

ARTICLES

Survey Mode and Item Non-Substantive Response: Insights from a Replication Study and Other Recent Trials

Michael Soszynski¹ , Ryan Bliss² 

¹ Senior Evaluation Researcher, ADM Associates, ² Director, ADM Associates

Keywords: item non-response, survey mode, non-substantive response, mode effects, web survey, phone survey, replication study, nonresponse, text-to-web survey

<https://doi.org/10.29115/SP-2023-0025>

Survey Practice

Vol. 16, Issue 1, 2023

This is a replication of a Survey Practice article that was published in September 2022 that explored survey administration mode and non-substantive response. It revisits the findings from the first study using a randomized controlled design. In this study, we found that on average phone respondents provided more non-substantive responses than web survey respondents, though the portion of phone and web respondents with one or more non-substantive response was similar. Another paper we published in Survey Practice in 2023 also explored mode effects with random assignment to text-to-web and phone groups. Thus, this is the third article by the authors on this topic released in Survey Practice. As part of this paper, we discuss findings from all three articles and contextualize them in relationship to other research on item non-substantive response and survey mode (specifically web and phone).

Introduction

Note: This paper's introduction section is abridged as our previous studies (Soszynski and Bliss 2022 and Soszynski and Bliss 2023) provided background information on mode effects.

In a September 2022 Survey Practice article, we explored mode effects using an English-language survey of residential energy efficiency program participants in a state located in the southern United States via phone and web. We assigned survey administration mode based on the availability of contact information in program participation records. Participants with email addresses were contacted by email, and those without were contacted by phone. Phone respondents had higher rates of item non-substantive response, even while controlling for demographic and background factors. We published a second similar study in August 2023 that compared text-to-web and phone survey administration. Consistent with the first study, we found evidence that phone respondents had more item non-substantive responses for several individual demographic questions and one background question.

This article is a replication study of our September 2022 paper, this time using random assignment, as one concern with our September 2022 study was that survey recipients were not randomly assigned to survey administration mode. It is possible that the method of assigning survey recipients in the September 2022 study could have systematically biased the sample in some unknown way. Although we controlled for several demographic and background variables in

that study's regression analysis, it is possible that we did not control for others that may have affected the results. This paper replicates that study using a randomized control trial design to address this concern.

Additionally, while both of our previous papers investigated item non-substantive response, the exploration was limited. In this paper, we provide a deeper analysis of item non-substantive response by looking at question-specific item non-substantive response rates, as well as multiple dependent variables (one or more non-substantive responses, portion of survey with non-substantive responses).

Our continued exploration of survey mode and item non-substantive or non-response is warranted for numerous reasons. Understanding the potential relationships between survey mode, question type, and likelihood of item non-substantive response is important when considering study design and data analysis. Missing data reduces sample and statistical power. Further, existing research presents divergent findings in this area. For example, a recent meta-analysis found no statistically significant difference in the average item non-response rate in web versus other survey modes (Čehovin, Bosnjak, and Lozar Manfreda 2022). Other studies have found lower rates of item non-response in face-to-face and telephone surveys when compared to web surveys (Lee et al. 2019 and Lesser, Newton, and Yang 2012).

Methods

From August 2021 through January 2022, we conducted an English-language survey of residential energy efficiency program participants in a southern state in the United States. The program offered customers in-home energy audits, no-cost installation of energy-efficient equipment, and rebates for higher-cost improvements. Program participants could broadly be characterized into two groups: 1) customers who received in-home energy audits with direct installation of efficient equipment (LED light bulbs, low-flow showerheads, high-efficiency faucet aerators, and/or advanced power strips); and 2) customers who received cash rebates for more major or significant improvements (ENERGY STAR® windows, ENERGY STAR® doors, or attic insulation).

Before reviewing program data for the presence of valid contact information, we randomly assigned 2,313 program participants to either a phone (1,108) or web group (1,205). Phone number or email address was absent from program records for 460 (20%) of the program participants. Phone numbers were unavailable for 87 (8%) of those assigned to the phone condition, and email addresses were unavailable for 373 (31%) of those assigned to the web group. This left 1,021 program participants in the phone group and 832 in the web group.

Table 1. Sampling and Recruitment Dispositions

Disposition	Web	Phone
Randomly assigned	1,205	1,108
Missing relevant contact information	373	87
Have relevant contact information	832	1,021
Not attempted (budgetary reasons)	0	243
Total attempted	832	778
Unable to connect ¹	67	209
Total contacted	765	569
Disqualified (incorrect contact, nonparticipants)	5	4
Qualified	760	565
Refused	n/a	78
No response (to phone message or email)	597	399
Invalid response ²	0	1
Completed	163	87
Completed as % of contacted and qualified	21%	15%

¹ For the phone group, “unable to connect” included wrong or bad (e.g., disconnected) number or full mailbox. For the web group, this was bounced emails.

² One phone response was deemed invalid and removed.

The phone group call list was randomized before calling. Call staff were instructed to leave one voicemail and call each customer until a terminal disposition was reached (refused, completed, disqualified) or called up to three times. We did not call 243 participants in the phone condition because we had reached the budget for survey completion incentives. Thus, we attempted to reach 778 in the phone condition.

After the removal of contacts we were unable to reach because of technical issues (wrong or disconnected phone numbers, full phone mailboxes, or bounced emails), 569 and 710 participants remained in the phone and web groups, respectively. Nine contacts (5 web, 4 phone) were disqualified from taking the survey as they indicated they were either not the correct contact or stated they had not participated in the program. Of the remaining 565 phone and 760 web contacts, the survey was completed by 87 contacts assigned to the phone group and 163 contacts assigned to the web group (15% and 21% response rates, respectively).

Both groups were offered a \$10 incentive (digital or physical gift card) for completing the survey and an additional \$10 if they uploaded photos of the items installed through the program. Phone and web respondents received instruments designed to be as similar as possible. One difference between the modes was that web respondents were visually presented with either “Prefer not to say” or “Don’t know” response options for each background/demographic question, whereas phone respondents were not explicitly offered non-substantive response options. Phone respondents were given “Prefer not to say” or “Don’t know” as options only after each question had been asked and a respondent indicated a lack of knowledge or willingness to provide a response. Another difference between phone and web administration was that

phone administration staff were not required to read all response options if a phone respondent answered a question before being prompted with response options.

The web respondents and phone administration staff received a prompt to answer the question if they left a question blank (questions were “soft-required”). Four percent of all respondents had one or more seen but unanswered questions (8 of the 250 respondents). The analysis presented below considers these skipped questions as refused or “Prefer not to answer.”

Respondents answered up to 54 questions: 43 program indicator and 11 demographic or background questions. The number of questions varied, depending on the manner in which they interacted with the program and their response patterns. For example, customers who had an in-home energy audit and multiple energy-efficient items installed were asked more questions than customers who purchased a single program-rebated item. The number of questions asked averaged 35 across all respondents.

The survey instrument used was generally consistent with the 2022 study, though two fewer demographic questions were asked in this iteration of the survey (neither income nor employment status questions were asked). Respondents were asked up to eleven background or demographic questions, and we assessed differences between phone and web respondents on seven of these (sex, race or ethnicity, homeownership, age, education home size, and household size). The Results section presents the demographic and background characteristics of the two groups. The primary purpose of this assessment was to determine whether it would be necessary to control for any such group differences in our analysis of how survey administration mode relates to item non-substantive response.

Four other background questions were asked in the survey (home type, home age, and space and water heating type). These four variables were excluded from the demographic bivariate comparison analysis, though they were included in the analysis of non-substantive response.

[Table 2](#) summarizes information about the dependent variables of interest, including number of response categories or levels, and the type of analysis performed to assess differences between phone and web respondents.

We used two approaches to investigate non-substantive responses. The first approach considered item non-substantive response as dichotomous: each respondent that provided one or more non-substantive response was given a value of 1, and each respondent who did not was given a value of 0. For this analysis, we compared the percentage of web respondents that provided *any* non-substantive response with the percentage of phone respondents who provided *any* non-substantive response.

Table 2. Dependent Variables and Type of Analysis²

Variable	Description	Analysis
Binary item nonresponse	At least one “Don’t know” or “Prefer not to answer” response, versus no such responses, to up to 11 demographic/background questions and up to 43 program indicator questions.	<ul style="list-style-type: none"> z-test for difference in mean percentage nonresponse between phone and web respondents. Logistic regression
Non-substantive response percentage	Percent of total responses provided marked “Don’t know” or “Prefer not to answer”	<ul style="list-style-type: none"> z-test for difference in mean percentage nonresponse between phone and web respondents. Linear regression
Sex	Two nominal categories	<ul style="list-style-type: none"> chi-squared tests for categories and z-test for differences of proportions for “don’t know” or “prefer not to answer” responses to test the difference between phone and web respondents in each category
Race or ethnicity	Five nominal categories	
Homeownership	Two nominal categories	
Age	Ordinal, eight levels	<ul style="list-style-type: none"> Mann-Whitney U to test the difference between phone and web respondents in the distribution of responses across all levels. z-test for difference of proportions to test difference between phone and web respondents in “Don’t know” or “Prefer not to answer”
Education	Ordinal, four levels	
Home size (square feet)	Ordinal, six levels	
Household size (number of people)	Ordinal, seven levels	

² All variables included “Don’t know” and/or “Prefer not to answer” options. These are not included in the number of response categories or levels

Table 3. Non-Substantive Response by Survey Mode

Categorization	Web	Phone	Statistic	p-value
At least one non-substantive response	67%	64%	0.52	0.61
Average portion of survey with non-substantive responses ¹	4%	9%	-3.95	<0.01
Minimum portion of survey with non-substantive responses	0%	0%	N/A	
Maximum portion of survey with non-substantive responses	29%	40%	-1.76	0.08

¹The average portion of survey with non-substantive response takes into account the variable number of questions asked to each respondent (depending on their engagement with the program). In other words, the denominator for each respondent depended on the number of questions asked.

The second approach investigated the mean percentage of non-substantive responses for web and phone respondents. For this analysis, we determined the percentage of each respondent’s responses that were non-substantive. This included “Prefer not to answer,” “Don’t know,” and skipped questions. For this study we reviewed a larger number of program indicator questions than the 2022 study, and two fewer demographic questions. We expanded the review of program indicator questions to more fully explore respondents’ tendency to provide non-substantive responses across the entire survey, rather than limiting the review to a subset of questions, as our initial study did.

Results

The phone and web groups had a similar portion of respondents who provided one or more non-substantive responses. Sixty-seven percent of the web responses and 64 percent of phone responses had one or more non-substantive responses. However, on average, phone responses contained a higher portion of responses that were non-substantive. Phone respondents’ surveys had an average of about nine percent of questions marked with non-substantive responses, compared to about four percent of web respondents (see [Table 3](#)).

As this study used randomized assignment, the phone and web populations should not have differed on any demographic or household characteristics variables. However, respondents in the phone group had lower response rates (17 percent phone response rate compared to 23 percent web), which may have created a nonresponse bias.

We thus examined possible differences between the web and phone respondents on those characteristics to determine whether it would be necessary to control for them in the analysis of non-substantive response. [Table 4](#) shows demographic, home, and household characteristics by survey mode.

To control for multiple comparisons, we divided the conventional level of statistical significance ($p < 0.05$) by the number of comparisons made (14) across the seven demographic/home characteristic variables (variables listed in [Table 4](#)), which produced a criterion significant level of $p < 0.0036$. For nominal level variables, we counted each level except “Prefer not to answer” and “Don’t know” and tested the differences with the two-sample z -test for proportions; for ordinal variables, we counted the entire item as one comparison and tested the differences with the Mann-Whitney U .

The groups had similar background and demographic characteristics. None of these variables showed differences (other than the percentages of non-substantive responses) that reached the criterion of $p < 0.0036$. Three items – sex, age, and education – showed differences that would be considered statistically significant under the standard criterion of $p < 0.05$. However, the effect sizes for these items were small. For age and education, the point-biserial correlations were 0.11 and 0.15, respectively. For sex, we used the z -test result to estimate a 0.27 Cohen’s D value. Thus, we deemed it unnecessary to control for them in the analysis of non-substantive response.

Table 4. Demographic Characteristics by Survey Mode

Question	Response	Web (n=163)	Phone (n=87)	Statistic ^a	p-value
What is your sex?	Male	48%	52%	2.64	0.104
	Female	49%	33%		
	Prefer Not to Answer	4%	15%	-3.19	<0.01
What is your age?	18-24 years	1%	0%	-1.73	0.04
	25-34 years	8%	8%		
	35-44 years	10%	14%		
	45-54 years	9%	10%		
	55-64 years	14%	10%		
	65-74 years	34%	18%		
	75-85 years	18%	9%		
	86 years or older	1%	2%		
	Prefer not to say	4%	28%	-5.32	<0.01
How would you identify your race or ethnicity?	Asian	3%	2%	1.46	0.83
	Black/African American	8%	3%		
	Caucasian/White	71%	61%		
	Hispanic or Latino	3%	2%		
	Native American or Alaska Native	5%	3%		
	Other (not specified)	1%	0		
	Prefer not to say	9%	28%	-4.12	<0.01
What's the highest level of education you've completed?	Less than high school	1%	0%	-2.25	0.01
	High School Graduate/ GED	11%	16%		
	Vocational/technical school, associate degrees, or some college	27%	15%		
	Four-year college degree	22%	30%		
	Graduate or professional degree	34%	10%		
	I prefer not to answer	6%	29%	0.88	0.35
Do you own or rent your home?	Own	96%	97%	0.96	1.67
	Rent	3%	1%	-0.76	0.45
	I prefer not to answer	1%	2%		
About how many square feet is your home?	Less than 1,000 square feet	2%	6%	-0.29	0.39
	1,000-1999 square feet	45%	44%		
	2,000-2,999 square feet	40%	29%		
	3,000-3,999 square feet	6%	3%		
	4,000-4,999 square feet	2%	1%		
	5,000 or greater square feet	1%	0%		
	Don't know	4%	17%	-3.68	<0.01
How many people, including you, currently live in your household?	1	15%	17%	0.75	0.23
	2	54%	39%		

Question	Response	Web (n=163)	Phone (n=87)	Statistic ^a	p-value
	3	9%	13%		
	4	9%	8%		
	5	1%	1%		
	6	1%	2%		
	7 or more	0%	2%		
	Prefer not to answer	10%	17%	-1.52	0.13

^a For nominal-level variables, the test statistics are chi-squared tests for categorical variables and two-sample z tests for proportions for “Don’t Know” and “Prefer not answer.” For ordinal-level variables, the test statistic is the z score for Mann-Whitney U.

^b Does not sum to 100% because respondents could select more than one race or ethnic background.

For this paper’s analysis of the relationship between survey mode and overall item non-substantive response, we first followed the strategy used in the September 2022 paper and used a logistic regression to explore the relationship between survey mode on a dichotomous dependent variable for non-substantive response. The logistic regression operationalized the non-substantive response variable as a dichotomous variable: each respondent who provided one or more “Prefer not to answer,” “Don’t know,” or skipped question was given a value of 1 for the non-substantive response variable, and each respondent who did not provide one or more of those responses was given a value of 0.

We found that the overall model was not significant, the pseudo-R-squared values indicate poor overall model fit, and the survey administration independent variable was not statistically significant (see [Table 5](#)). Earlier in the paper we showed that a similar, but smaller portion of phone respondents had at least one non-substantive response compared to web respondents (see [Table 3](#)). When we performed a logistic regression, the results were consistent, though neither analysis yielded statistically significant results. In other words, in this study, we found that survey administration mode did not have a statistically significant relationship with having at least one non-substantive response.

Table 5. Logistic Regression Model of Non-substantive Response and Survey Mode

Independent variable	B	SE	z ratio	Pr(> z)	OR
Model 1 (n=250)					
(Intercept)	0.73	0.17	4.37	<0.01	2.08
Survey Administration Mode (Coded as 1=Phone, 0=Online)	-0.19	0.28	-0.68	0.49	0.83
Model X2	0.46 (df=1)				
-2 Log likelihood	-160.43 (df=2)				
Hosmer and Lemeshow R-squared	<0.01				
Cox and Snell R-squared	<0.01				
Nagelkerke R-squared	<0.01				

Next, we used linear regression to explore if mode was related to the portion of each respondent's survey that was non-substantive. We ran a linear model and found the survey mode variable was significant and the linear model was significant overall, though the R-squared value indicated a somewhat poor model fit. These findings are consistent with our preliminary comparison of average item non-substantive between phone and web respondents (9% compared to 4%).

Table 6. Linear Regression Model of Non-substantive Response and Survey Mode

Independent variable	B	SE	t value	Pr(> z)
Model 3 (n=211)				
(Intercept)	0.043	<0.01	7.79	<0.01
Survey Administration Mode (Coded as 1=Phone, 0=Online)	0.037	<0.01	4.05	<0.01
Residual Standard Error	0.08 (df=249)			
R-squared	0.07			
Adjusted R-squared	0.06			
F statistic	16.42			
p-value	<0.01			

Discussion

This is our third article in Survey Practice examining survey administration mode and non-substantive responses. All three of our studies have found evidence that phone respondents had more item non-substantive response than web respondents, though the level and type of evidence has varied.

Broadly, we believe there are two main takeaways from our studies. Phone surveys tend to have higher rates of item non-substantive response than web surveys, and the way item non-substantive response is investigated matters. Considering the average portion of the survey with non-substantive response and the portion of survey takers with one or more non-substantive response for each mode, the results differed. Operationalization of the dependent variable as a binary may be less useful than looking at the average item non-substantive response for each respondent. Moreover, depending on the needs of the research, it may make more sense to focus on the likelihood of non-response for specific questions.

Our September 2022 study found that a larger portion of phone respondents had one or more non-substantive response compared to web respondents while controlling for demographic and background factors. The published article did not include a comparison of the portion of survey with non-substantive response for the two modes. When we went back and investigated, we found results that were fairly consistent with the present study. Though the difference was not significant, the average item non-substantive response was higher for

Table 7. Non-Substantive Response by Survey Mode for Three Trials

Study	Categorization	Web	Phone	Statistic	p-value
Study 1 (September 2022)	At least one non-substantive response	46%	67%	5.37	<0.01
	Average portion of survey with non-substantive responses	5%	9%	-2.75	<0.01
	Average portion of demographic/background section with non-substantive responses	8%	13%	2.72	<0.01
Study 2 (August 2023)	At least one non-substantive response	67%	70%	-0.53	0.59
	Average portion of survey with non-substantive responses	5%	5%	0.23	0.41
	Average portion of demographic/background section with non-substantive responses	7%	12%	-2.47	<0.01
Study 3 (current)	At least one non-substantive response	67%	63%	0.53	0.63
	Average portion of survey with non-substantive responses	4%	9%	-3.20	<0.01
	Average portion of demographic/background section with non-substantive responses	4%	11%	3.59	<0.01

phone survey respondents compared to email respondents (we found an average item non-substantive response rate of 9% for phone respondents and 5% for email respondents in the 2022 study).

The second study we published in *Survey Practice* in August 2023 compared text-to-web and phone survey administration. We found evidence that phone respondents had more item non-substantive response for several demographic questions and one background question. However, that article did not explore differences between mode and average portion of each survey response that was non-substantive or the portion with one or more non-substantive response.

When we reviewed the data, we found a similar portion of text-to-web and phone respondents had one or more non-substantive responses, and the average portion of each response that was non-substantive was similar. These findings seemed to contradict question-level differences, so we investigated further. We found that, although the average portion of survey with non-substantive response was similar, phone respondents provided more non-substantive responses to demographic and background questions (12% compared to 7%). [Table 7](#) provides a summary of the three trials.

All three of our studies found evidence that phone respondents provide more non-substantive responses than web respondents. Further, phone respondents were more prone to provide non-substantive responses to age (all three studies) and income questions (in both studies it was asked). These findings align with existing research indicating less willingness for respondents to answer sensitive questions during interviewer-led surveys. However, the limitation of a high portion of respondents refusing these questions begs for further investigation, ideally for a study of a population with known background characteristics. We searched for literature on the relationship between demographics and

propensity for item non-substantive response and found limited results, though we found literature indicating higher income individuals are more prone to refuse income questions (Yan, Curtin, and Jans 2010).

Reviewers might suggest a limitation of these studies is that differences between groups could be related to explicitly offering “Don’t know” or “Prefer not to say” as options. Web respondents seeing “Prefer not to say” or “Don’t know” on their screen and phone respondents not explicitly being offered non-substantive response options are different experiences. However, the finding that phone respondents tended to have a higher non-substantive response rate suggests that though web respondents were prompted visually for each question with an opportunity to provide a non-substantive response, interviewer effects may have persisted.

A lack of benchmarking data or underlying information regarding the population is a weakness for all three of our studies. In each of our studies, both modes could potentially underrepresent or overrepresent certain groups, leading to similar but somewhat biased samples. Additionally, though sample sizes for the studies were reasonably sized and large enough to meet 90–10 precision for the populations of interest (program participants), the generalizability of our studies may be limited as they were conducted on unique populations. The surveys examined in these studies were conducted on participants in energy efficiency programs in the southern United States, and one of the studies was limited to income-qualified households.

The primary finding of this paper is that phone respondents tend to provide more non-substantive responses than web respondents. Our studies’ limitations suggest future research is warranted and should seek to explore survey mode and item non-substantive response with larger sample sizes and a survey of a population that might provide more generalizable results. Researchers may also consider independent variables such as income and employment status or para data, such as time of day, day of the week, or number of contact attempts. Beyond exploring factors that could explain item non-response, it is important for researchers to continue to investigate methods for reducing item non-response (e.g., question wording, survey design).

Incomplete survey data is an ongoing concern. Questions with rates of high non-substantive response may hinder researchers’ ability to draw conclusions. For example, high demographic question refusal rates may make assessing the representativeness of a sample challenging. This issue warrants continued attention as technology use patterns and data collection methods shift.

.....

Lead Contact

Michael Soszynski
Senior Evaluation Researcher, ADM Associates
Michael Soszynski
mikesoszynski@gmail.com
609 864 1096

Submitted: March 24, 2023 EST, Accepted: October 24, 2023 EST

REFERENCES

- Čehovin, Gregor, Michael Bosnjak, and Katja Lozar Manfreda. 2022. "Item Nonresponse in Web Versus Other Survey Modes: A Systematic Review and Meta-Analysis." *Social Science Computer Review* 41 (3): 926–45. <https://doi.org/10.1177/08944393211056229>.
- Lee, Hana, Sunwoong Kim, Mick P. Couper, and Youngje Woo. 2019. "Experimental Comparison of PC Web, Smartphone Web, and Telephone Surveys in the New Technology Era." *Social Science Computer Review* 37 (2): 234–47. <https://doi.org/10.1177/0894439318756867>.
- Lesser, Virginia M., Lydia A. Newton, and Daniel Yang. 2012. "Comparing Item Nonresponse Across Different Delivery Modes in General Population Surveys." *Survey Practice* 5 (2): 1–4. <https://doi.org/10.29115/sp-2012-0009>.
- Soszynski, Michael, and Ryan Bliss. 2022. "Does Survey Administration Mode Relate to Non-Substantive Responses? A Comparison of Email versus Phone Administration of a Residential Utility-Sponsored Energy Efficiency Program Survey." *Survey Practice* 15 (1). <https://doi.org/10.29115/sp-2022-0009>.
- . 2023. "Demographic and Measurement Differences between Text-to-Web and Phone Survey Respondents: A Randomized Control Trial Using a Survey of Home Weatherization Program Participants." *Survey Practice* 16 (1): 1–11. <https://doi.org/10.29115/sp-2023-0012>.
- Yan, Ting, Richard Curtin, and Matthew Jans. 2010. "Trends in Income Nonresponse Over Two Decades." *Journal of Official Statistics* 26 (1): 145–64. https://www.researchgate.net/profile/Ting_Yan3/publication/272887076_Trends_in_Income_Nonresponse_Over_Two_Decades/links/555f779208ae86c06b636e34.pdf.