

ARTICLES

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

Michael Soszynski¹ , Ryan Bliss²

¹ Senior Evaluation Researcher, ADM Associates, ² ADM Associates

Keywords: mode effects, item nonresponse, program evaluation, non-substantive response

<https://doi.org/10.29115/SP-2022-0009>

Survey Practice

Vol. 15, Issue 1, 2022

Program evaluation research often relies heavily on surveys and increasingly self-administered online tools. Often, however, a portion of program participants or customers do not have or do not provide email addresses. This study investigates if the type of contact information that members of a population provide relates to non-substantive responses (choosing "don't know" or "prefer not to answer"). We found evidence that survey administration mode relates to non-substantive responses, even while controlling for demographic factors.

Introduction

Program evaluators and survey researchers often do not have the budget to conduct surveys using multiple modes. This limitation, choosing phone or web administration, may result in mode effects that could skew results, reduce generalizability, affect data quality, and lead to misleading conclusions.

Mode effects are the differences in responses to a survey or survey question attributable to the survey administration mode. Past research has found that survey mode can affect the demographics of survey respondents, specifically income and education (Datta, Walsh, and Terrell 2002). Further, a recent meta-analysis of survey characteristics and representativeness found that mixed-mode surveys and other-than-Web modes tend to be more representative (Cornesse and Bosnjak 2018).

Mode effects are composed of two components: differential non-observation and observation effects (de Leeuw 2018). Non-observation effects refer to who responds to different modes and observation effects refer to how respondents respond to questions presented via a given mode.

Non-observation effects may result in systematic underrepresentation when different types of respondents choose or are only able to complete the survey in a specific mode. As such, non-observation effects may result in a form of self-selection bias that can affect the sample frame if groups of individuals are

^a Corresponding author:
Michael Soszynski
mikesoszynski@gmail.com
609 864 1096

unable or unwilling to provide one or more forms of contact information. For example, some members of the population of program participants may be unwilling to provide email contact information or telephone contact information. A survey performed using a single mode of contact in these cases may result in biased results if those willing to provide a form of contact information differ systematically from those who are not willing to provide that information.

In contrast, observation effects occur when the mode of administration affects the types of responses provided. For example, responses gathered orally may provide different information than would be collected in a written or online survey because the interviewer may provide clarifying information to the respondent or probe the respondent to clarify their response. In addition, some individuals may be more hesitant to share private information or to disclose information that they fear may result in being stigmatized when speaking with another person. For example, mode effects were demonstrated in a comparison of Web and telephone responses that found that Web respondents gave less differentiated answers to sets of questions with the same response options (Fricker et al. 2005). The same study also found a mode effect for survey response time. Specifically, Web respondents took longer to complete the survey than telephone respondents, although the authors noted the slower completion times mainly related to entering open-ended responses.

Survey administration mode can also affect patterns of providing non-substantive responses (i.e., refusing to answer or selecting “don’t know”) or item nonresponse (skipping questions) to certain questions. For example, one study found that item nonresponse differed significantly across modes with telephone respondents having significantly lower item nonresponse than Web and mail respondents (Lesser, Newton, and Yang 2012). Another study found that an interviewer led survey had lower item nonresponse than PC or Web-based survey designs (Lee et al. 2019). A more recent meta-analysis of item nonresponse in Web versus other modes found no statistically significant difference in the average item nonresponse rate in Web versus other survey modes (Čehovin, Bosnjak, and Lozar Manfreda 2022).

The study presented here aims to add to the existing literature on survey administration mode effects by exploring a real-world program evaluation issue, namely if the type of contact information that members of the population provide relates to non-substantive responses (choosing “don’t know or “prefer not to answer”).

We examine this issue by comparing responses given by respondents who provided email contact information to those who did not provide email contact information but who did provide a telephone number. In this analysis, we compare the portion of respondents in each group that provided non-substantive responses while controlling for demographic factors.

Hello!

I am contacting you on behalf of [SPONSOR ORGANIZATION], which needs feedback from individuals like you who participated in the [PROGRAM NAME]. You are one of a select group of people we are contacting to help us learn about participants' experience with the program. To ensure that our findings are accurate, we need responses from as many people as possible. Can you be one of the people we hear from?

After you complete the survey, we will send you a link to a \$5 gift card valid at your choice from 80+ retailers or selected non-profits as our way of thanking you for answering our questions.

[Click Here to Start the Survey](#)

And use this access code: [CODE]

We would greatly appreciate your willingness to answer a few questions about your experience with this program. Your responses will be kept anonymous and will be used to improve [REDACTED] programs in the future.

If you have any questions, comments or require technical assistance feel free to contact me at [REDACTED] or [REDACTED]. If you do not wish to assist in this study, please click the "unsubscribe" button below.

Thank you so much!

Figure 1. Survey email message

Methods

In August, November, and December 2020, we conducted an English-language survey of residential energy efficiency program participants in the Southern United States via phone and Web. Program participants were either sent an email to complete a Web-based survey or contacted via telephone to complete the survey by telephone. The mode of administration was based on the available contact information: participants with email addresses were contacted by email and those without were contacted by telephone. In both modes, the contact was provided an overview of the study and asked if they would be willing to provide their feedback.

The email survey was launched in two batches for quarter one and quarter two participants. We emailed an invitation and one reminder to 1,619 participants who had valid email addresses. This represented all participants with a valid email address. Quarter two email invitations were sent on August 3rd and 5th, with reminders on August 7th. Quarter three email invitations were sent November 9th, with reminders on November 12th.

When administering the survey by telephone, we called 912 of the 1,102 participants without email addresses in two waves; the first wave was between August 9 and September 1, and the second wave was between November 11 and December 1. A total of 190 participants were not invited to take the survey because of budget and time constraints. These contacts only had telephone numbers in the tracking data. The list had been randomized before calling and did not systematically exclude certain customers in any manner. We left one voicemail and made one follow-up call to potential phone respondents. The data did not contain information regarding respondent device type for email responses nor did it contain information regarding phone service type (cell or landline).

We received 337 web and 165 telephone responses, for a 21% web and 22% telephone response rate, respectively (see [Table 1](#))

Table 1. Survey administration information

Survey delivery	Total
Web	
Initial email contact list	1,826
Invalid email addresses	29
Bounced email	105
Undeliverable email	73
Invalid email (%)	11%
Email invitations sent (unique valid)	1,619
Email completions	337
Email response rate (%)	21%
Phone	
Initial phone list	912
Disconnected/wrong number	158
Invalid phone (%)	17%
Phone calls (unique valid)	754
Phone completions	165
Phone response rate (%)	22%
Overall	
Total invites (unique)	2,373
Total completions	502
Overall Response rate (%)	21%

Phone and Web respondents received instruments designed to be as similar as possible. For all questions analyzed in this study, respondents were given “don’t know,” “prefer not to say,” or both response options. The most notable difference between phone and Web administration was that if a phone respondent answered a question before being prompted with response options, phone administration staff were not required to read all response options. The Web respondents received a prompt to answer the question if they left a question blank. (Questions were “soft-required.”)

The analysis for this study focused on 9 of the 13 demographic/home characteristics questions and up to 12 program indicator questions for non-substantive response. We excluded four home characteristics variables—home type (single-family detached, single-family attached, etc.); age of home; main space heating fuel; and main water heating fuel—as being unlikely to produce results that would be easily interpretable.

We operationalized the non-substantive response variable as a dichotomous variable: each respondent who provided one or more “don’t know” or “prefer not to answer” responses to a demographic or program indicator 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. In total, 12 program indicator questions included up to nine Likert scale questions and up to three yes/no questions. [Table 2](#) summarizes

Table 2. Survey variables

Variable	Description*	Analysis
Independent Variables (Demographic/Home Characteristic)		
Nominal Level		
Sex	Two categories	z-test for difference of proportions to test difference between phone and Web respondents in each category, including “don’t know” or “prefer not to answer.”
Race or identity	Five categories	
Employment status	Seven categories	
Homeownership	Two categories	
Home type	Eight categories	Not included in analyses
Home heating type	Three categories	
Water heating type	Three categories	
Ordinal Level		
Age	Ordinal, eight levels	Mann-Whitney <i>U</i> 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.”
Household income	Ordinal, ten levels	
Education	Ordinal, four levels	
Home size	Ordinal, six levels	
Household size	Ordinal, seven levels	
Home age	Ordinal, eight levels	Not included in analyses
Dependent Variable (Item Nonresponse)		
Item nonresponse level	Dichotomous, respondents with at least one “don’t know” or “prefer not to answer” response to up to 13 demographic questions and up to 12 program indicator questions.	z-test for difference in mean percentage nonresponse between phone and Web respondents.

*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

additional information about each of these dependent variables, including number of response categories or levels, and the type of bivariate analysis performed to assess differences between phone and Web respondents.

We used logistic regression analyses to assess the relationship of item nonresponse to survey mode. Demographic factors were included in the regression analysis to explore possible correlations with item nonresponse and to understand how mode may affect item nonresponse, while controlling for various background characteristics.

In addition to the outcomes previously discussed, we investigated survey response time. The survey collection tool collected start and submission date time. However, there are limitations to interpreting times. For both email and phone responses, it appears responses were not submitted promptly at survey completion (i.e., the survey-taker or survey administration staff may have failed to hit submit immediately after they had completed the survey). To address this, we removed 90 respondents’ response times that were determined to be outliers. A total of 63 Web (19%) and 27 phone (16%) response times were removed. After this preliminary step, we found a median survey response time of 17 minutes 35 seconds and an average response time of 24 minutes 16 seconds. The average and median telephone survey response times were about

7 and 3 minutes longer than the Web surveys, respectively. The questions used in this study were located in the latter half of the survey, with the demographic questions being the final section.

Results

[Table 3](#) and [Table 4](#) show demographic and home and household characteristics, respectively, by survey mode. To control for multiple comparisons, we divided the conventional level of statistical significance ($p < .05$) by the number of comparisons made (19) across the nine demographic/home characteristic variables (variables listed in [Table 3](#) and [Table 4](#), excluding home type, home age, space heating fuel type, and water heating fuel type) which produced a criterion significant level of $p < .0026$. For nominal level variables, we counted each level except prefer not to answer, and for ordinal variables, we counted the entire item as one comparison.

Among the demographic variables, age, education, household size, home ownership, and two of the employment status variables had statistically significant relationships to survey mode.

The phone survey gathered more responses from participants aged 65 or older, while the Web mode had a larger portion identifying as 44 years old or younger (36% vs. 22%). The phone survey gathered a larger portion of responses from people ages 65 or older (45% vs. 26%).

Web survey respondents noted having completed more education than phone respondents: 58% of Web respondents said they had a four-year degree or higher compared to 41% of phone respondents.

A higher percentage of Web than phone survey respondents (47% vs. 30%) reported working at least 30 hours per week, while phone survey respondents were more likely than Web respondents to report being retired (48% vs. 28%). A larger portion of Web respondents identified as male (51% vs. 41%), but this difference did not reach the above criterion for statistical significance.

We did not find a statistically significant difference between Web and phone respondents' reported household income, though a significant portion of respondents to both modes preferred not to share this information (see [Table 4](#)).

Regarding race, the portion of respondents that identified as white or Caucasian was larger in the Web survey; however, the magnitude of the difference was small. We did not find a statistically significant difference between the portion of respondents who identified as non-white.

Web respondents tended to own their homes at a higher rate than phone respondents, while phone respondents were less aware of their home size (see [Table 4](#)). Phone respondents reported living alone more than Web respondents.

Table 3. Demographic characteristics by survey mode

Question	Response	Web (n=337)	Phone (n=165)	Statistic ^b	p-value
What is your sex?	Male	51% (171)	41% (68)	2.082	0.039
	Female	42% (143)	50% (82)	-1.537	0.124
	Prefer Not to Answer	7% (23)	9% (15)	-0.479	0.632
What is your age?	18–24 years	1% (2)	1% (1)	-5.277	<0.0001
	25–34 years	14% (46)	10% (16)		
	35–44 years	22% (73)	12% (19)		
	45–54 years	16% (53)	7% (12)		
	55–64 years	17% (57)	11% (18)		
	65–74 years	21% (70)	16% (27)		
	75–85 years	5% (18)	25% (42)		
	86 years or older	0% (1)	3% (5)		
	Prefer not to say	5% (17)	15% (25)	-3.842	<0.0001
How would you identify your race or ethnicity? ^a	Asian	3% (10)	5% (7)	-0.782	0.435
	Black/African American	6% (20)	9% (15)	-1.375	0.1697
	Caucasian/White	76% (256)	68% (110)	1.9627	0.050
	Hispanic or Latino	4% (12)	4% (6)	-0.069	0.945
	Native American or Alaska Native	7% (25)	4% (6)	1.666	0.096
	Prefer not to say	9% (32)	13% (21)	-1.181	0.238
Which of the following categories best describes your employment status?	Working up to 30 hours per week	11% (36)	8% (13)	0.994	0.321
	Working 30 or more hours per week	47% (160)	30% (50)	3.664	<0.001
	Not employed, looking for work	3% (9)	1% (1)	1.555	0.121
	Not employed, Not retired or disabled	2% (8)	1% (2)	0.875	0.382
	Retired	28% (93)	48% (79)	-4.498	<0.001
	Disabled, not able to work	1% (5)	1% (1)	0.849	0.396
	Prefer not to answer	8% (26)	11% (19)	-1.400	0.162
What's the highest level of education you've completed?	High School Graduate/GED	10% (34)	20% (32)	3.255	0.001
	Vocational, technical or some college	24% (81)	27% (45)		
	Four-year college degree	30% (100)	21% (35)		
	Graduate or professional degree	28% (94)	20% (33)		
	I prefer not to answer	8% (28)	12% (20)	-1.365	0.173

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

^b For nominal-level variables, the test statistic is the two-sample z test for proportions. For ordinal-level variables, the test statistic is the z score for Mann-Whitney U.

We used regression analysis to examine whether response differences between phone and Web respondents reflect the demographic differences rather than mode effects. We regressed the non-substantive response variable on demographic characteristics and mode of administration using a series of hierarchical nested logistic regression models.

We included all demographic and home characteristics variables that had statistically significant bivariate relationships with survey mode in the regression analyses.

Table 4. Home and household characteristics by survey mode

Question	Response	Web (n=337)	Phone (n=165)	Statistic ^a	p-value
Do you own or rent your home?	Own	96% (322)	88% (145)	3.169	0.002
	Rent	3% (11)	5% (9)	-1.178	0.239
	I prefer not to answer	1% (4)	7% (11)	-3.387	0.001
About how many square feet is your home?	Less than 1,000 square feet	6% (19)	5% (8)	1.4280	0.1530
	1,000–1999 square feet	52% (176)	52% (85)		
	2,000–2,999 square feet	29% (98)	24% (40)		
	3,000–3,999 square feet	7% (23)	5% (8)		
	4,000–4,999 square feet	3% (11)	1% (8)		
	5,000 or greater square feet	1% (3)	0% (0)		
	Don't know	2% (7)	14% (23)	-5.266	<0.001
How many people, including you, currently live in your household?	1	16% (55)	30% (49)	3.302	0.002
	2	34% (116)	38% (62)		
	3	18% (60)	7% (12)		
	4	10% (33)	10% (17)		
	5	7% (22)	4% (7)		
	6	2% (7)	0% (0)		
	7 or more	0% (1)	2% (4)		
	Prefer not to answer	12% (42)	8% (14)	1.329	0.184
What is your approximate household income?	Less than \$10,000	1% (2)	0% (0)	0.967	0.332
	\$10,000 to \$19,999	4% (12)	4% (7)		
	\$20,000 to \$29,999	4% (14)	2% (4)		
	\$30,000 to \$39,999	9% (31)	7% (11)		
	\$40,000 to \$49,999	7% (23)	5% (8)		
	\$50,000 to \$74,999	15% (50)	15% (24)		
	\$75,000 to \$99,999	14% (48)	8% (13)		
	\$100,000 to \$149,999	13% (45)	12% (20)		
	\$150,000 to \$199,999	5% (17)	2% (3)		
	\$200,000 or more	3% (9)	0% (0)		
	I do not recall/Prefer not to say	26% (86)	45% (75)	-4.495	< 0.001

^a For nominal-level variables, the test statistic is the two-sample z test for proportions. For ordinal-level variables, the test statistic is the z score for Mann-Whitney U.

For these models, we dummy coded the nominal variables (i.e., education, sex, and employment). Age and household size were input into the model as ordinal variables.

In Model 1, we regressed the non-substantive response variable on the variables that had significant differences between Web and phone respondents (employment status, age, education, and household size). We found the overall model was statistically significant; however, only employment significantly predicted question non-response.

Table 5. Logistic regression models of non-substantive responses, demographic independent variables.

Independent variable	B	SE	z ratio	Pr(> z)	OR
Model 1 (n=431)					
(Intercept)	0.58117	0.68627	0.847	0.39708	1.788
Employed	-0.71725	0.22656	-3.166	0.00155	0.488
College	-0.19126	0.20508	-0.933	0.35101	0.826
Age	0.07339	0.07034	1.043	0.29677	1.076
Household size	-0.09231	0.08321	-1.109	0.26731	0.912
Model X2(1)	27.73631				
-2 Log likelihood	-282.9252 (df=5)				
Hosmer and Lemeshow R-squared	0.047				
Cox and Snell R-squared	0.062				
Nagelkerke R-squared	0.083				
Model 2 (n=431)					
(Intercept)	0.416	0.694	0.600	0.549	0.549
Survey Administration Mode*	0.704	0.223	3.161	0.002	0.002
Employed	-0.715	0.229	-3.118	0.002	0.002
College	-0.137	0.208	-0.656	0.512	0.512
Age	0.033	0.072	0.451	0.652	0.652
Household Size	-0.091	0.084	-1.091	0.275	0.275
Own	-0.348	0.503	-0.691	0.490	0.490
Model X2(1)	37.8015				
-2 Log likelihood	-277.8926 (df=6)				
Hosmer and Lemeshow R-squared	0.064				
Cox and Snell R-squared	0.084				
Nagelkerke R-squared	0.112				

* Coded as 1=Phone, 0=Online.

Model 2 regressed non-substantive response on demographic factors and a dummy variable for the administration mode. We found that the mode variable was statistically significant ($p < .002$) in this model and therefore remains a predictor of item nonresponse, even while controlling for demographic factors.

[Table 5](#) summarizes the model statistics. Both models were statistically significant overall. Pseudo R^2 values indicate the model with the administration mode had a better overall fit than the model with only demographic independent variables.

Discussion

This article explored mode effects on respondent characteristics, nonresponse/self-selection, and item nonresponse. We found that survey mode related to non-substantive responding. We found higher non-substantive response rates for phone respondents, while controlling for demographic and background factors.

This study's limitations include potential unmeasured differences between the two samples, sample size, and the formulation of non-substantive response (combining don't know and refused responses). Though the survey gathered demographic information and we sought to control for these factors through logistic regression, reviewers may identify additional factors that may influence the likelihood of item nonresponse that we did not investigate (for example time of day, quality of phone connection). Similarly, a larger sample size could potentially increase the strength of the analysis and future analysis may be more focused on parsing out the difference between non-substantive responses, whether they are don't know or prefer not to say.

Our findings contribute to a growing body of research on item nonresponse. Čehovin et al.'s meta-analysis found item nonresponse rates in web surveys were similar compared to other modes. Both Lesser and Lee et al. found that item nonresponse differed significantly across modes with phone respondents having significantly lower item nonresponse than web and mail respondents. Some research has suggested that interactions between survey mode, demographics, and questionnaire characteristics contribute to variations in item nonresponse (Messer, Edwards, and Dillman 2012). Our study found that phone respondents had higher item non-substantive response, even while controlling for respondent characteristics.

Strategies for reducing item nonresponse may take the form of forcing a response, prompting a response if someone attempts to skip an item, or offering "don't know" or "no response" options. The latter may be done in combination with forced or prompted responses. None of these is ideal. Forcing responses can increase drop-out rates and offering "no response" options can increase item nonresponse (Kmetty and Stefkovics 2021). Also, offering "don't know" responses in online surveys can result in a higher rate of this response than in a telephone survey that allows but does not explicitly offer it (Zeglovits and Schwarzer 2014). Our study's findings and other past research illuminates an area in need of continued research—if and how survey administration mode may relate to non-substantive responses and the most appropriate methods for addressing this data quality issue.

Submitted: July 29, 2021 EST, Accepted: August 11, 2022 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*. <https://doi.org/10.1177/08944393211056229>.
- Cornesse, Carina, and Michael Bosnjak. 2018. "Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis." *Survey Research Methods* 12 (1): 1–13. <https://doi.org/10.18148/SRM/2018.V12I1.7205>.
- Datta, Pratim, Kenneth R. Walsh, and Dek Terrell. 2002. "The Impact of Demographics on Choice of Survey Modes: Demographic Distinctiveness between Web-Based and Telephone-Based Survey Respondents." *Communications of the Association for Information Systems* 9. <https://doi.org/10.17705/1cais.00913>.
- de Leeuw, Edith D. 2018. "Mixed-Mode: Past, Present, and Future." *Survey Research Methods* 12 (2): 75–89. <https://doi.org/10.18148/SRM/2018.V12I2.7402>.
- Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. 2005. "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69 (3): 370–92. <https://doi.org/10.1093/poq/nfi027>.
- Kmetty, Zoltán, and Ádám Stefkovics. 2021. "Assessing the Effect of Questionnaire Design on Unit and Item-Nonresponse: Evidence from an Online Experiment." *International Journal of Social Research Methodology*, August, 1–14. <https://doi.org/10.1080/13645579.2021.1929714>.
- 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>.
- Messer, Benjamin L., Michelle L. Edwards, and Don A. Dillman. 2012. "Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Based Mixed-Mode Surveys of the General Public." *Survey Practice* 5 (2): 1–9. <https://doi.org/10.29115/sp-2012-0012>.
- Zeglovits, Eva, and Steve Schwarzer. 2014. "Presentation Matters: How Mode Effects in Item Non-Response Depend on the Presentation of Response Options." *International Journal Of Social Research Methodology* 19 (2): 191–203. <https://doi.org/10.1080/13645579.2014.978560>.