*Public Opinion Quarterly*, Vol. 72, No. 5 2008, pp. 914–934

# COMPLETION TIME AND RESPONSE ORDER EFFECTS IN WEB SURVEYS

NEIL MALHOTRA

**Abstract** The use of the World Wide Web to conduct surveys has grown rapidly over the past decade, raising concerns regarding data quality, questionnaire design, and sample representativeness. This research note focuses on an issue that has not yet been studied: Are respondents who complete self-administered Web surveys more quickly—perhaps taking advantage of participation benefits while minimizing effort—also more prone to response order effects, a manifestation of "satisficing"? I surveyed a random sample of the US adult population over the Web and manipulated the order in which respondents saw the response options. I then assessed whether primacy effects were moderated by the overall length of time respondents took to complete the questionnaires. I found that low-education respondents who filled out the questionnaire most quickly were most prone to primacy effects when completing items with unipolar rating scales. These results have important implications for various aspects of Web survey methodology including panel management, human–computer interaction, and response order randomization.

The use of the World Wide Web to conduct surveys has grown rapidly over the past decade, with researchers in government, academia, and the private sector increasingly making use of the Internet to collect data. The advantages of this mode have been extensively documented: cost-effectiveness, large samples, the ability to administer intricate graphics and multimedia to respondents, and efficiency of data management, among others (e.g., Couper 2000; Benfield and Szlemko 2006; Van Selm and Jankowski 2006). However, as Web

NEIL MALHOTRA is with the Department of Political Science, Stanford University, Encina Hall West, Room 100, Stanford, CA 94305-6044, USA. I gratefully acknowledge Jon Krosnick, Paul Sniderman, Jed Stiglitz, Daniel Schneider, John Bullock, Matt Levendusky, Elizabeth Popp, Alex Kuo, and anonymous reviewers for valuable suggestions. I thank Diana Mutz of TESS and Poom Nukulkij of Knowledge Networks for their time and assistance. The data collection was funded by Timeshare Experiments in the Social Sciences (TESS). Address correspondence to Neil Malhotra; e-mail: neilm@stanford.edu.

doi:10.1093/poq/nfn050 Advance Access publication December 4, 2008

surveys have become more popular, concerns regarding data quality and sample representativeness have increased in importance.[1]

This research note focuses on a concern that has not yet been studied: Do respondents who complete Web surveys more quickly also produce data of lower quality? This analysis provides initial evidence that addresses this question, laying a foundation for future research. Understanding the relationship between completion time and data quality is important because of the potentially perverse incentives of participants in Internet panels. Respondents in Web surveys generally receive some benefit in exchange for their continued participation in a panel.[2] Once respondents gain experience with Web surveys, some may perceive an incentive to complete subsequent questionnaires as quickly as possible and collect the benefit. In other words, thoughtfully considering the questions may become the altruistic choice; the purely rational approach is to "satisfice," or minimize effort in responding to surveys and simply provide the appearance of compliance (Krosnick and Alwin 1987; Krosnick 1991). Because Web surveys are self-administered, an interviewer is almost never present to pace the respondent. Although Internet data collection firms do their best to weed out panelists who produce "junk" data via *strong* satsificing, the detection of *weak* satisficers who are somewhat disengaged with the task may be more difficult.

There is a limited amount of existing literature to draw upon in addressing these questions. Yan and Tourangeau (2008) have found that response time to individual items presented over the Web is related to both respondent characteristics (e.g., age, education, experience with Internet-based questionnaires) and item characteristics (e.g., the number and type of response categories, question location).[3] Whereas the existing research has focused on response time at the *item* level, this note examines the effect of completion time on survey satisficing

1. Accordingly, research on Web survey methodology is extensive; the online database WebSM (www.websm.org) has collected over 2,000 articles written on the topic. Many of these studies focus on issues with Internet-based data collection such as coverage and nonresponse error (e.g., Lee 2006; Couper et al. 2007), strategies for increasing response rates (e.g., Jones and Pitt 1999; Crawford, Couper, and Lamias 2001), panel effects (e.g., Dennis 2001; Coen, Lorch, and Piekarski 2005), questionnaire design (e.g., Couper, Traugott, and Lamias 2001; Couper et al. 2004; Tourangeau, Couper, and Conrad 2004), and statistical inference (Malhotra and Krosnick 2007; Sanders et al. 2007).

2. For example, panelists may receive continued service such as WebTV (e.g., Knowledge Networks), or are encouraged to participate in future surveys through the offering of points in exchange for merchandise or chances to win lotteries (e.g., Harris Interactive). Some scholars have explored whether long-term participation in panels affects the response options chosen by "professional" survey respondents (e.g., Dennis 2001). This paper focuses on a different issue: Does panel participation create incentives for respondents to complete the questionnaire as quickly as possible without devoting enough attention to the items?

3. Other studies have also examined the relationship between response time and item design (e.g., Heerwegh 2002; Draisma and Dijkstra 2004; Tourangeau, Couper, and Conrad 2004). Another set of investigations has not directly examined data quality as a dependent variable, instead focusing on the relationship between response time and attitude stability (e.g., Heerwegh 2003).

at the *questionnaire* level, investigating whether respondents who rush through provide less thoughtful answers overall.

Research on survey satisficing has identified several signs of poor data quality and cognitive detachment from the task: response order effects, "don't know" responses, nondifferentiability in response, and mental coin flipping (Krosnick 1991; Chang and Krosnick 2002). This analysis focuses on the examination of response order effects because they can be *exogenously* induced by manipulating the order in which response options are presented to participants. In other words, if respondents are carefully considering the item and reporting their true attitudes, then the ordering of the response choices should have no impact on the options that are selected. In a meta-analysis of several studies of order effects in categorical items and ordinal rating scales, Krosnick and Fabrigar (forthcoming) demonstrate that primacy effects (bias toward selecting earlier response choices) occur in visually presented items, a sign that some respondents satisfice by selecting the first acceptable option and not considering the entire set. It is important to note that this analysis has implications beyond order effects, which I use as an indicator of data quality. Future research can explore explicitly other commonly identified signs of satisficing—large numbers of "don't know" responses, "straight lining" of answers, and low correlations between theoretically related variables.

To address the questions above, I conducted a survey administered by Knowledge Networks (KN) over the Web that included questions with both ordinal, unipolar rating scales as well as questions with categorical response options. I manipulated the order in which respondents saw the response options, and assessed whether primacy effects were moderated by the overall length of time respondents took to complete the questionnaires. Below, I present a theoretical overview of the hypothesized relationship between completion time and response order effects. I then describe the data, experimental design, methods of analysis, and the results. Finally, I discuss the implications of my findings for Web survey methodology.

## Theoretical Overview

Response order effects within questions are presumed to be mediated by three factors: survey satisficing, memory limitations, and cognitive elaboration (Sudman, Bradburn, and Schwarz 1996). Further, these mechanisms are contingent on survey mode. According to the satisficing explanation, respondents select the first alternative that is reasonably acceptable (and sometimes simply the first option period) and do not take the time to fully consider the entire set of response options, resulting in primacy effects in visually presented items (Krosnick 1991). The memory limitations mechanism posits that respondents are not able to remember full lists of items and are prone to recency effects in orally presented items because they are more likely to remember the last item they heard (Smyth et al. 1987). Finally, the cognitive elaboration model

hypothesizes that each response alternative induces argumentation within the respondent's mind, meaning that responses listed higher up on visually presented scales that elicit positive, agreeing thoughts are more likely to be selected (Krosnick and Alwin 1987; Schwarz, Hippler, and Noelle-Neumann 1992).

However, primacy effects in *visually presented*, *verbally labeled rating scales* likely are only the consequence of satificing, and not of memory or elaboration effects. According to Sudman, Bradburn, and Schwarz (1996): "Memory limitations are most likely to play a role in the emergence of response order effects when numerous or complex response alternatives are presented without the help of show cards, thus taxing respondents' memory" (p. 137). Additionally, the authors write: "Verbal rating scales do not present different substantive options that respondents need to elaborate on. Rather, terms ranging from 'like very much' to 'dislike very much' constitute a response continuum where the different options require far less elaboration than is the case with opinion questions [with categorical response options]" (p. 157). The hypothesized relationship between primacy effects and rapid completion time is based on the possibility that respondents satisfice and do not consider all available options. Conversely, the processes of memory limitations and cognitive elaboration suggest that primacy effects should be related to more extensive and lengthy search. Hence, I predict that: the relationship between primacy effects and faster completion time should exist for unipolar rating scales.[4]

Moreover, the cause of faster completion times may vary for different groups of respondents. Indeed, the relationship between completion time and data quality is complicated, and fast responses are not necessarily a perfect measure of low attention to the questionnaire. A respondent may respond to an item quickly because he or she has limited cognitive skills or motivation, and consequently does not think extensively about the question (Krosnick, Narayan, and Smith 1996; Narayan and Krosnick 1996). Alternatively, opinionated, knowledgeable respondents with strong attitudes may also report their responses more quickly (e.g., Krosnick 1989; Bassili 1993). Combining these two groups should therefore dilute the observed relationship between primacy effects and faster completion times. Hence, I predict that: the relationship between primacy effects and completion time should be strongest among respondents with limited cognitive skills.

## Data

I conducted a survey experiment on attitudes toward the government response to Hurricane Katrina in the city of New Orleans. The survey was administered by KN over the Internet between May 26 and May 31, 2006, using a nationally

4. This is not to say that the relationship between primacy effects and completion time should be completely absent in categorical items, but existing theory suggests that such a relationship should be weaker.

representative sample of 397 American adults recruited via random digit dialing (RDD).[5] All data are weighted by demographics.[6]

## Experimental Design

I asked five questions, measured on five-point ordinal, unipolar rating scales dealing with interest, attention, importance, anger, and sadness with respect to Hurricane Katrina. I also asked six questions, each with four categorical response options, about citizens' attitudes toward the responsibilities and trustworthiness of various political actors. The complete wordings of the questions and response options can be found in the Appendix.

Half of the respondents were assigned to receive the response options in the order listed in the first column of the Appendix. I refer to this as the "control" condition. The other half were assigned to the "reverse order" condition, in which I reverse the order in which response options are seen (see column two of the Appendix). Response options listed closer to the top of the computer screen in the control condition were coded with higher numbers. Therefore, in the reverse order condition, response options coded with higher numbers are located toward the bottom of the screen. Below, when I refer to the "top two" response options, I mean the options listed highest up in the control condition.

I also recorded the amount of time (in minutes) it took for respondents to complete the questionnaire.[7] Because completion time data are skewed, I normalized the values by taking their natural logs, following previous studies in cognitive psychology (e.g., Fazio 1990; Yan and Tourangeau, 2008). Additionally, I dropped extreme outliers—the 11 respondents whose completion times were more than one standard deviation greater than the mean (i.e., respondents whose completion times were greater than or equal to 51.7 minutes).[8] This breakpoint represents the main discontinuity in the

5. KN selects households using RDD and provides selected households with free hardware and Internet access in exchange for completing one questionnaire per week. Some 584 panelists were randomly drawn from the KN panel; 397 responded to the invitation, yielding a final stage completion rate of 68.5 percent. Five respondents were excluded because completion time data were not recoverable. The recruitment rate for this study, reported by Knowledge Networks, was .294 and the profile rate was .569, for a cumulative response rate of .294*.573*.685 = 0.1153 = 11.53 percent.

6. Poststratification weights are derived from weighted sample distributions along combinations of the following variables: gender, age groups, race, region, and education. Distributions are calculated using recent U.S. Census Bureau data and KN panel data. Cell-by-cell adjustments over the various univariate and bivariate distributions are calculated to make the weighted sample cells match those of the U.S. Census and KN panel.

7. There was variance in the speed of Internet connectivity across respondents, but KN recommends that all respondents use Internet Explorer. In correspondence with KN staff, I learned that in surveys without images and multimedia (such as the one analyzed here), differences in connectivity speed do not produce meaningful differences in completion times.

8. It is possible that these respondents might have been interrupted while filling out the questionnaire.

**Table 1.** Descriptive Statistics

| | Completion time (minutes) | Log completion time | Age (years) |
|---|---|---|---|
| Mean | 7.16 | 1.80 | 46.7 |
| Median | 5.85 | 1.77 | 44 |
| Standard deviation | 4.67 | 0.56 | 16 |
| Minimum | 1.57 | 0.45 | 18 |
| Maximum | 31.37 | 3.45 | 91 |
| 33rd percentile | 4.65 | 1.54 | 37 |
| 67th percentile | 7.65 | 2.03 | 55 |

NOTE.—$n = 381$.

distribution of completion times—there are no respondents with completion times between 31.4 and 51.7 minutes.[9]

I did not record completion times for each individual item, but believe that total completion time more accurately taps the construct of overall attention and time spent on the questionnaire. Yan and Tourangeau (2008) found several item-specific features related to response time such as question length, the substantive construct being measured, and the position of the item within the questionnaire. Total completion time averages out idiosyncrasies unique to each item and better taps the level of "satisficing" present throughout the survey. Moreover, because the analyses are conducted at the level of the questionnaire, total completion time is the relevant measure. Descriptive statistics for completion time and age are presented in table 1.

## Methods

To assess whether completion time moderated order effects in the rating scales, I estimated the following structural component of a Poisson regression model,[10] indexed by each respondent $i$:

$$E[Y_i \,|x_i\,] = \exp\{\beta_0 + \beta_1 R_i + \beta_2 T_i + \beta_3 (R_i \times T_i) + \beta_4 A_i + \beta_5 (R_i \times A_i)\} \tag{1}$$

where $Y_i$ represents the number of items in the questionnaire for which respondent $i$ selected one of the "top two" response options as they were presented in the control group, $R_i$ is a dummy variable representing whether the respondent received the reverse order treatment, $T_i$ represents the natural log of the time taken to complete the survey in minutes, and $A_i$ represents the respondent's

9. These procedures produce an unskewed distribution. The mean of log completion time is 1.80, and the median is 1.77.

10. A likelihood-ratio test revealed that there was no overdispersion in the data. I also estimated negative-binomal regression models and the results were statistically and substantively similar to those reported in this paper.

age.[11] I expect the estimate of $\beta_1$ to be negative and statistically significant if there is a primacy effect because response options listed closer to the top in the control condition (i.e., $R_i = 0$) are coded as "higher" categories. If respondents who complete the survey more quickly produce lower quality data, then the coefficient on the interaction between the reverse order treatment dummy and completion time ($\beta_3$) should be positive and statistically significant. In other words, spending more time on the survey should mitigate the primacy effect captured by $\beta_1$ (i.e., dampen the marginal effect of the reverse order treatment, $\beta_1 + \beta_3$). As discussed below, I found completion time to be positively correlated with respondent age. Hence, I control for any potential confounding effect of age by including it in the regression model, along with its interaction with the reverse order treatment dummy.

Additionally, as previously explained, the relationship between primacy effects and completion time should be strongest for the unipolar rating scales and among respondents with low cognitive skills, which I proxy using level of education.[12] Hence, I estimate equation (1) for various subsets of the data, including the subset of the five items using rating scales, and among low-education respondents. Throughout the paper, I classify low-education respondents as those with a high school diploma or less, and high-education respondents as those with schooling beyond high school.

To better understand the effect of completion time on satisficing in substantive terms, I calculate the "treatment effect" of reversing the order of the response options for values of log completion time between 0.5 and 3 (1.65 and 20.09 minutes), holding age at its mean. The treatment effect is defined as the difference between the control and reverse order conditions in the expected count of selecting one of the top two response options. Because the highest coded options are listed on the bottom in the reverse order condition, the treatment effect should be negative in the presence of primacy. Confidence intervals of the treatment effects are calculated via the delta method (Xu and Long 2005).

## Results

### DESCRIPTIVE STATISTICS

Before delving into the results from the statistical model, I present descriptive statistics on completion time. I estimated an OLS regression predicting log

11. Age is coded to lie between 0 and 1, with 0 being the youngest respondent in the dataset and 1 the oldest.

12. Respondents with strong attitudes may also answer questions quickly without satisficing. In Question 3, I measured how personally important Hurricane Katrina was to the respondent. However, because this item was measured using a unipolar rating scale, and because the order of responses was manipulated, it is not an ideal candidate to use as a moderating variable.

**Table 2.** OLS Regressions Predicting Log Completion Time with Demographics

| | |
|---|---|
| Age | 0.87** (0.15) |
| High school | 0.20* (0.11) |
| Some college | 0.08 (0.12) |
| College | 0.05 (0.12) |
| White | −0.04 (0.08) |
| Male | −0.02 (0.07) |
| Party identification | −0.04 (0.11) |
| Constant | 1.42** (0.14) |
| *n* | 381 |
| $R^2$ | 0.13 |

NOTE.—Data are weighted.
**$p < .01$; *$p < .05$ (one-tailed).

completion time with a set of available demographic and political variables: age, education, race, gender, and party identification (see table 2).[13] The coefficients on race, gender, and party identification are insignificant, and I fail to reject the joint test that the three education dummy variables are all equal to zero. However, older respondents took significantly longer to complete the questionnaire. Consequently, I control for the main and interactive effects of age in the regression models, as discussed above.

Respondents with low education who completed the questionnaire most quickly were more likely to select response options near the top of the list. Table 3 presents the average number of items (out of five unipolar rating scales and six categorical items) for which respondents selected one of the top two response options in the control condition (e.g., "extremely interested" or "very interested" for Question 1) by condition, education, and completion time tercile. For instance, in the control condition, low-education respondents in the quickest tercile selected one of the top two options for 2.25 (of five) unipolar rating scales (see the value in the top line of table 3). When the order of the response options was reversed, these respondents selected one of the top two options for only 0.63 items, yielding a primacy effect of 1.62 items. This effect is substantively large, encompassing over 40 percent of the total number of unipolar rating scales.

Consistent with my theoretical expectations, the negative relationship between primacy effects and completion time is only present for low-education

13. As mentioned above, age is coded to lie between 0 and 1. I tried including age squared in the regression model but it was not statistically significant. The other demographic variables are coded as dummy variables with respondents with less than a high school education, nonwhites, and females as the baseline categories. The standard seven-point party identification scaled used by the American National Election Study is coded to lie between 0 and 1, with 0 representing "strong Democrats" and 1 representing "strong Republicans."

**Table 3.** Moderating Effect of Completion Time on Primacy Effects by Level of Education

| | Completion time: Quickest tercile | | | Completion time: Middle tercile | | | Completion time: Slowest tercile | | |
|---|---|---|---|---|---|---|---|---|---|
| | Control | Reverse | Difference | Control | Reverse | Difference | Control | Reverse | Difference |
| Low education ($n = 175{,}171$) | | | | | | | | | |
| Unipolar rating scales | **2.25** | **0.63** | **1.62** | **2.60** | **1.43** | **1.17** | **3.10** | **2.25** | **0.85** |
| Categorical items | 2.73 | 2.30 | 0.43 | 2.09 | 2.72 | −0.63 | 3.28 | 3.20 | 0.08 |
| High education ($n = 204{,}201$) | | | | | | | | | |
| Unipolar rating scales | 2.02 | 1.87 | 0.15 | 2.93 | 2.40 | 0.53 | 3.11 | 3.04 | 0.07 |
| Categorical items | 3.62 | 2.67 | 0.95 | 3.14 | 2.62 | 0.52 | 3.23 | 2.72 | 0.51 |

NOTE.—Average number of items in which the respondent selected one of the top two response options as presented in the "control" condition. Data are weighted. Low-education respondents defined as those possessing a high school diploma or less.

respondents completing unipolar rating scales (see the bold values in the top line of table 3). For the other subgroups, the relationship is smaller and more inconsistent.[14] Finally, note that the strongest primacy effect is found among low-education respondents who were in the quickest tercile of completion time. Primacy effects decreased as these low-education respondents took more time to complete the survey. Furthermore, consistent with predictions, high-education respondents who completed the survey most slowly exhibit absolutely no primacy effect—the difference between the control and reverse order conditions is nearly zero.

Another way of presenting the data is to examine the average proportion of respondents who selected each response option (as listed in the control condition) averaged across the five unipolar rating scales. As shown in the top line of table 4, 24.9 percent of low-education respondents in the quickest completion time tercile selected the top response option when it was presented first, but 0.0 percent selected it when it was presented last, yielding a primacy effect of 24.9 percent. For each subgroup, I calculate an average primacy effect across the five response options by (1) multiplying the differences between conditions by 1 for the top two response options; (2) multiplying the differences between conditions by $-1$ for the bottom three response options; and (3) calculating the mean across the five items.

As shown in the top panel of table 4, the average primacy effect decreases monotonically for low-education respondents as they take more time in completing the questionnaire. Indeed, the effect among the quickest tercile is nearly double that among the slowest tercile. This substantively large, monotonic pattern is not evident among high-education respondents (see the bottom panel).

## MODEL ESTIMATES

To formally test the patterns observed in tables 3 and 4, I conducted Poisson regression models (as represented by equation 1) predicting the number of items for which respondents selected one of the top two response options (as presented in the control condition). The third column of table 5 presents the coefficient estimates for the model predicting responses to the unipolar rating scales for low-education respondents. Consistent with my theoretical predictions, the estimate of the variable representing the response order manipulation ("Reverse order") is negative and statistically significant ($\beta_1 = -1.10$, $p = .037$). Respondents who completed the questionnaire most quickly exhibited significant primacy effects—the likelihood of selecting one of the top two response options decreases when they are placed on the bottom. Also, as

14. The descriptive statistics suggest that there may be a non-linear relationship between completion time and primacy effects for high-education respondents. I included squared terms for completion time in equation (1) to test this possibility but found no statistically significant evidence of nonlinearity.

**Table 4.** Moderating Effect of Completion Time on Primacy Effects for Unipolar Rating Scales

| | Completion time: Quickest tercile | | | Completion time: Middle tercile | | | Completion time: Slowest tercile | | |
|---|---|---|---|---|---|---|---|---|---|
| | Control | Reverse | Difference | Control | Reverse | Difference | Control | Reverse | Difference |
| Low education ($n = 175$) | | | | | | | | | |
| First response option | 24.9% | 0.0% | 24.9% | 19.5% | 11.0% | 8.6% | 27.5% | 12.0% | 15.5% |
| Second response option | 20.0 | 12.6 | 7.4 | 32.5 | 17.6 | 14.9 | 34.5 | 33.0 | 1.5 |
| Third response option | 31.5 | 34.9 | −3.5 | 29.8 | 36.1 | −6.4 | 23.3 | 38.5 | −15.2 |
| Fourth response option | 14.6 | 29.3 | −14.7 | 9.9 | 25.2 | −15.3 | 6.7 | 11.5 | −4.8 |
| Fifth response option | 9.1 | 23.2 | −14.1 | 8.4 | 10.2 | −1.8 | 8.1 | 4.9 | 3.1 |
| Average primacy effect | | | 12.9% | | | 9.4% | | | 6.8% |
| High education ($n = 204$) | | | | | | | | | |
| First response option | 15.7% | 12.7% | 3.0% | 26.0% | 18.5% | −0.7% | 17.2% | 26.5% | −9.4% |
| Second response option | 25.6 | 24.7 | 0.8 | 32.6 | 29.6 | −4.9 | 45.0 | 34.3 | 10.7 |
| Third response option | 25.1 | 38.4 | −13.3 | 27.0 | 31.9 | −4.9 | 18.1 | 19.4 | −1.3 |
| Fourth response option | 21.6 | 17.3 | 4.3 | 9.2 | 14.1 | 3.0 | 13.9 | 10.0 | 3.9 |
| Fifth response option | 12.1 | 6.9 | 5.1 | 5.2 | 6.0 | 7.5 | 5.8 | 9.8 | −3.9 |
| Average primacy effect | | | 1.5% | | | 4.2% | | | 0.5% |

NOTE.—Average proportions of respondents selecting each response alternative across all five items. Data are weighted. Low-education respondents defined as those possessing a high school diploma or less.

**Table 5.** Poisson Regressions Predicting Number of Items for Which Respondents Selected Top Two Options

| | All items | | Unipolar rating scales | | Categorical items | |
|---|---|---|---|---|---|---|
| | Low education | High education | Low education | High education | Low education | High education |
| Reverse order | −0.44 (0.34) | −0.44* (0.20) | −1.10* (0.61) | −0.49 (0.34) | −0.02 (0.37) | −0.36 (0.29) |
| Log completion time | 0.07 (0.11) | 0.08 (0.08) | 0.01 (0.16) | 0.23* (.11) | 0.14 (0.14) | −0.04 (0.09) |
| Reverse order × Log completion time | 0.28* (0.16) | 0.05 (0.13) | 0.70** (0.27) | 0.08 (0.17) | −0.01 (0.18) | −0.01 (0.19) |
| Age | 0.82*** (0.18) | 0.17 (0.26) | 1.42*** (0.29) | 0.61 (0.38) | 0.25 (0.24) | −0.21 (0.30) |
| Reverse order × Age | −0.65* (0.38) | 0.51 (0.39) | −1.69** (0.63) | 0.67 (0.51) | 0.14 (0.39) | 0.38 (0.53) |
| Constant | 1.17*** (0.24) | 1.58*** (0.15) | 0.28 (0.34) | 0.30 (0.24) | 0.63* (0.30) | 1.36*** (0.15) |
| $n$ | 171 | 200 | 175 | 204 | 171 | 201 |
| $\chi^2$ (5) | 45.85*** | 30.29*** | 55.65*** | 43.01*** | 7.26 | 9.08 |

NOTE.—Data are weighted. Low-education respondents defined as those possessing a high school diploma or less.
***$p < .001$; **$p < .01$; *$p < .05$ (one tailed).

expected, the interaction term between the response order manipulation and log completion time is *positive* and statistically significant ($\beta_3 = 0.70$, $p = .005$), meaning that primacy effects were mitigated as low-education respondents took more time in answering the questions. As per theoretical expectations, similar effects were not observed for unipolar rating scales when examining high-education respondents because, in this subgroup, faster completion times may represent attitude strength and crystallized opinion formation. Moreover, I did not see primacy effects when predicting responses to the categorical items.

Because coefficient estimates from Poisson regression models are not easily interpretable, I plot the size of the treatment effects by education and completion time (holding age at its mean) for the unipolar rating scales in figure 1. The primacy effect is the *difference* between the control and reverse order conditions in the expected count (i.e., number) of items for which respondents selected one of the top two response options. If this difference is negative, then it means that a primacy effect is present—respondents select one of the top two options less when they are listed on the bottom.

As shown in the top panel of the figure, when low-education respondents fill out the questionnaire quickly (log completion time $\approx 0.5$), the size of the primacy effect is about two items, representing a substantively large 40 percent of the five unipolar rating scales. The shaded 90 percent confidence interval also suggests that this effect is significantly different from zero. However, as completion time increases, the primacy effect disappears until it is statistically indistinguishable from zero. Hence, low-education respondents who completed the survey more slowly were less likely to be affected by the response order manipulations. As illustrated in the bottom panel of figure 1, this was not the case for high-education respondents. The moderating effect of completion time on primacy effects is essentially flat—those who completed the questionnaire quickly were as likely to be unaffected by the response order manipulations as those who completed the questionnaire more slowly.

Lastly, I estimate individual OLS regressions predicting responses to each of the five unipolar rating scales for low-education respondents.[15] As shown in table 6, the interaction term between the order manipulation and log completion time is positive and statistically significant in four of five cases, producing the aggregate results I observed earlier.

## Discussion

These findings have important implications for designers, administrators, and users of Web surveys. First, they suggest that low-quality data may still be produced even when the most egregious panel participants are removed. Accordingly, extremely quick completion time may be a valuable criterion in

15. I also estimated ordered logistic regression models, and the results were statistically and substantively similar. I present OLS regressions for ease of interpretation.

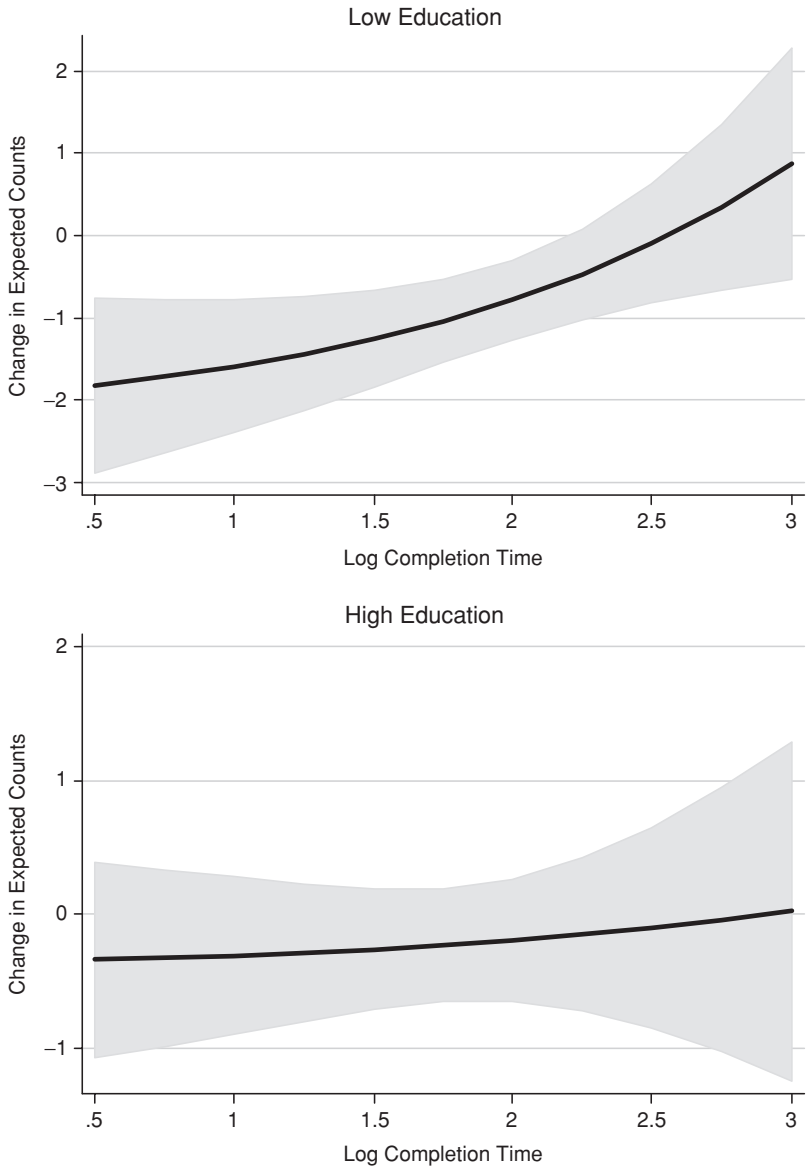


**Figure 1.** Change in Expected Counts of Number of Items in Which Top Two Options are Selected Due to Response Order Manipulation (Unipolar Rating Scales).

NOTE.—Difference between the control and reverse order conditions in the expected counts of items in which one of the top two response options is selected plotted on the *y*-axis, holding "age" variable at its mean. Ninety percent confidence intervals of treatment effects shaded. Low-education respondents defined as those possessing a high school diploma or less.

**Table 6.** OLS Regressions Predicting Rating Scale Responses for Unipolar Rating Scales (Low Education)

| | Interest | Attention | Importance | Anger | Sadness |
|---|---|---|---|---|---|
| Reverse order | −1.01 (0.66) | −1.63** (0.56) | −0.46 (0.67) | −1.68* (0.84) | −0.02 (0.81) |
| Log completion time | −0.12 (0.25) | −0.13 (0.22) | 0.04 (0.28) | −0.40 (0.32) | 0.62* (0.35) |
| Reverse order × Log completion time | 0.75* (0.32) | 1.00** (0.30) | 0.62* (0.35) | 1.05** (0.41) | 0.24 (0.40) |
| Age | 1.91*** (0.53) | 2.21*** (0.50) | 2.01*** (0.54) | 1.18* (0.63) | 1.52** (0.56) |
| Reverse order × Age | −2.51** (0.88) | −2.04** (0.63) | −2.83** (0.82) | −1.13 (0.83) | −2.14** (0.73) |
| Constant | 3.19*** (0.50) | 2.84*** (0.43) | 2.31*** (0.52) | 3.29*** (0.67) | 1.98** (0.70) |
| $R^2$ | .22 | .30 | .21 | .07 | .23 |

NOTE.—Data are weighted. $n = 175$ for all models. Low-education respondents defined as those possessing a high school diploma or less.
***$p < .001$; **$p < .01$; *$p < .05$ (one-tailed).

filtering out participants or their data. As shown in figure 1, the largest primacy effects were observed for respondents whose completion times were more than one-and-a-half standard deviations below the mean (log completion time <0.96 (2.61 minutes)), representing about 5 percent of the sample. However, the findings also indicate an important caveat. Completion time, *in and of itself*, may not be an optimal filtering criterion because it is only a proxy for satisficing among a select subgroup of respondents (those with the lowest cognitive skills) and for a select subgroup of items (unipolar rating scales). Nevertheless, my results suggest that researchers should include completion time as a control variable in statistical models, and reported results should be robust to the removal of completion time outliers.

Second, my results indicate that a better method of dealing with these issues is to introduce pacing in Web surveys. Unlike face-to-face or telephone modes, the absence of a live interviewer allows the respondent to complete the questionnaire at his or her own pace. However, it is possible that features that slow down their completion of self-administered questionnaires may frustrate some respondents. Recent work by Tourangeau, Couper, and Steiger (2003) and Conrad, Schober, and Coiner (2007) suggests that the introduction of human dialog (perhaps through voice and/or video of humans) to guide the respondent and lengthen completion time may result in higher data quality. Survey researchers may benefit from collaborations with experts on human–computer interaction to enhance the survey experience for respondents and simultaneously produce higher quality data for investigators.

Finally, the findings underscore the importance of including response order manipulations in questionnaires to mitigate the impact of survey satisficing, particularly among respondents who are rushing to complete the items. In the presence of primacy effects, randomizing order does not eliminate measurement error, but will reduce bias.[16]

The studies presented here lay the foundation for future research exploring the relationship between completion time and satisficing in Web surveys. First, it would be beneficial to replicate these findings in items that employ different types of tasks and deal with other substantive domains. Further, other signs of satisficing including nondifferentiability and "don't know" responses can be analyzed. Third, it may also be instructive to examine completion time of individual items (or groups of items). Following Yan and Tourangeau (2008), it may be that the moderating effect of completion time on data quality only exists for certain types of items located in certain portions of the questionnaire. Finally, replicating these analyses using data collection firms that employ convenience

16. Additionally, changing the incentive structure in Internet panels may also be necessary to ameliorate satisficing and discourage respondents from racing through the questionnaire to collect the benefit. However, because I am unsure of how this could be accomplished, I remain more tentative on this recommendation.

samples can assess whether the findings are robust to different participant pools and incentive structures.

Web surveys will become an integral part of the future of survey research, and have already presented researchers with unique opportunities to collect data that were unavailable in the pre-Internet world. However, along with the promise of Web surveys comes the peril that self-administration and panel structures create incentives for some respondents to quickly complete questionnaires without paying proper attention to the items. This paper provides evidence that respondents with relatively lower cognitive skills who take less time to complete Web surveys satisfice and produce lower quality data in the form of order effects. Identifying these respondents and attempting to make them more cognitively engaged with the task are important issues to address as Web surveys become more prevalent.

## Appendix: Wording and Coding of Items

In this appendix, I present the wordings and codings of the 11 items for the control and reverse order conditions.

1. How interested were you in the events surrounding Hurricane Katrina? ("Interest")

| Control condition | Reverse order condition |
|---|---|
| Extremely interested (5) | Not interested at all (1) |
| Very interested (4) | Slightly interested (2) |
| Moderately interested (3) | Moderately interested (3) |
| Slightly interested (2) | Very interested (4) |
| Not interested at all (1) | Extremely interested (5) |

2. How closely did you follow the media coverage of the events surrounding Hurricane Katrina? ("Attention")

| Control condition | Reverse order condition |
|---|---|
| Extremely closely (5) | Not closely at all (1) |
| Very closely (4) | Slightly closely (2) |
| Moderately closely (3) | Moderately closely (3) |
| Slightly closely (2) | Very closely (4) |
| Not closely at all (1) | Extremely closely (5) |

3. How personally important to you were the events surrounding Hurricane Katrina? ("Importance")

| Control condition | Reverse order condition |
|---|---|
| Extremely important (5) | Not important at all (1) |
| Very important (4) | Slightly important (2) |
| Moderately important (3) | Moderately important (3) |
| Slightly important (2) | Very important (4) |
| Not important at all (1) | Extremely important (5) |

4. How angry does Hurricane Katrina and its aftermath make you feel? ("Anger")

| Control condition | Reverse order condition |
|---|---|
| Extremely angry (5) | Not angry at all (1) |
| Very angry (4) | Slightly angry (2) |
| Moderately angry (3) | Moderately angry (3) |
| Slightly angry (2) | Very angry (4) |
| Not angry at all (1) | Extremely angry (5) |

5. How sad does Hurricane Katrina and its aftermath make you feel? ("Sadness")

| Control condition | Reverse order condition |
|---|---|
| Extremely sad (5) | Not sad at all (1) |
| Very sad (4) | Slightly sad (2) |
| Moderately sad (3) | Moderately sad (3) |
| Slightly sad (2) | Very sad (4) |
| Not sad at all (1) | Extremely sad (5) |

6. Who should be most responsible for responding to natural disasters? ("Level: Respond")

| Control condition | Reverse order condition |
|---|---|
| Federal authorities (4) | Private citizens (1) |
| State authorities (3) | Local authorities (2) |
| Local authorities (2) | State authorities (3) |
| Private citizens (1) | Federal authorities (4) |

7. Who should be most responsible for helping to protect citizens from natural disasters before they happen? ("Level: Protect")

| Control condition | Reverse order condition |
|---|---|
| Federal authorities (4) | Private citizens (1) |
| State authorities (3) | Local authorities (2) |
| Local authorities (2) | State authorities (3) |
| Private citizens (1) | Federal authorities (4) |

8. Which party do you trust more to effectively respond to natural disasters? ("Party: Respond")

| Control condition | Reverse order condition |
|---|---|
| Republican party (4) | Neither (1) |
| Democratic party (3) | Both equally (2) |
| Both equally (2) | Democratic party (3) |
| Neither (1) | Republican party (4) |

9. Which party do you trust more to help protect citizens from natural disasters before they happen? ("Party: Protect")

| Control condition | Reverse order condition |
|---|---|
| Republican party (4) | Neither (1) |
| Democratic party (3) | Both equally (2) |
| Both equally (2) | Democratic party (3) |
| Neither (1) | Republican party (4) |

10. Which party do you trust more to provide social services to American citizens? ("Party: Services")

| Control condition | Reverse order condition |
|---|---|
| Republican party (4) | Neither (1) |
| Democratic party (3) | Both equally (2) |
| Both equally (2) | Democratic party (3) |
| Neither (1) | Republican party (4) |

11. Which party do you trust more to protect the American people? ("Party: Security")

| Control condition | Reverse order condition |
|---|---|
| Republican party (4) | Neither (1) |
| Democratic party (3) | Both equally (2) |
| Both equally (2) | Democratic party (3) |
| Neither (1) | Republican party (4) |

## References

Bassili, John N. 1993. "Response Latency versus Certainty as Indexes of the Strength of Voting Intentions in a Cati Survey." *Public Opinion Quarterly* 57:54–61.

Benfield, Jacob A., and William J. Szlemko. 2006. "Internet-Based Data Collection: Promises and Realities." *Journal of Research Practice* 2:Article D1.

Chang, LinChiat, and Jon A. Krosnick. 2002. "Comparing Self-Administered Computer Surveys and Auditory Interviews: An Experiment." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL, USA.

Coen, Terrence, Jacqueline Lorch, and Linda Piekarski. 2005. "The Effects of Survey Frequency on Panelists' Responses." Paper presented at the ESOMAR Conference on Panel Research, Budapest, Hungary.

Conrad, Frederick G., Michael F. Schober, and Tania Coiner. 2007. "Bringing Features of Human Dialogue to Web Surveys." *Applied Cognitive Psychology* 21:165–87.

Couper, Mick P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464–94.

Couper, Mick P., Arie Kapteyn, Matthias Schonlau, and Joachim Winter. 2007. "Noncoverage and Nonresponse in an Internet Survey." *Social Science Research* 36:131–48.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65:230–53.

Couper, Mick P., Roger Tourangeau, Frederick G. Conrad, and Scott D. Crawford. 2004. "What They See Is What We Get: Response Options for Web Surveys." *Social Science Computer Review* 22:111–27.

Crawford, Scott D., Mick P. Couper, and Mark J. Lamias. 2001. "Web Surveys: Perceptions of Burden." *Social Science Computer Review* 19:146–62.

Dennis, J. Michael. 2001. "Are Internet Panels Creating Professional Respondents?" *Marketing Research* 13:34–8.

Draisma, S., and W. Dijkstra. 2004. "Response Latency and (Para) Linguistic Expressions as Indicators of Response Error." In *Methods for Testing and Evaluating Survey Questionnaires*, eds. S. Presser, J. Rogthgeb, M. Couper, J. Lessler, E. Martin, J. Martin, and E. Singer. Hoboken, NJ: Wiley.

Fazio, Russell H. 1990. "A Practical Guide to the Use of Response Latency in Social Psychological Research." In *Review of Personality and Social Psychology*, eds. C. Hendrick and M. S. Clark. Newbury Park, CA: Sage.

Heerwegh, D. 2002. "Describing Response Behavior in Web Surveys Using Client-Side Paradata." Paper presented at the International Workshop on Web Surveys, Mannheim, Germany.

Heerwegh, D. 2003. "Explaining Response Latencies and Changing Answers Using Client-Side Paradata." *Social Science Computer Review* 21:360–73.

Jones, R., and N. Pitt. 1999. "Health Surveys in the Workplace: Comparison of Postal, E-Mail, and World Wide Web Methods." *Occupational Medicine* 49:556–8.

Krosnick, Jon A. 1989. "Attitude Importance and Attitude Accessibility." *Personality and Social Psychology Bulletin* 15:297–308.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys."*Applied Cognitive Psychology* 5:213–36.

Krosnick, Jon A., and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:201–19.

Krosnick, Jon A., and Lee R. Fabrigar. In press. *The Handbook of Questionnaire Design*. New York: Oxford University Press.

Krosnick, Jon A., Somya Narayan, and William R. Smith. 1996. "Satisficing in Surveys: Initial Evidence." *New Directions for Evaluation* 70:29–44.

Lee, Sunghee. 2006. "An Evaluation of Nonresponse and Coverage Errors in a Prerecruited Probability Web Panel Survey." *Social Science Computer Review* 24:460–75.

Malhotra, Neil, and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 15:286–323.

Narayan, Somya, and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60:58–88.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, and Paul Whiteley. 2007. "Does Mode Matter for Modeling Political Choice? Evidence from the 2005 British Election Study." *Political Analysis* 15:257–85.

Schwarz, Norbert, Hans J. Hippler, and Elisabeth Noelle-Neumann. 1992. "A Cognitive Model of Response Order Effects in Survey Measurement." In *Context Effects in Social and Psychological Research*, eds. Norbert Schwarz and Seymour Sudman. New York: Springer.

Smyth, M. M., P. E. Morris, P. Levy, and A. W. Ellis. 1987. *Cognition in Action*. London: Erlbaum.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz. 1996. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass.

Tourangeau, Roger, Mick P. Couper, and Darby M. Steiger. 2003. "Humanizing Self-Administered Surveys: Experiments on Social Presence in Web and IVR Surveys." *Computers in Human Behavior* 19:1–24.

Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. 2004. "Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions." *Public Opinion Quarterly*.

Van Selm, Martine, and Nicholas W. Jankowski. 2006. "Conducting Online Surveys." *Quality & Quantity* 40:435–56.

Xu, Jun, and J. Scott Long. 2005. "Confidence Intervals for Predicted Outcomes in Regression Models for Categorical Outcomes." *Stata Journal* 5:537–59.

Yan, Ting, and Roger Tourangeau. 2008. "Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology* 22:51–68.