**Does a Split Questionnaire Design to Reduce Survey Length Reduce Nonresponse and Nonresponse Bias?**

Andy Peytchev and Emilia Peytcheva

RTI International

Acknowledgment: This material is based upon work supported by the National Science Foundation under Grant No.1259985 to Andy Peytchev and Emilia Peytcheva.

**Abstract**

Results on the relationship between survey length and nonresponse are quite mixed, yet the general expectation is that longer surveys lead to greater nonresponse and increased potential for nonresponse bias. Split questionnaire design is an approach to reduce the survey length, while collecting data for all questions included in a longer survey. The survey is divided into modules and each respondent is assigned to a subset of modules. Different subsets are used, so that all possible combinations of survey questions are observed. The missing data for the omitted modules for each respondent are filled in using multiple imputation, to achieve a complete rectangular dataset, while providing a means to propagate the uncertainty in the imputed values into the variance estimates.

We conducted an experiment with a full-factorial design, manipulating both the length of the survey and the order of the survey modules. A national sample of addresses was selected and sample members were mailed invitations to a Web survey. The survey questions were obtained from large national in-person surveys that provide benchmark estimates to estimate nonresponse bias in the survey. Data collection ended earlier this year. In this paper we compare response rates and nonresponse bias as a function of survey length. This is the first phase of evaluation, which will be followed by comparison of measurement error differences and an evaluation of multiple imputation to recover omitted information. The findings can help inform future survey designs that aim to balance data needs with respondent burden and different sources of survey error.

**Introduction**

**Increasing nonresponse rates allow greater potential for nonresponse bias in survey estimates, as the latter is the product of the nonresponse rate and the difference between respondents and nonrespondents. The increasing reluctance to participate in surveys may also translate into greater measurement error once the sample member agrees to participate. Despite the increased reluctance to survey participation, interviews are not becoming shorter to reduce respondent burden—in fact, many surveys have become longer due to the constant need of additional data. However, as described in the following subsections, survey length is likely a major contributor to nonresponse and measurement error.**

**Nonresponse and Nonresponse Bias. Unit nonresponse has been increasing in surveys, which is very evident in ongoing surveys that have maintained the same general design over the years. In the NHIS**, the household nonresponse rate increased on average by 0.68 percentage points per year between 1990 and 2009. **During the same period, the nonresponse rate increased by a similar 0.63 percentage points per year in another major national survey, the National Survey on Drug Use and Health (NSDUH) (see Peytchev forthcoming). That is, despite increasing efforts and resources used to maximize participation, such as the introduction of incentives in the NSDUH during that time period, response rates have declined by about 2 percentage points every 3 years.**

Nonresponse can profoundly bias survey estimates. For example, one evaluation found that nonrespondents to one component of the National Health and Nutrition Examination Survey III were 59% more likely to report being in poor or fair health than respondents to this component of the survey (Khare et al. 1994). Similarly, the Belgian National Health Interview Survey, with a 38.6% nonresponse rate, obtained a 19% lower estimate for reporting poor health compared to the Belgian census with a 3.5% nonresponse rate (Lorant et al. 2007).

**Measurement Error.** Measurement error, defined as obtaining an inaccurate response from the survey respondent, can also have a considerable biasing effect on survey estimates. Arguably, it is also more difficult to quantify than nonresponse rates, and even bias, because it requires the same information to be collected from another source or a different procedure. The difficulty in estimating measurement error is unfortunate as it likely leads to insufficient attention being paid to this major source of survey error. In fact, some methodological studies investigating both nonresponse and measurement error have found that measurement error can be the dominant source of error for some estimates (Biemer 2001, Olson 2006, Peytchev 2011, Groves and Magilavy 1984). Furthermore, nonresponse and measurement error can be causally related (e.g., Peytchev, Peytcheva, and Groves 2010), in which case addressing one without considering the other can have unforeseen consequences. Therefore, to improve survey results, it is imperative to appraise and address both nonresponse and measurement error.

**Reducing Survey Length to Reduce Nonresponse and Measurement Error. Survey length has been found to affect both nonresponse and measurement error. This link may not seem surprising. The length of the survey can be seen as the size of the survey request made to the respondent, and once the survey is started, fatigue and lack of motivation may settle in. Longer surveys elicit greater nonresponse in both self-administered surveys (Heberlein and Baumgartner 1978, Galesic and Bosnjak 2009) and interviewer-administered surveys (Groves et al. 1999). To the extent that those who are less likely to participate in a longer survey are different on measures of interest compared to those who are more likely to participate, a longer survey will lead to nonresponse bias in these estimates.**

**Placing questions toward the end of a self-administered survey (thus achieving the measurement properties of a long interview) has been linked to shorter answers (Galesic and Bosnjak 2009), faster responding and less variability across questions (Galesic and Bosnjak 2009, Peytchev 2007), and extreme straight-lining across items (Herzog and Bachman 1981). Most importantly, when using a substantive model regressing** body mass index **(BMI) on diet and exercise questions, placing the questions later in a self-administered survey has been found to elicit greater measurement error, diminishing the ability of the measures to predict BMI (Peytchev and Peytcheva 2017). Despite the ability of interviewers to engage respondents, this effect of survey length will likely persist for interviewer-administered surveys. For example, respondents have been found to be more likely to answer negatively to questions as the interview progresses in order to avoid additional follow-up questions on that topic (e.g., Biemer 2000). Thus, reducing the length of the survey can be expected to reduce nonresponse rates, nonresponse bias, and measurement error, yet dropping questions altogether is often unfeasible.**

**Split Questionnaire Design. The fundamental premise of split questionnaire design is to reduce respondent burden and improve survey estimates through amalgamation of study design and statistical methods, without altering the study objectives. By omitting subsets of questions for different respondents, the survey can be shortened for all respondents. By imputing the omitted questions, a full analytic dataset is created to meet the original study objectives. Repeating the imputation multiple times allows estimation of variances that reflect the degree of uncertainty in the imputed values. The approach has been shown to yield valid inferences through simulation studies involving actual and model-based generated datasets. However, the design needs empirical research to address three related issues: (1) an evaluation of nonresponse and measurement error resulting from survey length, (2) the ability to reduce error form each source through split questionnaire design, and (3) a demonstration of how it is applied using available statistical tools to help its widespread use.**

**As an approach, split questionnaire design is widely applicable and has the potential to become a ubiquitous methodology, improving survey estimates and even reducing survey costs linked to survey length. It can be used for three beneficial outcomes: to reduce error in survey-based results, to increase the amount of information that can be collected from respondents to inform multivariate analyses, and to reduce the cost of data collection by requiring the collection of less information.**

**The concepts behind split questionnaire design are not new. The first use of questionnaire splits dates back to a survey conducted in the late 1930s (Hill 1951), which simply involved the administration of different sets of survey questions to reduce the survey length and still obtain the desired descriptive statistics. A decade later, Herzog and Bachman (1981) advised long questionnaires be split in different parts that are administered in different order. These researchers, however, considered the splits as independent surveys, without incorporating them into a unified analysis. Moreover, little attention has been focused on split questionnaire designs until Raghunathan and Grizzle (1995) demonstrated the advantages of split questionnaire designs in two simulation studies. The method is based on multiple matrix sampling design (Shoemaker 1973, Munger and Lloyd 1988), where a random sample of items is presented to each sampled individual. In contrast, split questionnaire design imposes restrictions on which items are administered to which individuals and shares some similarities with educational test designs (Holland and Thayer 1985, Holland and Wightman 1982). Despite the promise of the method, no experimental research has been conducted to further examine the feasibility of implementing this design in large national surveys, the nonresponse and measurement error properties of such designs, or comparison of inferences obtained by administering the full questionnaire vs. split questionnaires.**

**Research Questions**

**This study experimentally addresses four research questions:**

1. **Does shortening the survey reduce nonresponse in a self-administered survey?**
2. **Is nonresponse bias altered as a function of survey length?**
3. **Is measurement error altered as a function of survey length?**

**These are the initial set of research questions in this study. Our next step is to evaluate the reduction of nonresponse and measurement bias and impact on mean square error from using split questionnaire design, after multiply imputing the full data for all respondents.**

**Experimental Design and Data Collection**

A shortcoming of the extant literature is the confounding of nonresponse and measurement error in designs manipulating survey length. While easily explained by the motivation to increase response rates by reducing survey length, it limits our ability to understand a potentially more important consideration for reducing survey length. We used a 2 by 2 factorial design, manipulating survey length (15- vs. 30-minute survey) and the order of modules (dividing the survey into 5 topical modules and switching places for modules 2 and 4). Survey questions were taken from major U.S. national surveys: the American Community Survey (ACS) and its topical supplements, the National Health Interview Survey (NHIS), the General Social Survey (GSS), and the American National Election Studies (ANES). The approach to the construction of the survey instrument had the dual objective of providing results of relevance to important data collection systems and providing benchmark estimates to gauge errors in survey estimates produced by the experimental manipulation.

The experimental design is shown in Figure 1. There are four randomly assigned experimental conditions. Groups 1 and 2 use the full 30-minute instrument, but switch the order of modules B and D. Groups 3 and 4 use the reduced 15-minute survey, including only module B or D, and also dropping module C. A total of 1,8880 surveys were competed.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ****Exp.****  ****Group**** | ****Survey Instrument**** | | | | | ****Length**** |
| **1** | **Module A:**  **Core health questions** | **Module B** | **Module C** | **Module D** | **Module E: Core demog. & phys. meas.** | **30 minutes** |
| **2** | **Module A:**  **Core health questions** | **Module D** | **Module C** | **Module B** | **Module E: Core demog. & phys. meas.** | **30 minutes** |
| **3** | **Module A:**  **Core health questions** | **Module B** | **Omitted** | **Omitted** | **Module E: Core demog. & phys. meas.** | **15 minutes** |
| **4** | **Module A:**  **Core health questions** | **Module D** | **Omitted** | **Omitted** | **Module E: Core demog. & phys. meas.** | **15 minutes** |

Figure 1. Experimental design, manipulating survey length and order of survey modules.

A nationally-representative sample of addresses in the U.S. was selected and sample members were sent invitation letters, including a unique link and password to log into a Web survey instrument. Sample members were promised $5 upon completion of the survey. Half of the addresses received invitations to the 15-minute survey and the other half to the 30-minute survey. A second phase of data collection was planned to follow-up on a subsample of nonrespondents, increasing the incentive to $20. However, the response rate to the first phase fell far short of expectations (4.5%) and the second phase would not have been able to provide the necessary number of interviews for the measurement error analysis, the target for which was 2,000 completed surveys. Instead, the data collection design was modified to a single-phase and inclusion of $1 in the initial mail invitations for a second sample of household addresses. Data collection for the first sample was from March to August 2017 and for the second sample from September 2017 to January 2018.

**Results**

We examined three sets of potential outcomes from reducing the length of the survey: decrease in nonresponse, change in estimates due to change in nonresponse bias, and change in estimates due to reduced measurement error.

***Nonresponse rates.*** Consistent with expectations, response rates were higher in the short survey condition. Shortening the survey by one half increased the response rate by one percentage point, shown in the last row of Table 1. The increase is statistically significant, yet somewhat smaller than expected. However, depending on the incentive condition, this represents 10% to 20% increase in response rate (.9/4.1=22% and .9/8.6=10%). Moreover, the substantive importance of the increase in response rates is from the effect on survey estimates, which we present in the next set of results.

Table 1. Nonresponse. Response rates by survey length and prepaid incentive.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | Survey Length | |  |
|  |  |  | Long (30 mins) | Short (15 mins) | Overall |
| Incentive | Only promised ($5) |  | 4.1% | 5.0% | 4.5% |
|  | Prepaid & promised ($1+$5) |  | 8.6% | 9.5% | 9.1% |
|  | Overall |  | 6.5% | 7.4% | 6.9% |

Note: The overall effect of length is significant at α=.05.

***Nonresponse bias.*** We divided the analyses by type of survey questions: demographic characteristics, behavioral questions, and attitudinal questions. Demographic characteristics as well as self-reported health, health conditions, physical activity, and smoking were all asked in Module A, the first module in the survey. Despite benefiting from the full sample by comparing groups 1&2 to groups 3&4, only one of these estimates was significantly different between the short and long survey conditions (under 40 years of age), shown in Table 2.

However, the short survey condition yielded estimates that showed more conservative responses on attitudinal questions, and for one of the questions—supporting preferential hiring and promotion of women—the estimate was six percentage points lower and statistically significant. This is despite the substantial sample size limitation as these questions were in Module D, which was asked in the same part of the questionnaire only in groups 2 and 4, i.e., a contrast that is based on only half of the respondents in the study.

Table 2. Survey length and nonresponse bias. Survey estimates from the long and short version of the survey and population benchmark estimates.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Long (30 minutes) | | Short (15 minutes) | | Benchmark |
|  | Percent | Standard Error | Percent | Standard Error | Survey  Estimates |
| *Demographic characteristics* |  |  |  |  | *2017 ACS* |
| Female | 57.2 | (1.7) | 56.1 | (1.6) |  |
| Under 40 years old | 34.1 | (1.6) | 39.1\* | (1.5) |  |
| Married | 53.5 | (1.7) | 57.4 | (1.6) |  |
| White | 82.6 | (1.3) | 80.3 | (1.3) |  |
| Hispanic | 7.4 | (0.9) | 8.3 | (0.9) |  |
| *Behavioral questions* |  |  |  |  | *2017 NHIS* |
| Excellent or very good health | 82.9 | (1.3) | 84.6 | (1.1) |  |
| High blood pressure | 29.9 | (1.6) | 29.2 | (1.4) |  |
| High cholesterol | 34.8 | (1.6) | 32.4 | (1.5) |  |
| Heart disease | 5.5 | (0.8) | 4.3 | (0.6) |  |
| Angina | 2.1 | (0.5) | 1.6 | (0.4) |  |
| Heart attack | 3.3 | (0.6) | 2.4 | (0.5) |  |
| Asthma | 15.5 | (1.2) | 16.3 | (1.2) |  |
| Ulcer | 8.3 | (0.9) | 9.1 | (0.9) |  |
| Cancer | 12.8 | (1.1) | 12.2 | (1.0) |  |
| Seizure disorder or epilepsy | 1.6 | (0.4) | 1.5 | (0.4) |  |
| Have health insurance | 92.0 | (0.9) | 92.7 | (0.8) |  |
| Walk or bike for at least 10 minutes | 44.2 | (1.7) | 46.5 | (1.6) |  |
| Vigorous exercise | 40.1 | (1.7) | 45.0 | (1.6) |  |
| Current smoker | 13.5 | (1.2) | 10.0 | (0.9) |  |
| *Attitudinal questions* |  |  |  |  | *2016 GSS* |
| Men are better suited emotionally for politics than are most women | 12.3 | (1.5) | 13.9 | (1.5) |  |
| Support preferential hiring and promotion of women | 35.3 | (2.2) | 29.3\* | (2.0) |  |
| A man will not get a job or promotion while an equally or less qualified woman gets one instead | 35.6 | (2.2) | 40.0 | (2.2) |  |
| Support sex education in the public schools | 89.7 | (1.4) | 86.4 | (1.5) |  |
| Support legal abortion if the woman wants it for any reason | 52.2 | (2.3) | 55.3 | (2.2) |  |

***Measurement error.*** Analyses for measurement error used the experimental manipulation of the order of the survey modules within the long survey conditions—group 1 vs. group 2. Table 3 presents the same attitudinal questions that were part of the nonresponse bias analysis. Here we see more conservative reports when the questions are asked late in the survey, with a seven percentage-point difference for men rated as better suited emotionally for politics than women and for support for preferential hiring and promotion of women. The latter is the same question that showed a significant difference for nonresponse bias, making it a potentially useful error indicator for future studies.

Table 3. Survey length and measurement error bias. Survey estimates when the questions are asked late or early in the survey, and population benchmark estimates.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Towards End | | Towards Beginning | | Benchmark |
|  | Percent | Standard Error | Percent | Standard Error | Survey  Estimates |
| *Attitudinal questions* |  |  |  |  | *2016 GSS* |
| Men are better suited emotionally for politics than are most women | 4.9 | (1.1) | 12.3\*\*\* | (1.5) |  |
| Support preferential hiring and promotion of women | 28.5 | (2.2) | 35.3\* | (2.2) |  |
| A man will not get a job or promotion while an equally or less qualified woman gets one instead | 37.5 | (2.4) | 35.6 | (2.2) |  |
| Support sex education in the public schools | 90.2 | (1.5) | 89.7 | (1.4) |  |
| Support legal abortion if the woman wants it for any reason | 55.9 | (2.5) | 52.2 | (2.3) |  |

**Conclusions and Next Steps**

Shortening the survey increased response rates regardless of the incentive protocol and led to differences in some estimates as a function of nonresponse and as a function of measurement error. Our next step is to complete this analysis for variables on other topics that were included and to bivariate and multivariate associations. In addition, we will compare the conditions using paradata that are designed to serve as proxy measures for measurement error. The main objective is to then multiply-impute the data and evaluate the split questionnaire design compared to a full survey design in terms of bias, variance, and mean square error.

**Discussion Questions**

* Would it be challenging to support reduction of survey length based more on measurement error than on nonresponse bias?
* Are there other outcomes as a function of survey length that warrant attention?
* A more controversial question for the group: what is the importance you would give to differences in response rates in terms of percentage points vs. relative increases in response rates?

**References**

Biemer, Paul P. 2000. An Application of Markov Latent Class Analysis for Evaluating Reporting Error in Consumer Expenditure Survey Screening Questions. RTI International, Research Triangle Park, NC: Technical Report for the US Bureau of Labor Statistics.

Biemer, Paul P. 2001. "Nonresponse Bias and Measurement Bias in a Comparison of Face to Face and Telephone Interviewing." *Journal of Official Statistics* 17:295-320.

Galesic, Mirta, and Michael Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73 (2):349-360. doi: 10.1093/poq/nfp031.

Groves, Robert M., and Lou Magilavy. 1984. "An Experimental Measurement of Total Survey Error." American Statistical Association.

Groves, Robert M., Eleanor Singer, Amy D. Corning, and Ashley Bowers. 1999. "A Laboratory Approach to Measuring the Joint Effects of Interview Length, Incentives, Differential Incentives, and Refusal Conversion on Survey Participation." *Journal of Official Statistics* 15 (2):251-268.

Heberlein, Thomas A., and Robert Baumgartner. 1978. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." *American Sociological Review* 43 (4):447-462.

Herzog, A. Regula, and Jerald G. Bachman. 1981. "Effects of Questionnaire Length on Response Quality." *Public Opinion Quarterly* 45 (4):549-559.

Hill, Bradford. 1951. "The Doctor's Day and Pay." *Journal of the Royal Statistical Society. Series A (General)* 114 (1):1-34.

Holland, P. W., and D. T. Thayer. 1985. "Section Pre-Equating in the Presence of Practice Effects." *Journal of Educational Statistics* 10:109-120.

Holland, P. W., and L. E. Wightman. 1982. "Section Pre-Equating: A Preliminary Investigation." In *Test Equating*, edited by P. W. Holland and D. B. Rubin, 271-297. New York: Academic Press.

Khare, Meena, Leyla K. Mohadjer, Trena M. Ezzati-Rice, and Joseph Waksberg. 1994. "An Evaluation of Nonresponse Bias in NHANES III (1988-91)." Proceedings of the section on Survey Research Methods, American Statistical Association.

Lorant, Vincent, Stefaan Demarest, Pieter-Jan Miermans, and Herman Van Oyen. 2007. "Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census." *Int. J. Epidemiol.* 36 (6):1292-1299. doi: 10.1093/ije/dym191.

Munger, Gail F., and Brenda H. Lloyd. 1988. "The Use of Multiple Matrix Sampling for Survey Research." *Journal of Experimental Education* 56:187-191.

Olson, Kristen. 2006. "Survey participation, nonresponse bias, measurement error bias, and total bias." *Public Opinion Quarterly* 70 (5):737-758.

Peytchev, Andrey A. 2007. "Participation Decisions and Measurement Error in Web Surveys." Ph.D. Doctoral Dissertation, Survey Methodology, University of Michigan.

Peytchev, Andy. 2011. "Correction for Survey Nonresponse and Measurement Error." 10th Conference on Health Survey Research Methods, Atlanta, GA, April 10, 2011.

Peytchev, Andy. forthcoming. "Consequences of Declining Survey Response Rates." *Annals and the American Academy of Political and Social Science*.

Peytchev, Andy, and Emilia Peytcheva. 2017. "Reduction of Measurement Error due to Survey Length: Evaluation of the Split Questionnaire Design Approach." *Survey Research Methods* 11 (4):361-368.

Peytchev, Andy, Emilia Peytcheva, and Robert M. Groves. 2010. "Measurement Error, Unit Nonresponse, and Self-Reports of Abortion Experiences." *Public Opinion Quarterly* 74 (2):319-327.

Raghunathan, Trivellore E., and James E. Grizzle. 1995. "A Split Questionnaire Survey Design." *Journal Of The American Statistical Association* 90 (429):54-63.

Shoemaker, David M. 1973. *Principles and Procedures of Multiple Matrix Sampling*. Cambridge, MA: Ballinger.