Survey Research Methods (2020)
Vol. 14, No. 1, pp. 31-42
doi:10.18148/srm/2020.v14i1.7294

© European Survey Research Association
ISSN 1864-3361
http://www.surveymethods.org

The Impact of Splitting a Long Online Questionnaire on Data Quality

Ioannis Andreadis
Aristotle University of Thessaloniki
Greece

Evangelia Kartsounidou
Aristotle University of Thessaloniki
Greece

Long self-administered questionnaires may suffer from lower response rates, higher drop-outs, and lower quality responses. A shorter questionnaire reduces the burden of respondents. Using this as a starting point, we test the following method: split the long questionnaire into sub-questionnaires; invite everyone to answer the first sub-questionnaire; when respondents complete the first sub-questionnaire, invite them to answer the second sub-questionnaire, and so on. We present evidence that after splitting a long questionnaire into two shorter parts, the response rates of these sub-questionnaires are significantly higher than the response rate of the original, undivided, long questionnaire. However, the cumulative response rate of both parts is lower than the response rate of the long undivided questionnaire. Finally, we show that the respondents of the survey using the original, long questionnaire: i) provide more non-substantive answers ("neither/nor") to the Likert-type scale items and ii) give shorter answers to the open-ended questions of the survey than the respondents of the split survey. On the other hand, there is no significant difference between the long and the split questionnaire on the other indicators of response quality we have tested: item-nonresponse, speeding and straight-lining. This paper presents some first insights on splitting a long questionnaire into shorter parts. For now, the results are not promising to suggest with confidence to split the long questionnaire for the purpose of obtaining high data quality. Further research is needed to find the optimal interval time between the sub-questionnaires or the optimal length of the sub-questionnaires in which the overall response rate is maximized.

Keywords: web surveys, long questionnaire, data quality, response rates, drop-outs

1 Introduction

Nowadays, web surveys are being used more and more often in social science, as a fast and low-cost mode of data collection. However, web surveys are associated with some serious drawbacks. Some respondents may become less motivated or less engaged at some point during the survey, especially when it is online (Chen, 2011; Fang, Wen, & Prybutok, 2014) and they may drop-out or provide responses of lower quality (i.e. satisficing).

The length of the survey instrument influences considerably the response rate, the drop-outs and the response quality of a survey (see, for instance, Ganassali, 2008). Although a large part of the respondents who abandon a questionnaire drop-out near the beginning of it, longer questionnaires are expected to suffer from more drop-outs. In addition, relying on the theory of satisficing (Krosnick, 1991), when respondents get tired, or lose their interest and motivation, they tend to respond less thoroughly to the questions of a survey (i.e. with minimal cognitive effort).Consequently, lengthy online questionnaires lead to lower data quality (see, for instance, Crawford, Couper, & Lamias, 2001; Galesic, 2006; Galesic & Bosnjak, 2009; Marcus, Bosnjak, Lindner, Pilischenko, & Schutz, 2007).

The need to create shorter survey instruments is becoming more urgent due to the increasing rate of people who respond to web surveys using their mobile devices. In general, mobile devices users are associated with lower completion rates and higher drop-out rates than PC users (Buskirk & Andrus, 2014; Guidry, 2012; Lattery, Bartolone, & Saunders, 2013; Mavletova, 2013; Mavletova & Couper, 2014; Sarraf, Brooks, & Cole, 2014). These findings have urged web survey designers to optimize their surveys for mobile devices (Andreadis, 2015b; Antoun, Couper, & Conrad, 2017; Lugtig, Toepoel, & Amin, 2016). Given that web surveys completed on mobile devices take longer than desktop or laptop surveys (Andreadis, 2015a; Cook, 2014; Couper, Kapteyn, Schonlau, & Winter, 2007; Lambert & Miller, 2015), the need to design shorter questionnaires is becoming more and more pertinent.

In some cases, researchers are not able to create short questionnaires. For instance, in many comparative studies, national teams are requested to include a translated version of the common core questionnaire of the project in its entirety.

Evangelia Kartsounidou Al. Fleming, 4, 54642, Thessaloniki, Greece (E-Mail: evankart@polsci.auth.gr)

In addition to the common questions, the final national questionnaire includes many questions of national interest. As a result, the final questionnaire becomes very long.

Scholars have explored different ways to create shorter survey instruments. The large variety of approaches highlights the methodological gap that exists in this area. One of the proposed methods is based on removing some of the questions included in the initial questionnaire, mainly by reducing the number of items used in the same scale (Maloney, Grawitch, & Barber, 2011; Mühlan, Bullinger, Power, & Schmidt, 2008). However, this process could undermine the reliability or the validity of the survey data (Maloney et al., 2011; Smith, McCarthy, & Anderson, 2000).

Another approach is to split the questionnaire into shorter sub-questionnaires retaining all the initial questions. The questions which are not displayed in the first part are simply moved to other sub-questionnaires. After splitting the questionnaire, we can follow two different methods. The first method called matrix design or sampling design or split questionnaire design (SQD) (Adigüzel & Wedel, 2008; Fricker et al., 2012; Herzog & Bachman, 1981; Raghunathan & Grizzle, 1995) involves the administration of the sub-questionnaires to different groups of the sample. The main drawback of this method is that a large number of missing data has to be imputed (Raghunathan & Grizzle, 1995). The second method of splitting tries to address the problem of missing data. Galesic and Bosnjak (2009) divided the questionnaire into three different parts (10, 20 and 30 minutes). Although they administered different parts of the questionnaire to different sample groups, they gave all respondents the opportunity to see and answer all questions of the questionnaire.

In this paper, we follow the approach used by Galesic and Bosnjak (2009) of giving respondents the opportunity to see all questions, but contrary to them, our respondents did not answer a sub-questionnaire immediately after completing the previous one. Our method allows respondents to have a long break between the sub-questionnaires and start fresh each part of the survey. The main objective of this paper is to examine if this break can improve the response rate and the response quality of a web survey.

The structure of the remaining sections of this paper is as follows. We start with a literature review on response rates and response quality of web surveys. Based on the literature review, we formulate our research hypotheses. We continue with a section that discusses the data and the experimental design we have used to estimate the impact of splitting a long questionnaire on the response rate and the response quality of the survey. Finally, we report the findings and present a conclusion with the discussion and implications for web survey designers.

2 Literature Review and Research Hypotheses

2.1 Response rate

Generally, there is evidence that response rates are declining over time both for online and offline surveys (Beebe et al., 2010; Curtin, Presser, & Singer, 2005). For online surveys, the problem of increased drop-out rates consequently limits the usage of survey data. For example, according to Hoerger (2010), approximately 10% of students participating in Internet-mediated university studies drop out almost immediately. After answering the first few items, the drop-out rate increases by 2% per circa 100 items of survey content. Andreadis (2013, 2014) also finds that web surveys suffer with many drop-outs in the first page of the questionnaire. The frequency of drop-outs is lower for the following pages and it can increase again when the respondents face a difficult question or when they lose their interest in the survey (Galesic, 2006).

In an experimental research on offline surveys, there is evidence that shorter surveys lead to higher response rates (Porter, 2004). More specifically, Yammarino, Skinner, and Childers (1991) found that response rates in mail surveys were reduced by 7.8% when using a questionnaire longer than four pages. In addition, Kalantar and Talley (1999) found that the response rate is higher in a short questionnaire of seven items than in a longer questionnaire of 32 items.

In online surveys, a negative relationship between the response rate and survey length is also confirmed (Liu & Wronski, 2018). Galesic and Bosnjak (2009) show that more respondents will start and fill in a survey that lasts 10 minutes than a longer survey that lasts 30 minutes. Mavletova (2013) examining drop-out rates in mobile and web surveys, found slightly (but not significant) higher drop-out rates in the longer survey (15 minutes) than in the shorter one (5 minutes) in both modes. Finally, Deutskens, de Ruyter, Wetzels, and Oosterveld (2004) observed that the respondents of a long internet-based survey abandoned the questionnaire earlier than the respondents of its shorter version.

Based on literature and given that the invitations sent to the respondents clearly stated a very short completion time, in our study we expect more completed questionnaires, fewer drop-outs and fewer refusals in the case of using the split questionnaires. Since the respondents who have completed the first sub-questionnaire are already engaged in our system and a relationship has been developed with them, we expect that the response rate of the next sub-questionnaire should also be very high. Consequently, we anticipate that the partial response rates of the sub-questionnaires will be significantly higher than the response rate of the long original questionnaire. In this case, their composite response rate will be also higher than the response rate of the long original questionnaire. Therefore, our first research hypothesis is:

H1 Splitting a long questionnaire will increase the response

rate.

2.2 Response Quality

Long questionnaires may result in satisficing behaviour (Barge & Gehlbach, 2012; Chen, 2011; Guidry, 2012). Thus, our second main research hypothesis is that splitting the questionnaire will also improve the response quality. In order to test this hypothesis, we use five indicators that we discuss in more details in the following paragraphs:

- Item nonresponse
- Mid-point response
- Speeding
- Straight lining
- Open-ended response

Galesic and Bosnjak (2009) argued that longer web questionnaires are associated with greater amounts of missing data on individual questions (item nonresponse). An increase of item nonresponse in longer surveys is also reported by Peytchev and Tourangeau (2005) and Galesic (2006). In a web experiment, Deutskens et al. (2004), found that the number of "don't knows" (which can be selected as an alternative to item nonresponse) is higher in a longer questionnaire. Thus, our first response quality hypothesis is:

H2a We expect to observe decreased item-nonresponse among respondents of the split questionnaire.

Choosing a mid-point response in scales is also an indicator of low interest or low effort (Weems & Onwuegbuzie, 2001). Respondents may choose mid-point responses when they do not process a question with the required effort. Hence, instead of expressing an opinion towards an issue, they prefer a more neutral stance. There is evidence that mid-point responses are similar to "No opinion" answers (Blasius & Thiessen, 2001). Gilljam and Granberg (1993) referred these responses as false negatives, arguing that people who choose a middle point answer, usually have an attitude towards the specific issue but they choose not to express it. Finally, Krosnick et al. (2001) in offline surveys observed more no-opinion responses among low-educated respondents to questions near the end of the questionnaire. Based on these reports, our second response quality hypothesis is:

H2b We expect that respondents of the long questionnaire will choose more often mid-point responses in the attitudinal items.

Speeding is an extreme type of satisficing and scholars explore ways to reduce it in web surveys (Conrad, Tourangeau, Couper, & Zhang, 2017). There is evidence showing that web respondents tend to give faster responses to questions closer to the end of the questionnaire (Andreadis, 2012; Yan & Tourangeau, 2008). Shorter response times can be a sign of burden and an indicator of low response quality. More specifically, it has been shown that very fast respondents (i.e. when their item response times are below specific thresholds) appear to give random answers and these cases are introducing noise to the dataset (Andreadis, 2014). Thus, we argue that as the time the respondents have to spend on the questionnaire increases, they will be less motivated, and they will start spending less time to answer each item. As a result, our third response quality hypothesis is:

H2c We expect more speeders among the respondents of the long original questionnaire.

Non-differentiation in the answers to grid questions, the so-called straight-lining, is another indicator of satisficing behaviour and low response quality (Greszki, Meyer, & Schoen, 2014; Schonlau & Toepoel, 2015). Galesic and Bosnjak (2009) observed limited variability in answers to grid-based questions (straight-lining) in a lengthy web questionnaire. Greater likelihood of straight-line responding, when the questionnaire is long is also pointed out by Herzog and Bachman (1981) in paper surveys. This leads to our fourth response quality hypothesis:

H2d We expect to observe decreased straight-lining in grid questions among respondents of the split questionnaire.

Finally, the length of the responses in open-ended questions could indicate the level of involvement or the effort made by the respondent (Deutskens, de Ruyter, & Wetzels, 2006; Schmidt, Calantone, Griffin, & Montoya-Weiss, 2005). Herzog and Bachman (1981) relying on an older paper survey, mentioned that open-ended questions near the beginning of the questionnaire receive more answers than those near the end of it. In addition, Galesic and Bosnjak (2009) observed that open answers are shorter near the end than near the beginning of the web questionnaire. Consequently, in longer surveys, we can expect shorter responses to open-ended questions, especially if they are placed near the end of the questionnaire. Thus, our final response quality hypothesis is:

H2e We expect that respondents of the split questionnaire will give longer answers to open-ended questions.

3 Data and Methodology

To deal with the aforementioned issues associated with long web questionnaires, we propose the following method: i) split the long online questionnaire into sub-questionnaires; ii) invite everyone to the first sub-questionnaire; iii) when

Table 1
Operationalization of the response quality measures

$$\text{Item Nonresponse ratio} = \frac{\text{Number of missing answers}}{\text{Total number of questions}}$$

$$\text{Mid point response ratio} = \frac{\text{Number of mid-point answers}}{\text{Number of valid answers}}$$

$$\text{Speeding} = \begin{cases} 1 & \text{Response time} < \text{Scanning threshold} \\ 0 & \text{Otherwise} \end{cases}$$

$$\text{Straight lining} = \begin{cases} 1 & \#\text{ of same response} = \#\text{ of valid answers} \\ 0 & \text{Otherwise} \end{cases}$$

$$\text{Open-ended response} = \#\text{ of characters}$$

respondents complete a sub-questionnaire, invite them to the following sub-questionnaire; and iv) repeat step iii for the rest sub-questionnaires (if any). In this method, contrary to the matrix design, the respondents will have the chance to answer all the sub-questionnaires provided if they complete each sub-questionnaire they are invited to. As a result, this approach reduces the number of missing data in the final dataset. In addition, it reduces the number of contacts with people who are not interested in participating: people who have not completed the first sub-questionnaire are not contacted again. Furthermore, with the proposed method, respondents have a break between the sub-questionnaires. This break ensures that every time they start answering a sub-questionnaire, they are not tired by their effort to answer the questions of the previous sub-questionnaire.

In this paper, we use the data from the 2015 Greek candidate survey that was conducted as a web survey after the legislative election of January 2015. According to the definitions given by American Association for Public Opinion Research (2016), the Greek candidate survey can be classified into the category: "Internet Surveys of Specifically Named Persons", i.e. "the target population is synonymous with the sampling frame and thus is defined as those persons on the list with Internet access and a working e-mail address". The target population was the group of all candidate MPs of the following Greek Parliamentary parties: The Coalition of Radical Left (SYRIZA), the New Democracy (ND), the Potami, the Independent Greeks (ANEL) and the Panhellenic Socialist Movement (PASOK).[1]

Most of the Greek candidate MPs have e-mail addresses, which are available online especially during the period of the electoral campaign. We have collected their email addresses using two methods: i) using search engines and visiting websites related to the candidates or to the Greek elections in general, and ii) we have asked the political parties to send us a list of their candidates along with their email addresses[2]. Combining the data from these two methods, we have created a list of candidates with a known e-mail address. Following AAPOR, we define this list as our target population, and we have sent them an invitation to participate in our survey.

Usually, the questionnaires of candidate surveys are very long. Like many other comparative projects, the candidate study uses a common core questionnaire to allow a comparative analysis across time and countries. In our study, we use the common core international questionnaire of the Comparative Candidate Survey (CCS)[3]. Each national team can include additional items in the questionnaire resulting in very long questionnaires. For instance, the final Greek Candidate Questionnaire consists of six thematic units and it requires approximately 50 minutes to complete it.

The units of our sample ($n = 1090$) were split randomly into two different groups, A and B. A web experiment was designed to study the effect of the questionnaire length on response rate, drop-outs and response quality. Hence, the respondents of Group A ($n_A = 533$) received the whole questionnaire (Questionnaire A), which includes 85 ques-

[1] The Communist Party of Greece (KKE) and Golden Dawn (GD) have always refused to provide a list of email addresses for their candidates. Thus, the Greek Candidate datasets could not include these parties.

[2] The second method of email collection was implemented according to our design for four parties (SYRIZA, ND, ANEL and PASOK). The fifth party (POTAMI) instead of sending us the requested list, asked its candidates to contact us directly (if they wanted to participate in our survey). These are the only candidates who can be classified as self-selected and we have excluded them from the analysis (see also footnote 6). Hence, the candidate MPs of Potami that we use in the analysis of this paper are only the candidates whose email address was found using search engines.

[3] National study directors of CCS project have developed a common core questionnaire that is used in the aftermath of the national elections. The questionnaire includes a variety of questions that cover a broad spectrum of politics.

tions[4]. The respondents of Group B (n_B = 557) received only the first part of the same questionnaire (Questionnaire B1), which includes 9 questions. This approach has allowed us to mention in the invitation to respondents in Group B that it will take them less than 10 minutes to complete it. The rest of the questions (76 questions) were sent later as a separate questionnaire (Questionnaire B2)[5]. The invitations for B2 were sent only to respondents who completed B1. Diagram 1 shows the design of this experiment.

The first invitations were sent by email in the period 16-18 February 2015 to the candidate MPs of both groups. The first reminder was sent on the 27th of February 2015. A second reminder was sent on the 19th of April 2015. On 31st of April, B1 was over and the candidates who have not completed the first part were never contacted again. As for Survey B2, we sent invitations on the 8th of April to the respondents who have completed B1 before the 7th of April, on the 21st of May to the respondents who have completed B1 between the 8th of April and the 20th of May and on the 25th of June to the respondents who have completed B1 after the 20th of May. For each respondent, before sending the invitation to Survey B2, we waited for at least fifteen days after they completed B1. We have followed this approach because we wanted to make sure that respondents perceive B1 and B2 as two separate surveys. Additional reminders were sent to both groups until the end of July[6], while keeping the maximum number of contacts for each respondent less than or equal to six.

To check the first research hypothesis (H1), we start by comparing three types of response behaviours: completed questionnaires, drop-outs and refusals between surveys A and B1. Given that B1 does not include all the questions, we repeat the comparison between surveys A and B.

As for the quality of the two surveys (H2), we use five indicators similar to the indicators other researchers used to explore response quality (Table 1); these are as follows: i) item-nonresponse, ii) mid-point responses in scale items, iii) the time spent on the questionnaire (speeding), iv) straight lining in the grid questions of the questionnaire (non-differentiation), and v) the number of characters in the open-ended questions of the questionnaire (Barge & Gehlbach, 2012; Chen, 2011; Guidry, 2012; Mavletova, 2013).

To check H2a, we calculate the ratio of missing answers of each respondent (i.e. we divide the number of questions that have not been answered by the respondent with the total number of questions) in each section of the questionnaire. To check H2b, we select only the Likert-type scale items of the questionnaire where the mid-point indicates a neutral position. Then, we calculate the ratio of mid-point responses (neither agree nor disagree) of each respondent (i.e. "number of mid-points"/ "number of valid answers") in each section of the questionnaire.

The time spent on answering a question is related to the response quality. This is not a simple relationship because both longer response times (Yan, 2015), and shorter response times (Conrad et al., 2017) may indicate responses of lower quality. However, extremely short response times, i.e. when the time spent on a question is not even enough to read and comprehend the question, we can safely conclude that the response is of low quality. To find these cases, we use the "scanning" threshold method. To calculate the scanning threshold for a question, we add the minimum time needed to read the question and the minimum time needed to answer the questions. For the former, we count the number of characters of the question and we divide it with the number of characters that can be "read" in one second when an average reader scans a document. For the latter, we use the minimum time reported by Bassili and Fletcher (1991). Thus, the "scanning" threshold method provides the minimum time needed to read and answer an attitudinal question given the length of the question text (Andreadis, 2014). Respondents who dedicate less time than the "scanning" threshold in any of the attitudinal questions of the questionnaire are flagged as speeders. To check H2c, we calculate the percentage of speeders in each survey.

Straight lining in grid questions is an indication of low response quality. It is assumed that respondents who straight-line do not pay the required attention to the questions, providing incorrect answers. To check H2d, we compare the patterns in the grid questions of the questionnaire and we calculate the percentage of straight lining in the grid questions of the questionnaire.

To check H2e, we measure the number of characters of the answers to the open-ended questions given by the respondents of groups A and B. Longer answers could indicate a more complete and well-thought answer. In this measurement, we exclude the cases where respondents have skipped the open-ended questions (i.e. the length of the answer is 0) because they have already been measured as item nonresponses (H2a).

Finally, in the tables included in the following section, we compare the mean and the median values for all the indicators of response quality between Surveys A and B. We check for significant differences between the surveys using the Mann-Whitney test. The latter was preferred instead of a

[4] One question corresponds to one screen page. In the total number of questions, we did not include questions about European Parliament Elections, which have been displayed only to respondents who have run as candidates both for the National Parliament in 2015 and the European Parliament in 2014. These respondents have been excluded from the analysis presented in this paper.

[5] Apparently, by combining questionnaires B1 and B2, we get questionnaire A.

[6] The reader should take into account that June 2015 was a very difficult month to conduct the survey within a turbulent economic and political situation, a potential Grexit at stake and a forthcoming referendum (5th of July).

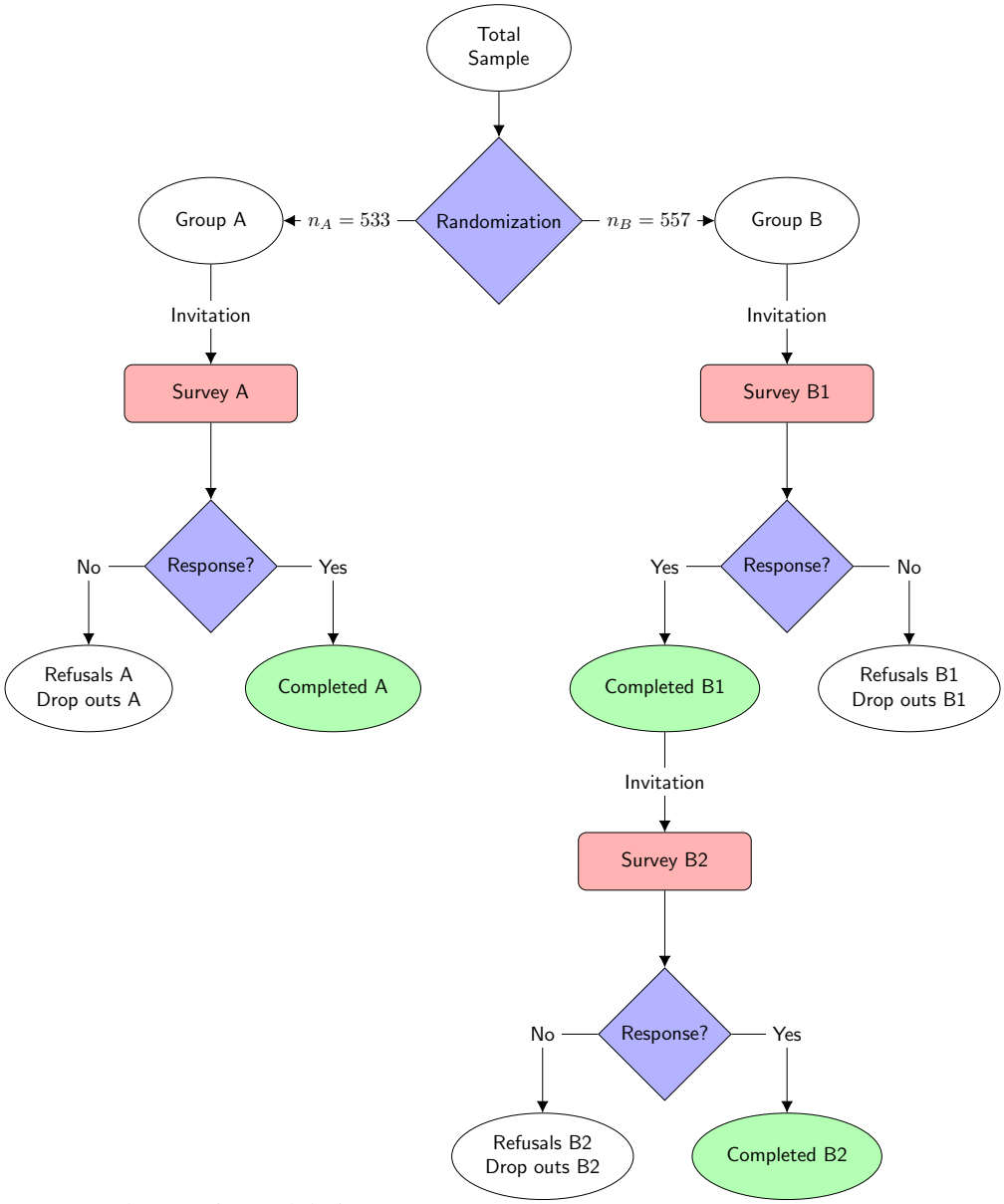

Figure 1. The experimental design

parametric test (e.g. t-test) because indicators in both groups do not follow a normal distribution.

4 Findings

In Table 2, we display the number of the invitations, the number of non-interviews, the number of completed questionnaires, the number of refusals and the number of drop-outs for surveys A and B1. A total of 533 invitations were sent to the candidate MPs of the first group (Survey A)[7]. In B1, 557 candidates were invited to participate[8]. However, many candidates did not click on the link in the invitation to the questionnaire neither of the Survey A (47.8%) nor of the Survey B1 (48.7%). We do not know the exact reasons

[7] In the beginning, we sent 611 invitations via email where only the 572 email addresses were active or correct. Therefore, the response rate of the survey is calculated based on the total invited candidate MPs with valid email addresses. There were some respondents (see footnote 2) who have contacted us asking to participate in the survey before we sent them the invitation. These people were already extremely motivated to participate in the survey and we have placed them in Group A but we exclude them from the analysis presented here. After their exclusion, the total number in Group A is 533.

[8] A total of 598 candidates were initially invited to B1; however, only the 557 email addresses were active or correct. Therefore, the response rate of the survey is calculated based on the total invited

for the lack of any reaction (we do not even know if they have received the invitation). We refer to these cases as "non-interview" with unknown eligibility (The American Association for Public Opinion Research, 2016). If we focus on the completed questionnaires, we can observe that almost one-third of the questionnaires (34.5%) of Survey A and 40.8% of the questionnaires of Survey B1 were completed. The difference of completed questionnaires between the two groups is approximately 6% points.

As refusals, we describe the candidate MPs who have clicked on the link in their survey invitation, but they have left the questionnaire without answering any of the questions. According to American Association for Public Opinion Research (2016), refusals belong to a larger category called "eligible, non-interview". In our case, refusals correspond to "implicit refusals", i.e. code 2.112 of AAPOR classification. Approximately, 10.1% of the candidate MPs who were invited to participate in Survey A and 7% in Survey B1 belong to this category.

There is a remarkable difference in the number of drop-outs between the two surveys. In group A, 7.5% of drop-outs are observed while in group B only 3.6%. This finding becomes more interesting if we examine the last question answered or seen before abandonment. In Survey A, only 25% of the drop-outs occur in the first five pages. On the other hand, in B1 almost all the drop-outs (90%) are noticed in the first five pages. We argue that drop-outs observed at the beginning of the questionnaire are the result of respondents who were less motivated to answer the questionnaire. On the other hand, drop-outs observed later in the questionnaire are mainly due to fatigue or burden. Hence, we argue that the reduced length of the Survey B1 encouraged the candidates to complete the questionnaire.

The comparison between surveys A and B1 shows that the length of the survey influences the response behaviour types. The smaller percentage of drop-outs along with the smaller percentage of refusals, explains the larger response rate in B1 than in A. We have calculated Pearson's Chi-Square to test if the distribution of responses differs significantly between Survey A and Survey B1. The test shows that the types of response behaviour are not independent of the length of the survey (Pearson's Chi Square = 12.38, $p = 0.004$).

As mentioned earlier, we have sent invitations to participate in Survey B2 only to the respondents who have completed B1[9]. According to Table 2, from a total of 225 invitations, we have 114 completed questionnaires, i.e. the response rate (50.7%) is rather high. However, when we compared the response rates of Survey A and Survey B consisting of the two parts (B1 and B2), we find that the composite response rate of Survey B is lower than the response rate of Survey A. In absolute numbers, in Survey A, we have 184 completed questionnaires out of 533 invitations, which leads to 34.5% response rate. In Survey B, we have 114 completed questionnaires out of 557 initial invitations, which gives us 20.46% composite response rate. We return to this finding in the final section of this paper.

As for hypothesis H2a, there are no significant differences between Survey A and Survey B regarding item nonresponse (Table 3). Looking closer to our data, we have noticed that most of the times, item nonresponse was not related to the burden of the respondents and it occurred at specific questions of the questionnaire. In our study, we have observed that many cases of item nonresponse occur when candidates are asked to provide information that they do not want to share, e.g. when we ask them details about the budget of their electoral campaign (section 6).

Table 4 shows the measures of central tendency (mean and median) of the share of mid-point responses in the attitudinal scale questions of the different sections of the questionnaire. For this table, we use questions with five-point scales, where the mid-point (three), refers to neither agree nor disagree. We do not observe any statistically significant differences between surveys A and B in the questions of sections one and two. However, we notice a significant difference in Section 3, where respondents of Survey A chose more often the middle points of the scale than respondents of Survey B: while in Survey A on average 17% of all responses were mid-point responses, in Survey B only 14% of all responses on average were mid-point responses. Finally, we observe that respondents tend to choose more often the "neither agree nor disagree" option in the scale questions which are placed later in the questionnaire.

Table 5 presents the percentage of respondents who answer faster (speeding) to at least one attitudinal question in Surveys A and B[10]. The percentage of speeders is relatively low in both surveys. Although we do not find any statistically significant differences between the two surveys, we observe that in Section 1, Survey B has more speeders (7%) than Survey A (5%) while in Section 3 there are more speeders in Survey A and has the highest number of speeders in both surveys. Especially, in Survey A the increase in speeders is sharper, which could be a sign of burden.

Another indicator of survey quality (H2d) that we exam-

---

candidate MPs with valid email addresses.

[9] At this point, it is worth mentioning that the time that B2 took place was far from optimal. Shortly after we started sending invitations for B2, the financial and political situation in Greece became extremely difficult. Greece was unable to pay the installment to the International Monetary Fund due on 5th of June 2015. The country was facing the danger of a Grexit, i.e. having to leave Eurozone and the European Union. Finally, the prime-minister announced a referendum, which has divided the Greek society, including of course, the respondents of our survey.

[10] We have excluded the first question of both questionnaires from this analysis because the response times of this item were not measured correctly due to technical reasons.

Table 2
Invitations, non-interviews, completed questionnaires, and drop-outs in Survey A, Survey B1, and Survey B2

	Survey A		Survey B1		Survey B2	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Unknown eligibility "Non-interview"	255	48	271	49	74	33
Completed questionnaires	184	35	227	41	114	51
Drop-outs	40	8	20	4	15	7
Refusals	54	10	39	7	22	10
Total Invited	533	100	557	100	225	100

The total invitations sent in B2 are 225, instead of 227, i.e. the number of completed questionnaires in B1. These two respondents of B1 answered almost all the questions of the survey and for this reason, we considered their questionnaires as completed. However, they have not clicked on the submit button, so we did not send them an invitation for B2.

Table 3
Item-nonresponse in the sections of the completed questionnaires in Survey A and Survey B.

	Survey A			Survey B			
	N	Mean	Median	N	Mean	Median	*p*-value[a]
Section1	220	0.06	0.00	227	0.08	0.00	0.914
Section 2	220	0.06	0.00	114	0.07	0.00	0.786
Section 3	220	0.08	0.05	114	0.08	0.05	0.301
Section 4	220	0.07	0.03	114	0.07	0.03	0.661
Section 5	220	0.06	0.00	114	0.07	0.00	0.161
Section 6	220	0.08	0.06	114	0.08	0.06	0.120

[a] Mann-Whitney-Test

Table 4
Mid-point responses in scale questions of the completed questionnaires in Survey A and Survey B

	Survey A			Survey B			
	N	Mean	Median	N	Mean	Median	*p*-value[a]
Section 1	217	0.14	0.13	219	0.12	0.10	0.072
Section 2	216	0.12	0.11	113	0.10	0.11	0.551
Section 3	214	0.17	0.13	114	0.14	0.13	0.025

Ratio of mid-point responses calculated in each section of the questionnaire.
[a] Mann-Whitney-Test

Table 5
Speeders in the sections of the completed questionnaires in Survey A and Survey B

	Survey A			Survey B			
	N	Mean	Median	N	Mean	Median	*p*-value[a]
Section 1	219	0.05	0.00	227	0.07	0.00	0.196
Section 2	219	0.05	0.00	114	0.06	0.00	0.805
Section 3	219	0.18	0.00	114	0.15	0.00	0.503

[a] Mann-Whitney-Test

Table 6
Straight lining in the grid questions of the completed questionnaires in Survey A and Survey B

	Survey A			Survey B			
	N	Mean	Median	N	Mean	Median	p-value[a]
Section 1	217	0.02	0.00	220	0.01	0.00	0.116
Section 2	217	0.10	0.00	113	0.10	0.00	0.394
Section 3	214	0.00	0.00	114	0.00	0.00	0.465

[a] Mann-Whitney-Test

Table 7
The number of characters in the open-ended questions of the completed questionnaires in Survey A and Survey B

	Survey A			Survey B			
	N	Mean	Median	N	Mean	Median	p-value[a]
Open ended 1	204	84.48	48.0	95	125.2	62	0.001
Open ended 2	172	189.56	108.5	93	222.4	136	0.2074

[a] Mann-Whitney-Test

ine is the straight lining in the grid questions. Table 6 shows that the percentage of straight lining is extremely low in all the grid questions regardless of the position of the grids in the questionnaire. In addition, the comparison of the respondents who provided straight-line responses between Survey A and Survey B does not exhibit a statistically significant difference (none of the p-values is less than 0.05). Hence, we cannot reject the null hypothesis that indicates that the percentage of straight lining does not differ significantly between groups A and B for all grid questions of our questionnaire.

In Table 7, we can observe the median number of characters of the responses to the open-ended questions of the questionnaire. The first open-ended question refers to the most important problems of the country and it is placed near the beginning of Survey B2[11]. The second question is the last question of the questionnaire of both Survey A and Survey B. Respondents were asked to mention what "left" and "right" means in politics for them[12]. Table 7 shows that the respondents of Survey B give constantly longer answers in both open-ended questions than the respondents of Survey A. Moreover, the statistically significant difference in characters in the first question could indicate that the respondents of Survey A are more tired than the respondents of Survey B, who have started completing B2 after a break. In general, the shorter responses in Survey A may be a result of the respondents' burden because they were answering a longer questionnaire without any breaks.

5 Discussion

In this paper, we have studied the impact of splitting a long questionnaire of a web survey on data quality. We have shown that there are significant differences in drop-outs, refusals and completed questionnaires between a long and a short survey. In addition, we have provided evidence that after splitting a long questionnaire into shorter parts (Survey B), the response rates of these sub-questionnaires (B1 and B2) are significantly higher than the response rate of the original, undivided, long questionnaire (Survey A). However, the composite response rate of Survey B is lower than the response rate of Survey A, due to the lower response rate of Survey B2.

We think that there are three possible factors that may have contributed to this result: i) the period that B2 was on the field, ii) the long break between B1 and B2, and iii) the length of B2. The first factor was exogenous: the economic, political and social turbulence in Greece while B2 was conducted could explain the lower than expected response rate of B2. The other two factors were results of our design. Our intention was to study how respondents would react to a second sub-questionnaire after allowing a long break (of more than 15 days) after the completion of the first sub-questionnaire. Although this long break allows the respondents to start fresh when they start answering the second survey, a shorter break could have given better results. In addition, in our experiment, the second sub-questionnaire was still very long because we wanted to have only two sub-questionnaires while keeping the first sub-questionnaire extremely short.

[11] The wording of this item is as follows: Which are the three most important problems of the country?

[12] The wording of these items is as follows: We have one last question for you: can you please, very briefly, indicate what "left" and "right" mean in politics to you nowadays?

As for the response quality of the surveys, we have shown that the quality of the responses in the split survey (Survey B) is slightly higher than in Survey A. In two out of the five quality indicators we have used, we observed statistically significant differences between the two surveys. The respondents of the long, undivided survey chose more frequently midpoint responses in the attitudinal questions and gave shorter answers to the open-ended questions of the survey than the respondents of split survey. On the other hand, there are no significant differences between the two surveys in terms of item-nonresponse, speeding, and straight-lining. However, if we focus on the rate of speeders, we also observe a tendency of lower response quality in Survey A than in Survey B. Therefore, we can conclude that there is an increase in the quality of the responses after splitting and our research hypothesis about the response quality is partially confirmed.

Concluding this paper, we present possible implications of our study for web survey designers. Our findings suggest that by splitting a long questionnaire, survey designers should expect less non-substantive answers ("neither/nor") to the Likert-type scale items and longer answers to the open-ended questions of the survey. In addition, by splitting the questionnaire into shorter parts, we can expect significantly higher response rates in each part, but, at least in our experiment, the cumulative response rate of both parts was lower than the response rate of the long undivided questionnaire. Thus, we invite the community of web survey practitioners struggling for higher response rates and better response quality to further investigate the method of splitting a long questionnaire into shorter parts. Further research is needed to find the optimal duration of the breaks between the sub-questionnaires and to investigate the optimal length of the sub-questionnaires, at which the overall response rate is maximized.

Acknowledgements

This work has been supported (grant number 714) by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI). In addition, we would like to thank the discussants and participants of the "Matrix design for social surveys" in ESRA Conference 2017 and two anonymous referees for their insightful and extremely valuable comments on our paper. Their feedback has helped us to significantly improve this paper. Needless to say, all remaining errors are the authors'.

References

Adigüzel, F. & Wedel, M. (2008). Split questionnaire design for massive surveys. *Journal of Marketing Research*, *45*(5), 608–617. doi:10.1509/jmkr.45.5.608

American Association for Public Opinion Research. (2016). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (9th ed.). Washington: AAPOR.

Andreadis, I. (2012). *To clean or not to clean? Improving the quality of VAA data.* IPSA World Congress.

Andreadis, I. (2013). Can timestamp analyses show the bottlenecks in web surveys. In *5th global wageindicator conference* (pp. 84–88). Amsterdam.

Andreadis, I. (2014). Data quality and data cleaning. In S. Garzia & D. Marschall (Eds.), (pp. 79–91). ECPR Press.

Andreadis, I. (2015a). Comparison of response times between desktop and smartphone users. In D. Toninelli, R. Pinter, & P. D. Pedraza (Eds.), *Mobile research methods: Opportunities and challenges of mobile research methodologies* (pp. 63–79). London: Ubiquity Press. doi:10.5334/bar.e

Andreadis, I. (2015b). Web surveys optimized for smartphones: Are there differences between computer and smartphone users? *Methods, Data, Analyses*, *9*(2), 213–228. doi:10.12758/mda.2015.012

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of mobile versus pc web on survey response quality. *Public Opinion Quarterly*, *81*(S1), 280–306. doi:10.1093/poq/nfw088

Barge, S. & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182–200. doi:10.1007/s11162-011-9251-2

Bassili, J. N. & Fletcher, J. F. (1991). Response-time measurement in survey research a method for CATI and a new look at nonattitudes. *Public Opinion Quarterly*, *55*(3), 331. doi:10.1086/269265

Beebe, T. J., Rey, E., Ziegenfuss, J. Y., Jenkins, S., Lackore, K., Talley, N. J., & Locke, R. G. (2010). Shortening a survey and using alternative forms of prenotification: Impact on response rate and quality. *BMC Medical Research Methodology*, *10*(1), 50. doi:10.1186/1471-2288-10-50

Blasius, J. & Thiessen, V. (2001). Methodological artifacts in measures of political efficacy and trust: A multiple correspondence analysis. *Political Analysis*, *9*(01), 1–20. doi:10.1093/oxfordjournals.pan.a004862

Buskirk, T. D. & Andrus, C. H. (2014). Making mobile browser surveys smarter. *Field Methods*, *26*(4), 322–342. doi:10.1177/1525822X14526146

Chen, P.-S. D. (2011). Finding quality responses: The problem of low-quality survey responses and its impact on accountability measures. *Research in Higher Education*, *52*(7), 659–674. doi:10.1007/s11162-011-9217-4

Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2017). Reducing speeding in web surveys by providing immediate feedback. *Survey Research Methods*, *11*(1), 45–61. doi:10.18148/srm/2017.v11i1.6304

Cook, W. A. (2014). Is mobile a reliable platform for survey taking? *Journal of Advertising Research*, *54*(2).

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an internet survey. *Social Science Research*, *36*(1), 131–148. doi:10.1016/j.ssresearch.2005.10.002

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perceptions of burden. *Social Science Computer Review*, *19*(2), 146–162. doi:10 . 1177 / 089443930101900202

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, *69*(1), 87–98. doi:10.1093/poq/nfi002

Deutskens, E., de Ruyter, K., & Wetzels, M. (2006). An assessment of equivalence between online and mail surveys in service research. *Journal of Service Research*, *8*(4), 346–355. doi:10.1177/1094670506286323

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters*, *15*(1), 21–36. doi:10.1023/B:MARK.0000021968.86465.00

Fang, J., Wen, C., & Prybutok, V. (2014). An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. *Computers in Human Behavior*, *30*, 335–343. doi:10.1016/J.CHB.2013.09.019

Fricker, S., Creech, B., Davis, J., Gonzalez, J., Tan, L., & To, N. (2012). *Does length really matter? Exploring the effects of a shorter interview on data quality, nonresponse, and respondent burden.* Paper presented on Federal Committee on Statistical Methodology 2012 Research Conference. Washington DC, USA.

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, *22*(2), 313–328.

Galesic, M. & Bosnjak, M. (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

Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, *2*(1), 21–32. doi:10 . 18148 / srm / 2008.v2i1.598

Gilljam, M. & Granberg, D. (1993). Should we take don't know for an answer? *Public Opinion Quarterly*, *57*(3), 348. doi:10.1086/269380

Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. In *Online panel research: A data quality perspective* (pp. 238–262). Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/9781118763520.ch11

Guidry, K. R. (2012). *Response quality and demographic characteristics of respondents using a mobile device on a web-based survey.* Annual Meeting of the American Association for Public.

Herzog, A. R. & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*(4), 549. doi:10.1086/268687

Hoerger, M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: Implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior, and Social Networking*, *13*(6), 697–700. doi:10.1089/cyber.2009.0445

Kalantar, J. S. & Talley, N. J. (1999). The effects of lottery incentive and length of questionnaire on health survey response rates: A randomized study. *Journal of Clinical Epidemiology*, *52*(11), 1117–1122. doi:10.1016/S0895-4356(99)00051-7

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. doi:10.1002/acp.2350050305

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., & Conaway, M. (2001). The impact of "no opinion" response options on data quality. *Public Opinion Quarterly*, *66*(3), 371–403. doi:10.1086/341394

Lambert, A. D. & Miller, A. L. (2015). Living with smartphones: Does completion device affect survey responses? *Research in Higher Education*, *56*(2), 166–177. doi:10.1007/s11162-014-9354-7

Lattery, K., Bartolone, G. P., & Saunders, T. (2013). *Optimizing surveys for smartphones: Maximizing response rates while minimizing bias.* In CASRO Online Research Conference. San Francisco, CA.

Liu, M. & Wronski, L. (2018). Examining completion rates in web surveys via over 25,000 real-world surveys. *Social Science Computer Review*, *36*(1), 116–124. doi:10.1177/0894439317695581

Lugtig, P., Toepoel, V., & Amin, A. (2016). Mobile-only web survey respondents. *Survey Practice*, *9*(3).

Maloney, P., Grawitch, M. J., & Barber, L. K. (2011). Strategic item selection to reduce survey length: Reduction in validity? *Consulting Psychology Journal: Practice and Research*, *63*(3), 162–175. doi:10.1037/a0025604

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schutz, A. (2007). Compensating for low topic interest

and long surveys: A field experiment on nonresponse in web surveys. *Social Science Computer Review*, *25*(3), 372–383. doi:10.1177/0894439307297606

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review*, *31*(6), 725–743. doi:10.1177/0894439313485201

Mavletova, A. & Couper, M. P. (2014). Mobile web survey design: Scrolling versus paging, SMS versus e-mail invitations. *Journal of Survey Statistics and Methodology*, *2*(4), 498–518. doi:10.1093/jssam/smu015

Mühlan, H., Bullinger, M., Power, M., & Schmidt, S. (2008). Short forms of subjective quality of life assessments from cross-cultural studies for use in surveys with different populations. *Clinical Psychology & Psychotherapy*, *15*(3), 142–153. doi:10.1002/cpp.573

Peytchev, A. & Tourangeau, R. (2005). *Causes of context effects: How questionnaire layout induces measurement error.* In Presented at the American Association for Public Opinion Research (AAPOR) 60th Annual Conference.

Porter, S. R. (2004). Raising response rates: What works? *New Directions for Institutional Research*, *2004*(121), 5–21. doi:10.1002/ir.97

Raghunathan, T. E. & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, *90*(429), 54–63. doi:10 . 1080 / 01621459.1995.10476488

Sarraf, S., Brooks, J., & Cole, J. S. (2014). *Taking surveys with smartphones: A look at usage among college students.* AAPOR Annual Conference, Anaheim.

Schmidt, J. B., Calantone, R. J., Griffin, A., & Montoya-Weiss, M. M. (2005). Do certified mail third-wave follow-ups really boost response rates and quality? *Marketing Letters*, *16*(2), 129–141. doi:10 . 1007 / s11002-005-2291-7

Schonlau, M. & Toepoel, V. (2015). Straightlining in web survey panels over time. *Survey Research Methods*, *9*(2), 125.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. doi:10.1037/1040-3590.12.1.102

Weems, G. H. & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development*, *34*(3).

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly*, *55*(4), 613. doi:10.1086/269284

Yan, T. (2015). Assessing quality of answers to a global subjective well-being question through response times. *Survey Research Methods*, *9*(2), 101–109. doi:10 . 18148/srm/2015.v9i2.6107

Yan, T. & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *22*(1), 51–68. doi:10.1002/acp.1331