examine use of color and graphics, various item response option formats, government sponsorship identification and suggested time needed to complete the survey as possibly influential characteristics for online survey cooperation. The respondents' perceptions of the survey's quality are also compared across these variable conditions.

[1] American Institutes for Research, Education Statistics Services Institutes, 1990 K Street, NW, Suite 500, Washington, DC 20006, U.S.A. Email: jwalston@air.org
[2] University of Maryland, Department of Measurement, Statistics and Evaluation, 1230 Benjamin Building, University of Maryland, College Park, MD 20742, U.S.A. Email: rlissitz@umd.edu
[3] Graduate Management Admission Council (GMAC), 1600 Tysons Blvd., Ste. 1400, McLean, VA 22102, U.S.A. Email: lrudner@gmac.com

The data for this study consisted of responses to sixteen variations of an on-line survey form. Surveys were presented to over 21,000 people during their visits to a web-site for the Educational Resources Information Center's Clearinghouse on Assessment and Evaluation (ERIC/AE) sponsored by the U.S. Department of Education. (All ERIC clearinghouses were terminated as of December 2003. A new centralized ERIC database web-site became available in September 2004 and is sponsored by the U.S. Department of Education's Institute of Education Sciences.) The survey was administered as part of the ongoing effort to measure ERIC users' level of satisfaction with various aspects of the ERIC web-sites. This type of survey is sometimes referred to as an "intercept" web-based survey because the request to participate occurs during a web-site visit rather than arriving in an e-mail. Web-based surveys in general, and intercept surveys in particular, are especially prone to low response rates due to noncooperation (Couper 2000).

2. Background

There is a large research literature that examines strategies to increase cooperation rates of mailed surveys. For reviews see: Linsky 1975; Heberlein and Baumgartner 1978; Harvey 1987; Goyder 1982; and Dillman 1991. Unfortunately, two successful methods for increasing cooperation rates – prenotification letters (Heberlein and Baumgartner 1978), and follow-up requests for mailed surveys (Dillman 1991) and e-mailed surveys (Schaefer and Dillman 1998) – do not transfer easily to a survey administered directly and immediately to web-site visitors. Monetary incentives, which can be very effective in mailed surveys (James and Bolstein 1990) have also been used in e-mailed surveys via the web-based service Paypal. Using this method, Bosnjak and Tuten (2003) found that potential survey respondents were no more likely to participate with pre-paid incentives or with the promise of a payment than those with no incentive, although those offered a chance for a cash prize upon completion were most likely to participate. Three factors – 1) appearance, 2) sponsorship and 3) time burden – are associated with effects on cooperation rates for traditional surveys and are considered in this study as potential influences for online survey participation.

Childers and Skinner (1996) suggest that color, attractive design and other factors associated with the appearance of a questionnaire affect respondents' perception of the survey's professionalism. This perception, they argue, results in a greater feeling of trust and higher levels of cooperation. Dillman (1978) explains that a professional-looking paper-and-pencil survey conveys seriousness and enhances the perception that it is important for the respondent to comply. Fowler (1993, p. 45) sums up the research regarding the appearance of a mailed questionnaire this way: "Generally speaking, almost anything that makes a mail questionnaire look more professional, more personalized, or more attractive will have some positive effect on response rates." One of the decisions facing a developer of a web-based survey is how best to enhance the visual appeal of the survey using the wide array of possibilities this medium allows. Dillman (2000) cautions that some efforts to enhance the appeal of a web-based survey may backfire. Advanced web features such as video clips, animation and sound may increase the time needed to load these surveys or keep some respondents from being able to access the survey at all and may increase the impression of the complexity of the survey. When these advanced

features increase the time the survey takes to load onto the respondents' computer, the result can be lower response rates (Dillman, Tortora, Conradt, and Bowker 1998). Dillman (2000, p. 384) provides many useful design strategies for providing clear instructions and design features that visually guide respondents through a web-based survey.

Web-based technology allows the survey designer to choose from a variety of item formats. Drop-down boxes, fill in the blank spaces, radio buttons, single buttons, and slider bars are some of the item formats used in surveys administered on the web. The item format and arrangement of the items has an effect on the appearance of the survey and, of great importance, can affect the values of the responses obtained (Couper, Traugott, and Lamias 2001; Tourangeau, Conrad, and Couper 2004). For example, Tourangeau, Crawford, Conrad, and Couper (2004) found that a primacy effect – a tendency for respondents to favor options presented near the beginning of a list – is apparent in web-based surveys as it is in paper-and-pencil surveys, but in addition found that the initial visibility of items in the drop-down box format was a more important factor in respondents' selections. The item format variations possible in web-based surveys exceed those available in paper-and-pencil surveys and the bias associated with web-based formats is an important area for inquiry. The results presented in this article, however, focus on differences in cooperation rates and perceptions of survey quality associated with varying item formats.

Surveys with government or university sponsorship generally have higher response rates than commercial surveys (Linsky 1975; Heberlein and Baumgartner 1978; Goyder 1982). Groves, Cialdini, and Couper (1992, p. 483) explain that higher compliance for these types of surveys is due to the respondents' perception that the request for participation is coming from "someone who is sanctioned by the society to make such requests and to expect compliance." Additionally, this effect is attributed to the respondents' perception that such studies are important, worth the respondent's time, and that the data will not be misused (Childers and Skinner 1996; Dillman 1991). No studies were found that examine this effect in web-based surveys.

Lengthy surveys have lower cooperation rates than shorter surveys because the time commitment and cognitive effort involved in completing a longer survey represents more burden for the respondent (Dillman 1978; Dillman 2000; Childers and Skinner 1996). In one national face-to-face survey, 27 percent of contacted individuals asked the interviewer how long the survey would take (Groves and Couper 1998), indicating that length is an important survey feature. Results from a number of studies suggest that longer surveys tend to achieve lower response rates (Dillman, Sinclair, and Clark 1993; Heberlein and Baumgartner 1978; Goyder 1982). In a review of literature on survey length, Bogen (1996) concludes that there is fairly clear, although inconsistent, evidence that the length of mailed surveys is inversely related to cooperation rates. The length of an Internet-based survey will only be apparent if the respondent is able to, and chooses to, scroll through the entire survey, if an initial time estimate is given, or if a progress indicator is used. In a web-based survey of college students that used e-mailed requests and multiple follow-up reminders for participation, time-to-complete estimates and a progress indicator were experimentally tested for their influence on response rates (Crawford, Couper, and Lamias 2001). These authors found that, of those that logged on to the survey, completion rates were significantly higher for surveys without the progress bar than for surveys with this

indicator (74.4 vs 68.5 percent). They also compared two time estimates given in the e-mail introduction to the survey: 8 to 10 minutes vs 20 minutes. The actual survey length was identical for both conditions (lasting about 20 minutes for complete responders). As expected, the lower time estimate had a significantly higher rate of respondents that logged on to the survey (36.6 percent) than the 20-minute one (32.5 percent). But the final completion rates were not significantly different for the 8–10 minute condition (23.5 percent) as compared to the 20-minute condition (25.2 percent). Crawford and his colleagues speculate that, while the lower time estimate encouraged initial cooperation, those that expected the survey to be shorter than it actually was tended to drop out at a higher rate than those whose survey lasted about as long as expected.

3.   Methodology

3.1.   Instrument

The web-based survey used for this study was a web-site-user satisfaction survey administered to visitors to the Education Resources Information Clearinghouse (ERIC) web-site. Sixteen versions of the survey – identical in item content but different with respect to the variables under investigation – were randomly assigned to the site visitors. As the ERIC visitor entered the site, the survey appeared in a separate window. The visitors could respond or exit at that time or could switch to the main ERIC window, proceed with their task and then respond or exit the survey later. Respondents could exit the survey at any time by either clicking on the "x" at the top right of the window, clicking a "decline to participate" button at the top of the survey page or clicking the "submit" button at the bottom of the survey page.

There were no skip patterns; each respondent was presented with the same number of items. The navigation approach was also identical for all respondents. Respondents moved from item to item by moving the scroll bar along the side of the screen or using the down arrow – much like navigating in a word-processing application. In the case of all respondents that exited the survey an electronic cookie was sent to the visitor's computer which prevented the survey from launching again.

Thirteen ERIC items made up the main content of the survey, followed by three background items, gender, occupation, and frequency of Internet use. The last set of items on the questionnaire required the respondent to report their perceptions, not of ERIC, but of the survey itself. The wording and format of these items was taken from a set of items used in a U.S. Census Bureau study that compared respondent satisfaction in respect of different versions of a computer-administered instrument (Zuckerman et al. 1999).

3.2.   Independent Variables

3.2.1.   Overall Appearance, Plain/graphic

There are two levels of the overall appearance variable, plain and graphic. The "plain" surveys were programmed in HTML and appeared to the visitor with black text on a white background and in a single font type and size (except for the larger font size set by HTML

used for the heading, "ERIC User Satisfaction Survey"). All plain surveys used the common radio button item type format. These plain surveys represent a relatively easily programmed web-based survey and can be prepared with the least amount of knowledge of HTML programming and little concern for graphic design. The "graphic" surveys were created to represent "professionally-designed" web-based surveys. These graphic surveys made use of various colors, font types and sizes, and images. Some of the decisions regarding colors, fonts, image placement and other design elements were made in consultation with a professional web-site graphic designer. These surveys represent the type of product that would require more time and resources to create than the plain version of the survey. Exhibit 1 displays the beginning of a "plain" survey. Exhibit 2 displays the first screen view of a version of the "graphic" survey.

3.2.2. Item Format, Radio Buttons/Big Buttons/Slider Bar

All of the surveys with a "plain" overall appearance had radio buttons for each of the questionnaire items. Radio buttons appear as a row of small circles, each circle corresponding to a response option. The respondent answers each question by clicking in one of the circles, which places a black dot in the circle to indicate that it has been selected. Clicking in another circle within an item's row changes the answer as the dot

ERIC User Satisfaction Survey

Your input is important to us. Please help us serve you better by filling out this survey. It should take no more than 5 minutes.

The U.S. Department of Education, National Library of Education

[I decline to take this survey.]

1. How frequently have you used the following ERIC products and services?
(Click in one circle on each row.)

	Never	Infrequently	Occasionally	Frequently	Very Frequently
ERIC database	○	○	○	○	○
AskERIC	○	○	○	○	○
ERIC Digests	○	○	○	○	○
ERIC Websites	○	○	○	○	○
Direct ERIC Clearinghouse Contact	○	○	○	○	○

2. How helpful have you found each of the following ERIC products and services?
(Click in one circle on each row.)

	Not at all helpful	Somewhat unhelpful	Have not used/ no opinion	Somewhat helpful	Very helpful
ERIC database	○	○	○	○	○
AskERIC	○	○	○	○	○

Exhibit 1.   Initial view of survey: Plain, radio button, government, 5 minutes

Exhibit 2.   Initial view of the survey – graphic, radio button, government, 5 minutes

automatically moves to the new answer. The radio button item type is commonly used for many interactive web applications and is an easily programmed feature in the HTML language.

There are three item format conditions within the graphic overall appearance condition. The item format variation applies only to the first thirteen items on the survey – the main ERIC web-site satisfaction survey questions. The radio button format in the graphic condition functions in the same way as the radio buttons in the plain version. Exhibit 2 shows the initial screen view for one of the graphic radio button surveys. The second item type within the overall graphic condition is individual large buttons. This item type presents each response option as an individually labeled large button that changes to a darker color when clicked. The clickable action area for each response option is much larger for this item type than for the radio buttons. Changing an answer is done in the same way as in the case of radio buttons: by selecting another button in the same row the new answer is highlighted and the old answer automatically goes back to its unselected color. This item type can be seen in Exhibit 3.

The third item type within the graphic condition is the slider bar. Each of these items presents a small vertical slider bar that sits on a horizontal line. The line has tick-marks corresponding to five labeled response options but the respondent can drag the slider bar to any position along the scale. The items for each format type have similar vertical spacing and placement on the screen and the positions of the horizontal response options are spaced similarly for the three item formats. See, Exhibit 4 for a screen view of a survey with slider bars. The item content and response option labels are consistent across the item

Exhibit 3.   Initial view of the survey – graphic, large buttons, government, 5 minutes

formats. The last eight questions on every survey, three demographic questions and five items about perceptions of survey quality, appear in radio button format for all versions of the survey (in the "plain" format for the plain surveys and in the "graphic" radio button format for all the graphic surveys).

3.2.3.   Sponsorship: Government/Nongovernment

For all surveys under the government-sponsored condition, the statement "ERIC is a project of the Department of Education's National Library of Education" appears in the introduction. For those in the nongovernment condition, this statement does not appear. The sponsorship statements have the same screen placement and font size in the plain and the graphic government-sponsored conditions. In the graphic condition the government surveys have the Department of Education seal as the illustration for the introduction (see, Exhibit 2) while the nongovernment sponsored graphic surveys have a similarly placed and sized image of a photograph of hands typing on a computer keyboard (see, Exhibit 3).

3.2.4.   Time to Complete Estimate, Five Minutes/Fifteen Minutes

To manipulate the potential respondents' expectation of how long the survey might take to complete, the introduction included either the statement, "This survey should take no more than 5 minutes to complete," or "This survey should take no more than 15 minutes to complete." Again, the placement of these two statements was identical across all survey conditions.

Exhibit 4.   Initial view of the survey – graphic, slider, government, 5 minutes

3.3.   Dependent Variables

i) *Survey started:* Surveys are classified as having been begun (at least one item answered) or not begun (regardless of how the respondent exited the survey).

ii) *Survey completed:* Surveys are classified as "complete" if they are returned with at least 90 percent of the items answered – no more than two items are missing. All other launched surveys are classified as not completed (including partial completes and nonresponders).

iii) *Perceptions of survey quality*: These are the respondents' perceptions about the survey along five dimensions: attractive-unattractive, worthwhile-waste of time, stimulating-dull, easy-difficult, frustrating-satisfying. These are the last five items to appear on the survey.

3.4.   Design

This study uses an experimental design to investigate the effect on cooperation rates and perceptions of survey quality due to variations in a web-based questionnaire. All independent variables are crossed, except that the item format variable is not fully crossed,

with the overall survey appearance variable. The graphic condition has three item formats while plain surveys have only the radio button format. A total of 21,588 surveys were launched. The surveys were launched so that equal numbers of plain vs graphic, government vs nongovernment, five- vs fifteen-minute time estimate surveys would occur. Fifty percent of the surveys are graphic, and one third of these are in each item format conditions.

3.5. Subjects

The participants in this study were visitors to the Education Resources Information Center, Assessment and Evaluation (ERIC/AE) Clearinghouse web-site. ERIC/AE was one of the 16 subject-specific ERIC clearinghouses and it provided user assistance to locate reports, journal articles and commercial books related to assessment and evaluation. Seventy-one percent of the respondents to this survey were women, 37 percent identified themselves as researcher or professor, 23 percent as a K-12 teacher or administrator and 24 percent as a college student. Over 90 percent of the respondents indicated that they access the Internet at least a few times a week.

4. Results

4.1. Initial Exit

Table 1 presents the percent of launched surveys that were exited when the site visitor closed the window by clicking on the "x" at the top right corner or by clicking on the "I decline to take this survey" button near the top of the survey page.

The most striking difference in the outcomes is that 80.2 percent of those receiving a slider bar survey exited the survey as compared to 61.6 to 64.7 under the other appearance/item format conditions. Conversely, 9.0 percent with a slider bar survey explicitly declined to take the survey by selecting the "I decline to take this survey" button near the top of the survey screen, as compared with 17.9 to 20.8 for the other conditions. The load time for the surveys was not registered in this study but during technical pretesting it was noted that slider surveys typically took around five seconds to load but that on a few computers the load time was around ten to fifteen seconds. The other item

Table 1.   *Percent of launched surveys that were closed or declined*

Exit mode	Appearance/Item format				Total graphic	Sponsor		Time		Total
	Plain	Graphic								
	Radio	Radio	Buttons	Slider		None	Gov't	5	15	
Closed window	64.7	62.5	61.6	80.2	67.9	66.7	66.0	65.6	67.0	66.3
"Decline" exit button	17.9	18.4	20.8	9.0	16.1	16.8	17.2	16.6	17.4	17.0

Note: Total number of launched surveys $= 21,588$.

format types typically loaded within a couple seconds. It appears likely that many potential respondents were lost in the first few seconds while the survey was loading.

4.2. Initial Cooperation

These analyses consider how likely a visitor is to respond to the survey regardless of the number of items that were completed or how they exited the survey. For example, even if a respondent answered just a couple of questions and then hit the "decline" button at the top of the survey, they are included as partial completers because at least one survey item response was captured. Analyses comparing the likelihood that a respondent returned a survey with at least one item completed is done to evaluate the variables' influence on encouraging respondents to at least begin the survey (Table 2).

Logistic regression analyses are conducted to determine if there are significant differences in these rates across survey conditions. Logistic regression is similar to linear regression, but is appropriate for models with a dichotomous dependent variable; in this model, the survey variables are used to predict whether the respondent begins or does not begin the survey. An analysis of only radio surveys (Table 3) provides comparisons involving the influence of the overall appearance variable, plain vs graphic, apart from effects associated with the various graphic item formats.

A main effect for the time estimate ($B = .238$) for these radio surveys suggests that the shorter time-to-complete statement positively affected the decision to begin the surveys (14.4 percent for "5 minutes" vs 10.8 percent for "15 minutes"). There is a significant interaction between appearance and sponsorship ($B = .345$); this is illustrated in Figure 1.

The graphic appearance has a positive influence in that visitors will begin a radio button type survey only for the government-sponsored surveys. For surveys without government sponsorship identified in the survey introduction, completion rates are similar for graphic radio surveys and plain radio surveys.

In order to examine the effect of various item formats, the same analysis was done with only the "graphic" surveys. Table 4 shows the logistic regression results for this analysis.

Results of the analysis of the likelihood that a respondent will begin one of the graphic surveys indicate that the main effect of the time estimate ($B = .319$) is similar to that

Table 2. *Response rates: percent returning at least a partially complete survey*

	Plain	Graphic			Total graphic	Total
	Radio	Radio	Buttons	Slider		
Gov't sponsor	11.6	15.3	11.3	8.7	11.8	11.7
5 min.	12.7	17.8	12.4	9.3	13.2	12.9
15 min.	10.5	12.7	10.1	8.2	10.4	10.4
No sponsor	11.1	10.8	13.1	6.9	10.3	10.7
5 min.	11.9	12.4	15.3	8.1	11.9	11.9
15 min.	10.3	9.3	10.8	5.6	8.6	9.4
Time 5 min.	12.3	15.1	13.8	8.7	12.5	12.4
Time 15 min.	10.4	11.0	10.5	6.9	9.5	9.9
Total	11.4	13.1	12.2	7.8	11.0	11.2

Note: Total number of launched surveys = 21,588.

Table 3. Radio button surveys: Logistic regression analyses, likelihood that respondent returned at least a partially complete survey

	Block 1:		Block 2:		Block 3:	
	Main effects		2-way interactions		3-way interactions	
	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio
Appearance (graphic)	.157 (.06)	* 1.17	− 126 (.11)		− .120 (.14)	
Sponsor (gov't)	.143 (.05)	* 1.15	.020 (.08)		.023 (.09)	
Time (5 min.)	.238 (.05)	* 1.27	.163 (.08)		.165 (.09)	
Appearance * Sponsor			.345 (.12)	* 1.41	.335 (.18)	
Appearance * Time			.175 (.12)		.164 (.18)	
Sponsor * Time			.055 (.11)		.050 (.13)	
Appearance * Sponsor * Time					.019 (.24)	
Constant	− 2.251 (.05)	*	− 2.161 (.06)	*	− 2.163 (.07)	*
Chi-square value for step (df)	*33.8 (3)*	*	*10.7 (3)*	*	*.006 (1)*	

Note: The reference condition for each main effect is noted in parentheses. Values are regression Beta weights; standard errors are in parentheses; ∗ $p < .05$; likelihood ratios (Lk.) are provided for significant effects. Total number of launched surveys = 21, 588.

found for the analysis of radio surveys. The five-minute condition has a higher rate of surveys that were begun (12.5 percent) than the fifteen-minute condition has (9.5 percent). Within the graphic conditions, there were significant interactions between item types and sponsorship with regard to the likelihood that a respondent begins the survey (slider vs non slider by sponsor, $B = −.431$; radio vs button by sponsor, $B = .566$). For the government
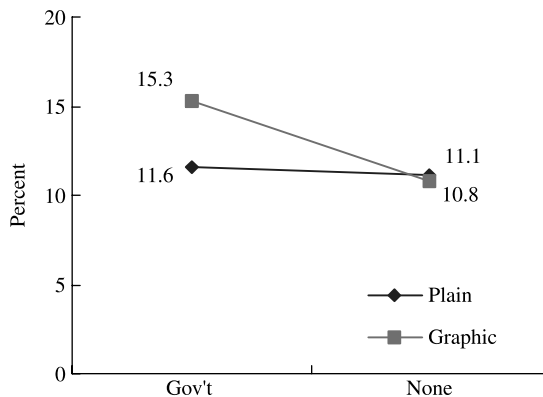

Fig. 1. Radio button surveys: Percent of surveys that were at least partially completed, interaction between appearance and sponsorship

Table 4. *Graphic surveys: Logistic Regression Analyses: Likelihood that respondent returned at least a partially complete survey*

	Block 1:		Block 2:		Block 3:	
	Main effects		2-way interactions		3-way interactions	
	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio
Item (Radio vs Button)	.082 (.07)		− .249 (.14)	* .78	− .174 (.16)	
Item (Nonslider vs Slider)	.492 (.08)	* 1.64	678 (.15)	* 1.97	.710 (.19)	* 2.04
Sponsor (gov't)	.156 (.06)	* 1.17	.292 (.10)	* 1.34	.309 (.11)	* 1.36
Time (5 min.)	.319 (.06)	* 1.38	.354 (.10)	* 1.43	.371 (.11)	* 1.45
Sponsor * Item (r/b)			.566 (.15)	* 1.76	.431 (.22)	
Sponsor * Item (slider)			− .431 (.17)	* .65	.480 (.25)	
Time * Item (r/b)			.063 (.15)		− .067 (.21)	
Time * Item (slider)			− .060 (.17)		− .005 (.24)	
Sponsor * Time			− .100 (.13)		− .134 (.15)	
Sponsor * Time * Item (r/b)					.238 (.30)	
Sponsor * Time * Item (slider)					− .088 (.34)	
Constant	− 2.445 (.06)	*	− 2.512 (.08)	*	− 2.522 (.09)	*
Chi-square value for block (df)	*88.32 (4)*	*	*16.79 (5)*	*	*1.13 (2)*	

Note: The reference condition for each main effect is noted in parentheses. Values are regression Beta weights; standard errors are in parentheses;* $p < .05$; likelihood ratios (Lk.) are provided for significant effects. Total number of launched surveys = 21,588.

sponsorship condition, radio surveys are more likely to be started than the other item types within the graphic condition, (4 percent higher than button surveys and 6.6 percent higher than slider surveys). Within the nongovernment surveys, the button surveys are the most likely started (2.3 percent higher than radio surveys and 6.2 percent higher than sliders). The button surveys are the only item type where the likelihood of starting the survey is higher in the nongovernment condition. Possible reasons for such an interaction between sponsorship and item types were explored in subsequent one-on-one cognitive interviews with a small sample of survey takers ($n = 9$) and no reasonable explanation was apparent. Attaching theoretical significance to this statistically significant 3-way interaction seems unwarranted.

In an analysis of all survey types considered together, it emerges that there is a significantly higher rate of those beginning the survey for the five-minute surveys (12.4 percent) than for the fifteen-minute ones (9.5 percent) across all appearance and item format types ($B = .253$). (Full regression output for this analysis is not shown in tables.)
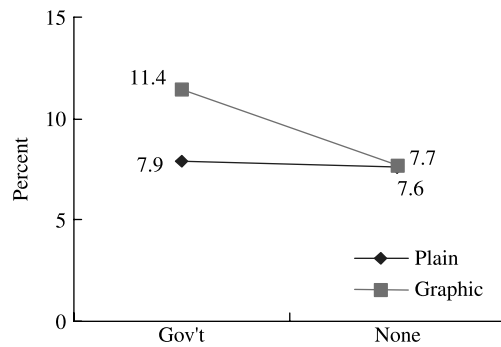
Fig. 2. Radio button surveys: Percent of all surveys that were completed – interaction between appearance and sponsorship

4.3. Response Rates

The next set of analyses investigates influences associated with the likelihood that a visitor will fully cooperate with the request and return a completed survey. A survey is considered complete for these analyses if at least 90 percent of the items are answered and no more than 2 out of the 21 items are missing. Table 5 shows the percent of surveys that were completed across each of the survey presentation conditions.

Overall appearance interacts with both sponsor ($B = .398$) and time ($B = .323$) in respect of the likelihood that a respondent completes the survey. Figure 2 illustrates the interaction effect for appearance and sponsorship. Graphic radio button surveys are associated with a 3.5 percent improvement in completion rates as compared to plain radio button surveys in the government condition. In the nongovernment surveys, overall appearance has no effect. Figure 3 illustrates the interaction effect for appearance and time. Graphic radio surveys are associated with a 3.2 percent improvement in completion rates as compared to plain surveys in the five-minute condition. When the survey introduction states that the survey might take fifteen minutes, the overall appearance does not significantly affect completion rates (see Table 6).

Table 5. Percent returning a completed survey

	Plain	Graphic			Total graphic	Total
	Radio	Radio	Buttons	Slider		
Gov't sponsor	7.9	11.4	8.5	8.1	9.3	8.6
5 min.	8.2	13.7	9.2	8.7	10.5	9.4
15 min.	7.6	9.2	7.8	7.5	8.2	7.9
No sponsor	7.6	7.7	8.9	6.2	7.6	7.6
5 min.	7.7	8.8	10.5	7.3	8.8	8.3
15 min.	7.4	6.5	7.3	5.2	6.3	6.9
Time 5 min.	8.0	11.2	9.8	8.0	9.7	8.8
Time 15 min.	7.5	7.9	7.5	6.3	7.2	7.4
Total	7.7	9.5	8.7	7.2	8.5	8.1

Note: Total number of launched surveys $= 21,588$.
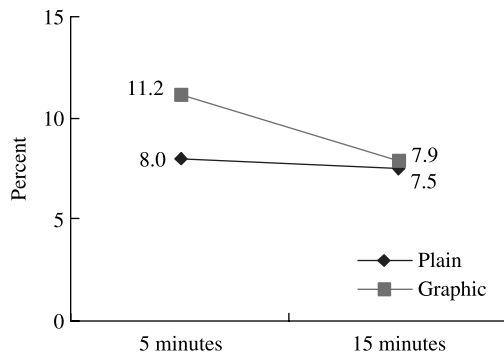
Fig. 3. Percent of all surveys that were completed – interaction between appearance and time, radio button surveys

Table 7 presents results for comparisons of completion rates for all graphic surveys.

Time is a significant main effect across all graphic surveys ($B = .318$); 6.9 percent of those with the fifteen-minute statement in the introduction returned a completed survey, as compared to 9.3 percent with the five-minute statement. There is a significant interaction between sponsorship and item type when the radio and button surveys are compared ($B = .497$). The direction of this interaction is similar to what was found for the item

Table 6. Radio button surveys: Logistic regression analyses, likelihood that respondent returned a complete survey

	Block 1:		Block 2:		Block 3:	
	Main effects		2-way interactions		3-way interaction	
	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio
Appearance (graphic)	.232 (.07)	* 1.26	− 163 (.13)		−.137 (.16)	
Sponsor (gov't)	.157 (.06)	* 1.17	.008 (.10		.020 (.11)	
Time (5 min.)	.161 (.06)	* 1.18	.034 (.10)		.045 (.11)	
Appearance * Sponsor			.398 (.14)	* 1.49	.351 (.21)	
Appearance * Time			.323 (.14)	* 1.38	.276 (.21)	
Sponsor * Time			.068 (.13)		.045 (.15)	
Appearance * Sponsor * Time					.083 (.28)	
Constant	− 2.645 (.06)	*	− 2.519 (.07)	*	− 2.525 (.08)	*
Chi-square value for step (df)	*23.81 (3)*	*	*14.05 (3)*	*	*.086 (1)*	

Note: The reference condition for each main effect is noted in parentheses. Values are regression Beta weights, standard errors are in parentheses,* $p < .05$, likelihood ratios (Lk.) are provided for significant effects. Total number of launched surveys $= 21,588$.

Table 7.   Graphic surveys: Likelihood that respondent returned a complete survey

	Block 1:		Block 2:		Block 3:	
	Main effects		2-way interactions		3-way interactions	
	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio	B(s.e.)	Lk. ratio
Item (nonslider vs slider)	.209 (.09)	* 1.23	.363 (.17)	* 1.44	− .368 (.20)	
Item (radio vs button)	.105 (.08)		− .229 (.16)		− .120 (.19)	
Sponsor (gov't)	.229 (.07)	* 1.26	.332 (.12)	* 1.39	.335 (.12)	* 1.40
Time (5 min.)	.318 (.07)	* 1.37	.355 (.11)	* 1.43	.356 (.12)	* 1.43
Item (slider) * Sponsor			.333 (.18)		.302 (.26)	
Item (r/b) * Sponsor			.497 (.17)	* 1.64	.331 (.27)	
Item (slider) * Time			.115 (.17)		− .074 (.25)	
Item (r/b) * Time			− .035 (.18)		− .028 (.27)	
Sponsor * Time			− .087 (.15)		− .096 (.16)	
Item (slider) * Sponsor * Time					.003 (.36)	
Item (r/b) * Sponsor * Time					.340 (.34)	
Constant	− 2.712 (.07)	*	− 2.766 (.09)		− 2.766(.09)	
Chi-square value for step (df)	*42.91 (4)*	*	*9.81 (5)*		*1.26 (2)*	

Note: The reference condition for each main effect is noted in parentheses. Values are regression Beta weights, standard errors are in parentheses,* $p < .05$, likelihood ratios are provided for significant effects. Total number of launched surveys = 21, 588.

formats and sponsorship interaction with regard to initial cooperation rates (Table 2), although each estimate of course is lower (see Table 5).

The "time by appearance" interaction effect on completion rates, which is significant when only radio surveys were compared, is also significant ($B = .250$) when all item types are considered together. (Full regression output for this analysis is not shown in tables.) The five-minute condition has a higher completion rate for graphic surveys (all three item formats combined) (9.7 percent) than the fifteen-minute condition (7.2 percent), while the plain surveys had more similar completion rates for the two time conditions (five-minute, 8.0 percent; fifteen-minute, 7.5 percent).

4.4.   Perceptions of Survey Quality

The last five items on the survey capture the respondents' ratings of five qualities of the survey. These items were presented with a semantic differential scale item format.

Responses for these items were coded on a scale of one to five with higher values associated with more positive ratings and a rating of three indicating the midpoint between two labeled endpoints. Every survey presented these items with a radio button format (the item format variables listed in the table refer to the format of the main survey items – the preceding 13 ERIC evaluation items). Means across survey variable levels are presented in Table 8. Analyses comparing responses across levels of the survey presentation variables give an indication of whether these features influence perceptions of survey quality. It was expected that respondents under the government sponsorship condition would rate their surveys more positively along the "worthwhile" dimension and those with the graphic surveys would rate the surveys as more "attractive."

Multiple analysis of variance (MANOVA) procedures were conducted to test the influence of the survey features on respondents' perceptions of the quality of the survey. This general linear model tests the effects of each of the survey variables on the joint distribution of the five perception items simultaneously. Post hoc tests for the influence of the survey variables on individual perception item means are examined and described below only when the omnibus F-test indicates a significant overall effect.

The MANOVA conducted on all surveys with radio button items indicates a significant effect on the overall survey appearance (plain vs graphic) ($F = 15.62$, $df = 5$) on survey perceptions. The time and sponsorship variables were not associated with survey perception differences. The subsequent univariate tests show that only the attractive-unattractive item has a significant difference between the mean responses relating to the appearance of the survey (plain vs graphic). Those with the graphic radio surveys rated the survey as more attractive (mean = 3.86) than did those responding to the plain surveys (mean 3.32; $F = 65.32$, $df = 1$). This difference represents a moderate effect size of .51 or about half a standard deviation. A separate MANOVA analysis of graphic surveys shows

Table 8.   Survey satisfaction item means

	Appearance/Item format				Total graphic	Sponsor		Time		Total
	Plain	Graphic								
	Radio	Radio	Buttons	Slider		None	Gov't	5	15	
Attractive – Unattractive	3.32	3.86	3.86	4.00	3.89	3.57	3.65	3.63	3.58	3.61
Worthwhile – A waste of time	3.50	3.47	3.42	3.62	3.49	3.46	3.52	3.49	3.49	3.49
Stimulating – Dull	2.86	2.98	2.91	3.07	2.98	2.88	2.96	2.96	2.87	2.92
Easy – Difficult	4.39	4.44	4.38	4.57	4.45	4.42	4.43	4.43	4.41	4.42
Satisfying – Frustrating	3.48	3.55	3.47	3.73	3.57	3.45	3.58	3.53	3.51	3.52

Note: listwise $N = 1,814$. Standard deviations range from .81 for the easy-difficult scale for the slider format to 1.15 for the worthwhile-waste of time scale for the slider format. The easy-difficult variable has a skewness statistic of about 1.5, which differs from the expected value of 0.0 for a normally distributed variable. This departure from normality is not considered severe enough to jeopardize the MANOVA test. The other four items have skewness statistics less than $+/- 0.5$.

no significant main or interaction effects due to time, sponsorship or the different item formats within the graphic condition.

Results of the MANOVA procedure for all surveys reveal a significant overall appearance effect (F = 34.8, df = 5). The univariate tests show significant effects for the attractive-unattractive item (F = 146.3, df = 1) and the stimulating-dull item (F = 5.0, df = 1). Those with graphic surveys (all three item formats combined) had a higher mean rating for the "attractive – unattractive" scale (3.89) compared to those in the plain survey condition (3.32; F = 146.26). The effect size for this difference is .54. A small but significant effect (d = .12) is noted for the higher rating for graphic surveys along the "stimulating – dull" scale (2.98) as compared to the plain surveys (2.86).

5. Discussion

The overall response rate for this web-based survey is low. Only 11.2 percent of visitors began the survey and 8.1 percent completed it. The lowest response rates were obtained in the slider bar, nongovernment, fifteen-minute time estimate condition (5.6 percent initial cooperation, 5.2 percent completes) and the highest in the graphic/radio, government, five-minute condition (17.8 percent initial cooperation, 13.7 percent completes). While the significant main and interaction effects on cooperation rates noted in the following discussion are due to modest changes in percentage points, the differences represent substantial differences in the likelihood of participation.

5.1. Survey Appearance

The graphic versions were rated more "attractive" by the respondents, as anticipated, and when all graphic item types are considered together, also slightly more "stimulating." While the nonrespondents' perceptions of the survey are unknown, the responses provided by the respondents suggest that the effort to make the graphic surveys more attractive than the plain surveys was successful. It was anticipated that the visual appeal of the graphic surveys could increase the cooperation rate, across all levels of the other survey variables, but this was not the case. There was no difference in the likelihood of starting or completing the survey that could be attributed solely to overall appearance of the survey. However, when the survey also conveyed other information expected to increase response rates, government sponsorship or lower time estimate, the radio item graphic surveys had a higher completion rate than the plain surveys. It appears that a survey's attractiveness alone is not likely to affect response rates for all online surveys, but can have a positive effect for surveys that are reasonably short or affiliated with an official sponsor.

5.2. Item Formats

Those receiving slider surveys were most likely to close the window (80.2 percent) and these surveys had the lowest cooperation rates. System response time is considered to have a great influence on users' satisfaction with any computer application (Schneiderman 1998, p. 354). Given the setting of the request for survey participation in this study, these five to fifteen seconds were probably perceived by many to be an unacceptably long time to wait for what was essentially an interruption of their visit to the web-site. This

conclusion supports Dillman's (2000) caution that advanced web-based features can have a negative effect on cooperation rates. It is unclear how well this finding may generalize to other web-survey applications. In surveys where the invitation to participate is delivered via e-mail, the respondents are in control of when they view the survey. Perhaps they would tend to choose to view the survey during a time in their day's schedule when seconds in wait time may be less of a deterrent. However, minimizing the wait time required to view a survey should probably be a priority for any type of web-based survey.

   The radio button item format within the graphic overall appearance condition had the highest overall response rates. Of the three item types compared in this study, the radio buttons are the only format that could be programmed exclusively in HTML. Given these two considerations, it seems clear that web-based surveys with the radio button item formats along with judicious use of color and images should be considered an optimal design for web-based survey items.

5.3.   *Time Burden*

Across all other levels of the survey variables, the likelihood that a potential respondent began the survey was higher for surveys estimated to take up to five minutes than for surveys that suggested a completion time of up to fifteen minutes. Time did not interact with appearance for the likelihood that the respondent *began* the survey, but the effect of time expectation interacted with appearance for the likelihood that the respondent *completed* the survey. It seems apparent that when the survey is expected to be burdensome (fifteen minutes), its attractiveness is irrelevant to the decision to begin or to complete it. Those unwilling to participate because of the fifteen-minute time commitment implied in the introduction were not enticed to begin the survey due to a more appealing look. In the case of the five-minute surveys, where time burden is probably less of a salient feature, some respondents were more motivated to begin and complete the more attractive surveys than the plain ones. It appears that time burden was a primary consideration for the respondents in this study though it is not certain whether everyone read the relevant statement. Given the high rate for closing the survey (Table 1), there was a relatively low difference (1.4 percent) for this exit behavior between the two time conditions (65.6 percent closed the 5-minute survey and 67.0 percent closed the 15-minute survey). It seems likely that many of those that closed the survey did so before completely reading the introduction.

5.4.   *Sponsorship*

In the comparison of all radio button type surveys, response rates were higher for those receiving surveys with the Department of Education identified in the introduction than for those receiving surveys without this identification, but this effect was only apparent in the graphic surveys. The more pronounced display of government affiliation conveyed in the graphic surveys may account for its positive influence in these graphic surveys but not in the plain surveys. The interaction between the large button item format and the sponsorship condition was unexpected and no obvious explanation is apparent.

6. Further Study

There are a wide variety of design elements, color schemes and item formats that can be incorporated into experimental designs of web-based surveys. Drop and drag type features might be useful for items where the respondent ranks a series of options. Slider bars may be used more successfully perhaps as faster computers become more prevalent. While surveys with the shortest expected time had the highest response rates in this study, it would be interesting to examine the effect of other time estimates or no time estimate. Given that the expected time burden has a relatively strong influence on response rates, perhaps a different approach for surveying web-site visitors should be investigated. Rather than administering an entire set of survey questions during a specific survey collection period, an ongoing series of very short surveys may prove useful. If the respondent could see the entire survey within one small window, perhaps response rates would be substantially higher than what was found in this study.

Given the immediate nature of the invitation regarding survey participation when web-site visitors are the target population, it may be that reaching high response rates will always be a major challenge. This study suggests that characteristics of the survey can have an effect on response rates, but the upper limit of what might be reasonably expected is still unknown.

Research in this area should continue to recognize that high response rates are not sufficient for collecting valid data. Studies like this one that focus exclusively on cooperation rates provide insight into potential respondents' decision to participate, but leave questions relating to response bias unanswered. Important differences between respondents and nonrespondents can introduce bias even when relatively high response rates are achieved. A complete picture of the influence of web survey features on data quality will probably take many years to establish and will come about through a variety of studies on different aspects of survey data quality – cooperation rates being one piece of this picture.

Appendix A: List of Survey Items

1. How frequently have you used the following ERIC products and services?
   [Never, Infrequently, Occasionally, Frequently, Very Frequently]
   ERIC Databases
   AskERIC
   ERIC Digests
   ERIC web-sites
   Direct ERIC Clearinghouse Contact
2. How helpful have you found each of the following ERIC products and services?
   [Not at all helpful, somewhat unhelpful, Have not used/No opinion, Helpful, Very helpful]
   ERIC Databases
   AskERIC
   ERIC Digests
   ERIC web-sites
   Direct ERIC Clearinghouse Contact

3. Based on your overall experience with ERIC web-based systems, including other ERIC web-sites you've visited, indicate your level of agreement with the following statements.
   [Strongly disagree, Disagree, No opinion, Agree, Strongly agree]
   The ERIC online system is easy to use.
   The ERIC online system is well organized.
   The ERIC online system is attractively designed.
4. What is your gender?
   [Male, Female]
5. In what capacity are you visiting the ERIC site today?
   [Researcher/Professor, K-12 Teacher/Administrator, K-12 Student, College Student, Other]
6. Not including e-mail, how often do you access the Internet at work or home?
   [Almost every day, A few times a week, A few times a month, About once a month, Less than once a month]
7. Lastly, we'd like to get your reaction to this electronic survey. Please rate this survey across each of these five dimensions.
   Attractive – Unattractive
   Waste of time – Worthwhile
   Stimulating – Dull
   Easy – Difficult
   Frustrating – Satisfying

7. References

Bogen, K. (1996). The Effects of Questionnaire Length on Response Rates – A Review of the Literature. Paper presented at the annual meeting of the American Association for Public Opinion Research.

Bosnjak, M. and Tuten, T. (2003). Prepaid and Promised Incentives in Web Surveys – An Experiment. Social Science Computer Review, 21, 208–217.

Childers, T.L. and Skinner, S.J. (1996). Toward a Conceptualization of Mail Survey Response Behavior. Psychology and Marketing, 13, 185–209.

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

Couper, M.P., Tragoutt, M.W. and Lamias, M.J. (2001). Web Survey Design and Administration. Public Opinion Quarterly, 65, 230–253.

Crawford, S.D., Couper, M.P., and Lamias, M.J. (2001). Web Surveys: Perceptions of Burden. Social Science Computer Review, 19, 146–162.

Dillman, D.A. (1978). Mail and Telephone Surveys, The Total Design Method. New York, NY: John Wiley.

Dillman, D.A. (1991). The Design and Administration of Mail Surveys. Annual Review of Sociology, 17, 225–249.

Dillman, D.A. (2000). Mail and Internet Surveys: The Tailored Design Method. New York, NY: John Wiley.

Dillman, D.A., Sinclair, M. D., and Clark, J. R. (1993). Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. Public Opinion Quarterly, 57, 289–304.

Dillman, D.A., Tortora, R.D., Conradt, J., and Bowker, D. (1998). Influence of Plain vs Fancy Design on Response Rates for Web Surveys. Proceedings of the American Statistical Association, Survey Research Methods Section, Dallas, TX.

Fowler, Jr., F.J. (1993). Survey Research Methods. Newbury Park, CA: Sage.

Goyder, J. (1982). Further Evidence on Factors Affecting Response Rates to Mailed Questionnaires. American Sociological Review, 47, 550–553.

Groves, R.M. Cialdini, R.B., and Couper, M.P. (1992). Understanding the Decision to Participate in a Survey. Public Opinion Quarterly, 56, 475–495.

Groves, R.M. and Couper, M.P. (1998). Nonresponse in Household Interview Surveys. New York, NY: John Wiley.

Harvey, L. (1987). Factors Affecting Response Rates to Mailed Questionnaires: A Comprehensive Literature Review. Journal of Marketing Research Society, 29, 342–353.

Heberlein, T.A. and Baumgartner, R. (1978). Factors Affecting Response Rates in Mailed Questionnaires: A Quantitative Analysis of Published Literature. American Sociological Review, 43, 447–462.

James, J.M. and Bolstein, R. (1990). Effect of Monetary Incentives and Follow-up on the Response Rate and Response Quality in Mail Surveys. Public Opinion Quarterly, 54, 346–361.

Linsky, A.S. (1975). Stimulating Responses to Mailed Questionnaires: A Review. Public Opinion Quarterly, 39, 82–101.

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

Scheiderman, B. (1998). Designing the User Interface. Reading, MA: Addison Wesley.

Tourangeau, R., Conrad, F.G., and Couper, M.P. (2004). Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. Public Opinion Quarterly, 68, 368–393.

Tourangeau, R., Crawford, S.D., Conrad, F.G., and Couper, M.P. (2004). What They See Is What We Get, Response Options for Web Surveys. Social Science Computer Review, 22, 111–127.

Zuckerman, A., Nichols, E., and Tedesco, H. (1999). Designing Surveys for the Next Millennium: Internet Questionnaire Design Issues. Paper presented at the Annual Meeting of the American Association of Public Opinion Research.