methods, data, analyses | Vol. 11(2), 2017, pp. 135-162 DOI: 10.12758/mda.2017.02

Analyzing Survey Characteristics, Participation, and Evaluation Across 186 Surveys in an Online Opt-In Panel in Spain

Melanie Revilla
RECSM-Universitat Pompeu Fabra

Abstract

Survey designers often ask about the best length for their questionnaires and the best format for their questions. Much research has already addressed these issues. However, the answers to these questions may vary with the population of interest, the mode of data collection used, and other factors.

The goal of this paper is twofold:

1. To give an overview of the present situation in opt-in online panels, in terms of survey characteristics, participation, and evaluation, by reviewing 186 surveys managed by the panel company Netquest in Spain in 2016. This will be useful to determine areas where further research needs to focus.
2. To study how key characteristics of questionnaires impact survey evaluation and levels of survey break-off. This will allow us to highlight the characteristics that best reduce break-off and improve respondents' survey evaluation.

Based on these results, we will propose practical recommendations for future survey design within the framework of opt-in online panels.

Keywords: opt-in online panels, questionnaire characteristics, participation, survey evaluation, mobile devices

1 Introduction

Research is increasingly relying on survey data, and thus on individuals' willingness to participate in surveys and provide quality responses. Designing and implementing a survey requires numerous decisions, all of which may affect respondents' willingness to participate in a given survey or accept future invitations, as well as the overall quality of the data obtained. In order to achieve high participation and high overall data quality, when designing new surveys, frequent questions include: What is the advised length of a questionnaire?[1] How should questions be formatted? What can be done to limit break-off?

The goal of this paper is twofold:

1. By reviewing 186 surveys run by the panel company Netquest in Spain in 2016, the paper aims to provide an overview of the current situation in opt-in online panels, in terms of: survey characteristics (e.g., target populations, quotas, survey content, including topic, question formats, estimated survey length, and incentives), participation (i.e. the number of panelists invited, the number that began the survey, and the numbers that screened out, broke off, or completed the entire survey), and evaluation of the survey itself (each survey included a final question allowing respondents to evaluate the survey they just finished, on a scale from "1-survey very badly done" to "5-survey very well done").

As Netquest provides data to all kinds of clients, agencies and researchers, we expect this overview to allow us to identify which target populations, survey topics, and question formats are most commonly used over a large range of research. This can be useful for at least two reasons: first, it helps us determine areas where further research is needed; second, it helps us identify areas where there are large disparities between the knowledge found in literature and what is done in practice. For instance, if the literature provides clear evidence that a specific question format performs worse than another, but we observe that the less efficient format is used more often in practice, we know where to channel our efforts when transferring knowledge from the academic world to the real practice of online surveying.

---

1 This is, for instance, what the author claims in this post: https://www.surveygizmo.com/survey-blog/how-long-can-a-survey-be/

Acknowledgements

I would like to acknowledge Netquest for providing me with the necessary data, and thank Gerardo Ortiz in particular for his huge support in getting the information needed, as well as Carlos Ochoa for his useful advice throughout the process. I also thank two anonymous reviewers and the Editor for their useful comments on a previous draft of this manuscript.

Direct correspondence to

Melanie Revilla, RECSM-Universitat Pompeu Fabra, Mercè Rodoreda 24.406, Ramón Trías Fargas, 25-27, 08005 Barcelona, Spain
E-mail: melanie.revilla@hotmail.fr or melanie.revilla@upf.edu

This overview also presents the overall level of participation and break-off, as well as respondents' average evaluation of surveys. These aspects may be more specific to the panel studied, as practical decisions concerning incentives or whether to announce survey length could affect these variables.

2. The paper also studies how break-off levels and survey evaluation are related to key characteristics of the questionnaires: topic, question type, and estimated survey length.

This second part seeks to identify whether survey-design decisions affect the break-off rate and participants' evaluation of the survey, and, if so, which decisions matter more. There is a tremendous unmet demand from online opt-in panels for survey-design guidelines. Based on our results, we will make some practical recommendations for future survey design within the framework of opt-in online panels.

The remainder of this paper is organized as follows: Section 2 offers background information, Section 3 provides information about the methodology, data and analyses conducted, Section 4 presents the main results, and Section 5 concludes.

2 Background

Much research has been done on survey characteristics, and it has chiefly focused on the effects of the topic, question format, survey length, and incentives on survey participation (mainly in terms of response rates) and on other aspects of data quality, such as break-off rates and survey evaluation (see Schuman & Presser, 1981; Sudman & Bradburn, 1982; Oppenheim, 1992; Tourangeau, Rips, & Rasinski, 2000; Brace, 2004; Saris & Gallhofer, 2014).

Most of these studies investigated face-to-face, telephone or postal mail surveys. However, in the past 15 to 20 years, web surveys have gained increasing traction. This new data collection mode differs at several levels from more traditional modes (de Leeuw, 2005): for instance, web surveys are computer-assisted and self-completed, and the stimulus is usually visual. Because it is easier to close a tab than to ask an interviewer to leave your home half-way through the questionnaire, it is expected that more respondents will break off online than during face-to-face surveys. Besides, with web surveys, respondents cannot turn to the interviewer if they have difficulty understanding or experience technical problems (e.g., if the webpage does not load). Thus, question layout and formulation could be even more important than in a face-to-face survey. In addition, different recommendations were needed about how to design these web surveys. This generated a lot of new research (e.g., Couper, 2000; Couper, Traugott, & Lamias, 2001; Dillman, 2000; Dillman & Bowker, 2001; Lozar Manfreda, Batagelj, & Vehovar, 2002; Marcus,

et al., 2007; Couper, 2008; Galesic & Bosnjak, 2009; Bethlehem & Biffignandi, 2011; Tourangeau, Conrad, & Couper, 2013), from which lists of recommendations have been extracted: see, for instance, Parsons' paper (2007) on web survey best practices.

Nevertheless, as Couper and Miller (2008) have pointed out, web surveys can be extremely different, and what applies in one case does not necessarily hold in another. One crucial distinction, in particular when studying survey participation, is the difference between one-time surveys and surveys done within the framework of online panels, in which the same group of people who agreed to participate in surveys are regularly contacted by the same online panel company to complete questionnaires, usually in exchange for money or gifts, as this helps increase collaboration (Göritz, 2006). By nature, panels need to retain respondents. Consequently, survey experience is more important for panels, as it can affect future participation. Cape (2012, p.6) stresses the need to find better ways to motivate online panelists, and recommends moving them toward "intrinsic motivation" to keep them active: "[Online panel companies] have a finite resource, which costs money to build and develop, and the industry as a whole is forcing down revenues per interview. The more we can do to motivate our panelists, the easier they will be to recruit and retain."

In addition, within online panels, it is important to differentiate between probability-based panels and opt-in or access panels (Callegaro, Lozar Manfreda, & Vehovar, 2015, chapter 5.2.). In probability-based panels, a random sample is drawn from the population and the selected units are contacted and invited to participate in the panel. Units who do not have Internet access are usually provided with it. On the other hand, in opt-in or access panels, individuals volunteer to participate. If they do not have Internet access, they cannot be part of the panel. This raises the issues of how representative for the target population different online panels truly are, and of how to deal with samples whose panelist profile may differ from that of the target population (e.g., using weighing; Callegaro, Lozar Manfreda, & Vehovar, 2015, 5.2), in terms not just of socio-demographics, but also of attitudinal variables. Because people volunteer to participate in opt-in panels, there is also a risk of professional respondents, that is respondents who frequently participate in surveys and are mainly doing so for incentives (Mathijsse, de Leeuw, & Hox, 2015). This could affect data quality in a number of ways (e.g., if these respondents are speeding through the questionnaire; see, e.g., Zhang & Conrad, 2014).

Probability-based and opt-in panels usually differ in several additional respects, including: a) the frequency of contact with panelists (more frequent in opt-in panels), b) panel management (sending similar surveys to all panelists versus sending completely different surveys to subgroups of panelists depending on their profile), c) the kind of survey sent (mainly academic versus mainly commercial), d) their goals (to represent the general population or to cover very specific target pop-

ulations needed by the client), etc. On the one hand, all this suggests that different recommendations might be needed for these opt-in panels. On the other hand, in the last few decades, this form of collecting web survey data has become common, in particular in market research, but also in other areas, such as social and political science. According to the AAPOR Standards Committee (2010), the majority of online research is based on non-probability panels. Thus, we believe that they require special attention.

Nevertheless, the opt-in online panels have not been studied much yet, although in a few cases methodological research is moving in that direction. For instance, Stenbjerre and Laugesen (2005) offer a summary of five years' worth of lessons learned while working with the Zapera online access panels in the Nordic region (Denmark, Sweden, Norway, Finland and Estonia). They approach the issue from several different directions, including recruitment, participation and incentives. More recently, Cape and Phillips (2015) examined the effects of questionnaire length on data quality for an opt-in online panel. Two books (Callegaro et al., 2014; Callegaro, Lozar Manfreda, & Vehovar, 2015) that include several chapters focused on non-probability-based online panels were recently published; they cover issues such as panelists' motivation for joining non-probability online panels, speeding, and professional respondents. These books include a summary of the results of the NOPVO study, the first large-scale commercial study to compare different non-probability panels in the Netherlands. This study a) made an inventory of all online panels in the Netherlands and b) compared the results across 19 online panels that conducted a similar survey (see also Vonk, Van Ossenbruggen, & Willems, 2006). Later, similar comparisons were conducted in the USA (Walker, Pettit, & Rubinson, 2009) and in Canada (Chan & Ambrose, 2011). All these studies revealed significant differences across online opt-in panels.

In addition, web surveys are increasingly completed on tablets and smartphones (Callegaro, 2010; De Bruijne & Wijnant, 2014; Revilla et al., 2016), which differ from traditional PCs in important ways: they are smaller, have touch-screens, are portable, etc. (Peytchev & Hill, 2010; Wells, Bailey, & Link, 2013). Thus, different recommendations may be needed when these mobile devices are used by at least some respondents.

In this paper, we focus on opt-in online panels, as these are increasingly common, differ from other means of collecting data on many levels, and have not yet received much academic attention. When available, our analyses also consider information on the devices panelists used to complete their surveys.

3 Methodology

3.1 Data: All Surveys Programmed by Netquest and Answered by Netquest Panelists in Spain Over a Period of About 6 Months

Our data comes from the Netquest online fieldwork company (www.netquest.com). Netquest has opt-in online panels in several countries since 2006. Netquest sends panelists survey invitations via email, using a list of individuals who have agreed to receive emails after answering a short satisfaction survey on a website belonging to one of the company's many collaborators. For each survey completed, panelists are rewarded with points, based on the estimated length of the questionnaire. These points can be exchanged for gifts. The company has panels in 23 countries. In this study, we focus on Spain, where the current panel counts 117,001 active panelists.

Our first goal was to get a good overview of what is the current situation in the Netquest panel in Spain in terms of survey characteristics, participation, and evaluation. To do this, we considered all surveys implemented by Netquest in Spain for a period of about six months (from mid-February 2016 to beginning of August 2016). We were interested in surveys that (1) were programmed by Netquest, so that we could have access to all necessary information, and (2) were sent to Netquest panelists (not to external databases provided by clients). A total of 216 surveys corresponded to this target of interest. However, we excluded 30 surveys, because of different reasons:

- Two were sent to Netquest panelists but were completed by their children; thus, they studied a different population.
- One study wanted only 15 interviews; as this was really a special case, we preferred to discard it.
- In 27 surveys, metadata was not properly collected, so we could not access necessary information on survey evaluation, devices used, etc.

In the end, 186 surveys were included in the database that we created by coding the characteristics of each individual survey. Five surveys were missing information on some aspects of interest, but the absences were minimal, so we kept them.

3.2 Aspects considered for the overview

We were interested in different aspects of each survey.

First, who is the target population? Besides the text description of each target population, we coded the following aspects:

- General or specific target population. We counted the target as the "general population" even when age limits were defined if these ages were between 16 (or 18) and 65 (or more). We also counted surveys targeting the general Internet population as the "general population". Thus, this is a quite broad definition of the "general population."
- Populations including only one gender: Surveys targeting only men or only women.
- Populations including limits on age, besides the 16+ or 18+.
- More than one target populations: For instance, surveys asking for 500 male respondents from 25 to 50, and 500 respondents who used product X at least once a week.
- We also research quotas used (if any).

***Second, what are the questionnaires' characteristics*?** In this case, we used the questionnaires to determine the main topics of the surveys as well as the main question formats used:

- Grids (also called "battery"), in which several items are presented together in a matrix format. Many studies contrast grid questions with item-by-item formats (see Tourangeau, Conrad, & Couper 2013, p. 72-76 for a summary). Even if results from the literature are mixed, many practitioners argue against the use of grids (e.g., Poynter, 2001 or Dillman, Smyth, & Christian, 2009), in particular when there are smartphone respondents (Lorch & Mitchell, 2014).
- Open-ended questions, in which respondents have to type in text as an answer. While closed questions have the advantage of being easier to analyze (they do not need to be coded) and may require less effort from respondents, open questions allow more elaborate answers. However, there are concerns that these questions might not provide all the information expected, particularly when respondents use mobile devices (e.g., Lambert & Miller, 2015).
- Multiple-response questions, in which the respondents can/must select all options they want or all options that apply. The instructions do not always explicitly state that respondents must "check all that apply," but there is no limit on the number of items the respondents can select. Previous research has usually recommended avoiding multiple-response questions and using "forced-choice" formats (e.g., asking to say yes or no for each item) instead (e.g., Smyth et al., 2006).
- Sliders, in which respondents have to position themselves on a sort of line. Again, results from previous research usually suggest that simpler alternative scales like radio buttons perform better (e.g., Funke, Reips, & Thomas, 2011).
- Dropdowns menus, in which respondents must click to make the menu appear and then select the most adequate option. Once more, there is some evidence against the use of drop-down menus (e.g., Healey, 2007).

- Ordering questions, in which respondents must rank different items from a list. Concerns have been raised about the measurement properties of ranking versus rating tasks (see, e.g., Ovadia, 2004).

We also coded if the questionnaires included some "agree/disagree" questions, that is questions asking explicitly if respondents agree or disagree with certain statements. Indeed, previous research suggests that this format creates a higher cognitive burden (Fowler, 1995) and acquiescence bias (Krosnick, 1991), as well as lower measurement quality (Saris et al., 2010).

Furthermore, we checked surveys to see if videos were present, as this can lead to more technical problems (i.e., panelists having troubles to viewing videos, in particular on smartphones).

We should note that, in general, in questionnaires programmed by Netquest, respondents cannot continue to the next question without providing an answer to the current question. Nevertheless, because of the presence of filter questions, all respondents in a given survey do not always get the same questions. We considered a format to be present if the highest proportion of respondents within a given survey got at least one question in that format. Thus, if 80% of respondents did not get a slider, and 20% did, we coded the survey as having "0 slider".

Finally, in this section we also consider the estimated length of the survey, which Netquest uses to determine the incentive respondents receive for each survey. The question of whether an ideal questionnaire length exists was already discussed in 1981 by Herzog and Bachman (p. 549). While some researchers "are convinced that survey instruments have a maximum length beyond which there is an increasing probability of premature termination, random responding, or other behavior patterns which result in data of lower quality," others "argue that a survey can be quite long without serious loss of respondents or deterioration in the quality of the responses." These authors found a tendency of somewhat lower quality answers toward the end of long questionnaires.

Third, what did respondents receive in exchange for their participation? Here, we focus on the incentives participants received for completing the entire survey. Incentives are in the form of points, which can be exchanged for gifts.

Fourth, what happened during the fieldwork? Panelists are invited to participate in a given survey. In Netquest's case, profiling information (i.e. information on different aspects of the panelists' lives, in particular behaviors and buying habits, which the panel organization has already collected and stored) is used, when available, to invite individuals who are expected to fit the target population. Once they receive the invitation, panelists can decide to start the survey (we will refer to this case as "started") or not. In the case of Netquest, panelists normally do not get any information about the survey in the invitation, so the decision to participate cannot

be linked to the survey's characteristics. Once a panelist starts the survey, different scenarios are possible:

- The panelist does not fit the population of interest or does not fit the set quotas (some quotas are already full). Thus he/she will be excluded from the survey and redirected to a profiling module. We refer to this case as "screened out."
- The panelist decides by him/herself to abandon the survey. This can occur at any moment after the panelist has started. We refer to this as "break-off."
- The panelist reaches the survey's final question. We refer to this as "complete."

We report the number of invitations, surveys started, panelists screened out, break-offs, and surveys completed across all surveys. From these numbers, we also calculated the following:

- Participation Rate = *(number started / number invited)*100*
- Screen-out Rate = *(number screened out /number started)*100*
- Break-off Rate = *[number of break-offs / (number of completes + number of break-offs)]*100*

Furthermore, for the panelists who completed the whole survey, we also considered the type of device (PC, tablet or smartphone) they used and the number of sessions in which they completed the survey (recorded automatically).

Fifth, what was the average evaluation of each survey? At the end of each survey, we added a question asking respondents to evaluate the survey, from (1) very badly done to (5) very well done. We considered the average across all respondents (PC, tablet, and smartphone) in each survey, as well as the average for PC-only and Smartphone-only respondents. We did not consider tablets separately, as they were used in a low number of cases.

3.3 How the Break-off Rate and Survey Evaluation Relate to key Questionnaire Characteristics

After our overview, we examined the relationships between some of the aspects considered. We do not study the decision to start the survey, as this is cannot be related to survey characteristics (no information is provided before the survey starts), nor did we study the screen-out rate, as this depends on the population of interest and the quotas required. On the contrary, the break-off rate is determined by panelists' decisions, and can be affected by survey characteristics such as the topic, the question format, and the estimated length. Galesic (2006) found that break-off is related to low interest (which can be linked to the topic) and higher reported burden (which can be linked to question format). Yan et al. (2010) consider the link between break-off and the interaction among the task duration announced, the real number of questions, and the presence of a progress indicator. In the book

by Tourangeau, Conrad, and Couper (2013), chapter 3.6 is dedicated to the "Factors Affecting Break-offs in Web Surveys". However, most of this research is not primarily or not at all focused on the case of online opt-in panels.

We also consider the impact of survey characteristics on respondents' evaluation of the survey. Indeed, if respondents do not like the survey, they may abandon it. In addition, if they do complete it, their satisfaction with the survey experience is expected to be lower. In this case, we expect both their probability of accepting their next survey invitation and the quality of their answers in the current survey to be reduced. This idea is supported by Cape and Phillips (2015), who found that longer surveys in an online opt-in panel do not lead to increased break-off rates, but are correlated with people speeding up during the survey, with higher satisficing, and thus lower data quality. Therefore, it is also important to study respondents' opinions of the survey.

We should note, however, that if the survey evaluation can be determined by the general characteristics of the survey, the break-off rate can only be affected by the characteristics of the questions prior to the break-off point. Nevertheless, we are not able to take this into account in our analyses, as we only possess information aggregated at the survey level. This is a key limit to these analyses. We are also limited by the fact that we possess information about the survey evaluation from the panelists who finished the survey, but not from those who did broke off or were screened out.

4 Main Results

4.1 Overview of the Current Situation in an Online Opt-in Panel

4.1.1 What is the Target Population for these Surveys?

First, we looked at the target population of the 186 surveys in our database. Table 1 presents a few examples of target population definitions. Table 2 gives the proportions of surveys that: have the general population as their target population; are limited to one gender; have age limits besides 16+ or 18+; and have more than one target of interest.

The examples in Table 1 give an idea of how specific the target populations can be in surveys run in online opt-in panels such as Netquest. It also shows how problematic getting a representative sample of such populations can be. For instance, individuals who need orthodontic work may not be aware of that fact. These hypothetical people would answer that they do not need an orthodontist, and could be screened out of surveys trying to target them. How can researchers

Table 1 Population of interest: a few examples

Definition of target population
- Men and women from 25-50 years old who play sports at least twice a week with an intensity of one hour of athletic activity and who bought detergent in the last two months.
- Population that has suffered from or is suffering from gout.
- 30-65 year-olds who consume oat, almond or rice drinks.
- Women from 25 to 45 years old who have dyed their hair blond in the last year.
- 18-50 year-olds who have bought rum in a supermarket of the brand of interest.
- Individuals between ages 25 and 65 who have cholesterol problems and consume cocoa powder.
- People who need orthodontic treatment but are not receiving it.
- Individuals who have drunk whiskey in the last three months, 25% of the brand of interest, 75% not against this brand.

Table 2 Population of interest: Proportion of surveys (of the 186 studied) with certain characteristics

Characteristic	%
Target population is the general population*	13.4
Target limited to only one gender	15.0
Target limited to some age group(s) (besides the 16+ or 18+)	52.1
More than one target of interest (within the same study)	19.3

* We count individuals between ages 16 (or 18) and 65 (or older) as the general population. For our purposes, the general Internet population is counted as the general population too.

acquire a sample of a population based on survey responses if respondents themselves do not know that they are part of the target population?

As we can see in Table 2, only 13.4% of surveys are interested in the general population, even very broadly defined (accepting age limits from 16 or 18 through 65 or older, and accepting the general Internet population). Furthermore, 15.0% of surveys are interested in only men or only women. 52.1% limit the population of interest to some age groups, besides the 16 or 18+ limit. 68.8% of surveys explicitly include a minimum age limit that spans from 8 to 55 years old. This limit is between 18 and 25 in 72.7% of cases. Besides, 47.8% of surveys include a maximum age limit that spans from 21 to 75 years old. The maximum age is 65 or older in 47.2% of cases. All this indicates that most surveys in the opt-in panel studied target

Table 3 The five most used quotas

Quotas on ...	Proportions of the surveys using these quotas (in%)
... Gender	78.5
... Age	72.6
... Geographical area	52.7
... Level of urbanization	8.7
... Social class	7.6

Note: For gender, N=158 because a quota is only possible when the population of interest includes both genders. For the others, N=186 (even for age, since even when the population is limited for some age groups, there are often still quotas within the rank of ages allowed)

very specific populations. In addition, 19.3% of the surveys define more than one target population, complicating matters even further.

Finally, most surveys also define some quotas. The goal of these quotas is usually to guarantee that the sample will be similar to the target population with respect to certain predefined variables. However, as we have just seen, target populations are often very specific. Most of the time, this means that we do not know the composition of the target population in terms of the main socio-demographics variables usually used as quotas. For instance, what is the gender or age distribution of the population of "people who need orthodontic treatment but are not receiving it"? In some cases, researchers have some ideas based on previous research. In others, quotas are used to make the sample similar to the Internet population or the whole panel, even if this does not correspond to the population of interest. It is therefore unclear the extent to which quotas are truly useful in improving the representativeness of the sample as it relates to the target population. Still, quotas were used in 95.2% of our 186 surveys. Table 3 shows the five most used quotas, with the proportions of surveys using each of them.

The most used quota is gender (78.5%), followed by age (72.6%) and geographical area (52.7%). Then come level of urbanization and social class, though their proportions are much smaller (8.7% and 7.6%, respectively). Variables such as having children, education or occupation are used in less than 3% of the surveys. We should note that these results may be strongly related to the country studied. For instance, in Latin American countries, the proportions of surveys that use quotas for social class in the Netquest panels is much higher, as habit differences across social classes are usually larger in Latin America.

4.1.2 What are the Characteristics of the Questionnaires?

After considering these surveys' targets, we researched the questionnaires' characteristics, in terms of topics, question formats and estimated length.

Survey topic. Table 4 presents the proportions of surveys dealing with various topics.

Up to 29.0% of surveys studied concern food or beverages. This is by far the most common topic. Surveys on society or politics come in second at 14.0%, whereas 11.8% of the surveys are about health, 8.6% are about insurance or banks, and 7.5% are about media, the Internet or new technologies. 71.0% of the surveys fit into one of these five categories. Some of the topics were more concrete and did not require prior knowledge (e.g., food) whereas others were more abstract and could have been affected by the respondents' level of knowledge on the topic (e.g., politics).

Question formats. Table 5 shows the proportions of surveys that made use of different question formats, from most to least common.

83.9% of the surveys include at least one multiple-response question, in which the respondents can/must select all the options they want/that apply. 76.3% of the surveys include at least one grid. Although some earlier research recommends avoiding multiple-response questions and grids (cf. Section 3.2), our study suggests that both are still very present. This is also true, in a lower proportion, for agree/disagree questions, which are present in 39.2% of the surveys. Three more formats are found quite frequently: open-response questions asking for a text answer (35.5% of the surveys), ordering questions (23.1%), and drop-down menus (18.3%). Again, this is the case despite evidence against these formats in academic literature. However, in the last case, further analyses would be needed to identify exactly which questions were asked using drop-down menus; in most cases it may only be a question on the province where the respondents live. Finally, videos are present in 7.5% of the surveys and sliders in only 2.7%. It is interesting to see that although web

Table 4 Main survey topics

Main topics	Proportions of the surveys within this topic (in%)
Food / Beverages	29.0
Society / Politics	14.0
Health	11.8
Insurance / Bank	8.6
Media / Internet / New Technologies	7.5
Others	29.1

Table 5 Proportions of surveys including different questions formats

Proportions of the surveys (in %) with at least one ...	
... Multiple-response question	83.9
... Grid	76.3
... Agree/Disagree question	39.2
... Open-text question	35.5
... Ordering question	23.1
... Drop-down menu	18.3
... Video	7.5
... Slider	2.7

surveys may allow these new features, they are not used much in practice in a panel like Netquest.

This overview shows that there is a clear gap between the academic guidelines on which question formats are best and which are actually used in online opt-in panel surveys. This suggests that the link between academic findings and their application in the practice of web survey administration must be improved. This overview also provides an indication of where further research could be useful, in order to study question formats that are often used in practice: for instance, even if forced-choice questions are recommended (Smyth et al. 2006; Revilla, 2015), more research about the evaluation of the quality of multiple-response questions could be useful, as this format continues to be used often.

Estimated survey length. In web surveys, completion time can vary greatly from respondent to respondent. Indeed, length depends on the rate at which respondents can read, process information, and answer questions; on the device used and the respondent's familiarity with the device, the speed of the Internet connection, the presence or frequency of interruptions, whether the respondent is multi-tasking or not, and so on. It also depends on the presence of filter questions. The estimated survey length (in minutes) can actually be very different from a given respondent's actual completion time. Thus, the estimated survey length gives more of an idea of the estimated complexity of the survey itself than of the experience of a given respondent. For this reason, we examine estimated survey length in this subsection on questionnaires' characteristics. Table 6 displays the minimum, maximum, average and median estimated survey length across all 186 surveys studied, as well as the proportions of surveys of different lengths.

Across all 186 surveys, the shortest had an estimated length of one minute, whereas the longest had an estimated length of 40 minutes. The average was 12

Table 6 Estimated length of the surveys in minutes

		Proportions (in %) of surveys with estimated length of ...	
Minimum across all surveys	1 minute	... 1-4 minutes	8.6
Maximum across all surveys	40 minutes	... 5-9 minutes	26.5
		... 10-14 minutes	30.3
Average for all surveys	12 minutes	... 15-19 minutes	16.8
		... 20-24 minutes	10.8
Median for all surveys	10 minutes	... 25-29 minutes	3.8
		... 30-40 minutes	3.2

minutes and the median was 10 minutes. This is much shorter than the average length across the surveys studied by Cape and Phillips (2015), which is 23 minutes in 2015. Cape and Phillips (2015) mention that the average adult attention span is around 20 minutes, and that 20 minutes is often considered the maximum questionnaire length for web surveys. This rule of thumb is actually used commonly.[2] In our study, 82.2% of surveys' have an estimated length of below 20 minutes, and only 7% have an estimated length of 25 minutes or more. Overall, the surveys in the panel we studied are quite short. We should mention, however, that the estimated length of the surveys programmed by Netquest (our focus here) is normally shorter than the estimated length of surveys sent to Netquest panelists but programmed directly by Netquest's clients. Thus, the average length would be slightly higher if we considered all surveys sent to Netquest panelists.

4.1.3 What did Respondents Get in Exchange for their Participation?

In general, the number of points respondents receive as an incentive corresponds to the estimated length in minutes, plus two. However, if the survey's estimated length is greater than 25 minutes, the incentive is increased further. Additional points are also sometimes awarded if a survey has specific requirements, such as two-wave surveys in which the researcher wants panelists to participate twice. Consequently, the correlation between estimated length and incentivization is very high, but

2 We found this rule discussed in many posts online, although some posts also discuss the pertinence of such a rule of thumb. See, for instance: http://blog.questionmark.com/how-many-questions-should-you-have-in-a-web-survey or http://researchaccess.com/2013/12/survey-length/.

Table 7 Incentives received (in number of points)

		Proportions (in %) of surveys with incentives of...	
Minimum across all surveys	4 points	... 4 points	0.5
Maximum across all surveys	58 points	... 5-9 points	25.3
		... 10-14 points	38.7
Average for all surveys	14 points	... 15-19 points	17.7
		... 20-24 points	9.7
Median for all surveys	12 points	... 25-29 points	5.4
		... 30-58 points	2.7

lower than one (around .95). Table 7 gives more information about the incentives received, in points.

Incentives span from four to 58 points, with an average of 14 and a median of 12. The highest proportion of surveys (38.7%) has an incentive between 10 and 14 points. To give some orientation on these points' value, we could mention that, for example, a panelist can acquire an e-book for 20 points, an online film for 40 points, a movie theater ticket for 120 points, or an eight gigabyte pen drive for 165 points.

4.1.4 What Happens During the Fieldwork? From Invitation to Completion

The participation process. Once the target population, questionnaire characteristics, and incentives are defined, panelists are invited to participate in the survey. In Netquest, profiling information is used to invite those panelists who are most likely to fit the target population. When these target populations are very specific, many panelists may need to be invited so that the study ultimately has enough respondents who fit the desired profile. In addition, most surveys use quotas: if the quotas are full, participants may be excluded. Finally, respondents might decide to abandon the survey, because they are experiencing some problems, because they do not like it, or simply because they have other things to do and forget to return to the survey. Table 8 provides further information on each step of the process, from invitation through completion. It also gives information about the participation, screen-out and break-off rates.

Table 8 shows major differences across surveys. The minimum number of invitations is 220, the maximum is 28,062, and the median is 2,239. Of the panelists invited, a minimum of 164 started, with a maximum of 18,019 and a median of

Table 8 Survey participation: from invitation to completion

	N	Minimum	Maximum	Average	Median
Number invited	182	220	28,062	3,437	2,239
Number started	182	164	18,019	2,131	1,450
Number screened out	185	1	14,291	1,105	466
Number of break-offs	185	2	2,261	131	49
Number of completes	186	90	5,015	875	602
Participation rate: (number of started / number of invited)*100	182	37.3	90.7	63.4	64.5
Screen-out rate: (number of screened out/number of started)*100	181	0.1	90.8	43.6	39.4
Break-off rate: [number of break-offs / (number of completes+ number of break-offs)]*100	185	1.1	62.1*	11.8*	6.7

Note: * These numbers are obtained excluding one special case: a survey where a product is sent to panelists' residences for testing. If we would include this survey, the maximum would be 88.9%, the average would be 12.2%, and the median 6.7%.

1,450. Then, 1 to 14,291 panelists were screened out, with a median of 466, and 2 to 2,261 break-offs, with a median of 49. In the end, 90 to 5,015 completed the full survey, with a median of 602.

The participation rate spans from 37.3% to 90.7%, with a median of 64.5%. Cross-survey differences are even more pronounced when considering the screen-out rate, which spans from 0.1% to 90.8%. This is related to the specificity of the target populations and to the profiling information available during sample selection. Overall, the median screen-out rate is high (39.4%) even if the company uses profiling information. This is an important problem for a fieldwork company because: a) it can affect the panelist's satisfaction and willingness to continue participating in the panel; indeed, it can be frustrating to discover you have been screened out; b) if panelists are rerouted to a profiling module, as they are in Netquest surveys, then the company must award them points even though they are screened out of the initial survey, resulting in a significant increase in costs; c) it is preferable not to send too many invitations to the same panelists; for instance, using opt-in panels data in the Nordic region, Stenbjerre and Laugesen (2005) found that six to 12 invitations per year is the frequency that leads to the best participation levels. However, if many panelists are screened out, more invitations must be sent to achieve a similar final number of completes. This can lead to panel overuse.

Finally, the break-off rate is low across the board, with a median of 6.7% and an average of 11.8% (excluding one survey where a product is sent to panelists' residences for testing). This is lower than the averages reported in many other studies: 30% for "general invitation surveys" and 15% for "individually targeted web surveys" (Galesic, 2006, p. 313). However, here again, there are huge variations across surveys (from 1.1% to 62.1%). We will try to explain these differences in Section 4.2 by examining the relationships between this break-off rate and various survey characteristics.

Number of sessions. Panelists can complete the questionnaire all at once or in several sessions, that is stopping and coming back later to continue. For respondents who completed the entire questionnaire ("completes"), the average number of sessions per survey is generally quite close to 1 (with a 1.2 average and a 1.1 median across all surveys), though it varies across surveys, from 1.0 to 2.9.

***Participation devices*.** For the "completes," we also had information on the type of device used to answer the survey. If respondents completed the survey in more than one session, we have information on only the device used in the first session. Table 9 gives the minimum, maximum, average and median proportions of PCs, tablets and smartphones used to participate across the 186 surveys.

Overall, PCs are still the main device of completion, but smartphone participation is not negligible. On average, across all surveys, 68.8% of panelists used a PC, 5.9% used a tablet and 25.2% used a smartphone. Again, there are large differences between surveys: some still have no mobile participation at all, whereas others have up to 39.7% tablet participation and 52.7% smartphone participation. It should be clear that surveys with no mobile participation at all are surveys in

Table 9 Proportion (in %) of different device types

Device	Minimum across all surveys	Maximum across all surveys	Average for all surveys	Median for all surveys
PC	40.9	100.0	68.8	67.4
Tablet	0	39.7	5.9	6.2
Smartphone	0	52.7	25.2	25.5

Table 10 Evaluation of the survey: average on a scale from (1) very badly done to (5) very well done

	Minimum across all surveys	Maximum across all surveys	Average for all surveys	Median for all surveys
All completes	3.2	4.5	4.0	4.1
PC respondents only	3.3	4.6	4.1	4.1
Smartphone respondents only*	3.1	4.5	4.0	4.0

* N=170 in the case of smartphone respondents only, as the surveys with no or very few smartphone respondents are not considered in this row.

which mobile devices were not allowed; if mobile devices were always allowed, their usage rate would be higher.

4.1.5 Evaluation of the Survey

The last step in this overview is to examine how respondents evaluated the surveys they took. Respondents who completed the entire survey were prompted to answer a final question: "Finally, what did you think about this survey? Select from 1 to 5 stars to indicate if you think the survey was (1) very badly done to (5) very well done." In this case, we also had information that allowed us to look at the evaluation for PC respondents only and for smartphone respondents only. We did not consider tablets only, because of the low rate of tablet usage. Table 10 reports the results across the 186 surveys.

The survey evaluations overall are quite positive: the median across the 186 surveys is 4.1 on a scale from 1 to 5. Moreover, variation across surveys is small, with a minimum of 3.2 and a maximum of 4.5. The results are similar when considering only PC and only smartphone respondents.

4.2 How Break-off Levels and Survey Evaluations Relate to Key Questionnaire Characteristics

In this section, we study how respondents' break-off levels and survey evaluations relate to key questionnaire characteristics, namely 1) their topic (dummies for the five main topics, "other" being the reference category), 2) the presence of different question formats (eight dummies; 1 meaning that the format is present at least once) and 3) the estimated survey length (continuous variable).

Looking at the distribution of the break-off rate across the 186 surveys shows a very skewed distribution toward the left, with two outliers on the right side (see Appendix 1a). In addition, the residuals of a simple OLS regression are not normally distributed (Appendix 1b). In order to deal with this, we use the logarithm of the break-off rate as a dependent variable.[3] This helps resolve the problem of the outliers, non-normality of the residuals (Appendix 1c), and heteroscedasticity (p=.49 for the Breusch-Pagan test with the transformed variable).

Concerning the survey evaluation, we use the average evaluation across all respondents in a given survey as a dependent variable; this takes 14 different values, ranging from 3.2 to 4.5. In this case, examination of the standardized normal probability plot suggests that residuals approximate normal distributions (Appendix 1d), so we use an OLS regression.[4] The results are presented in Table 11.

First, in the case of break-off, the whole model is significant (p<.01), and the model explains 28.2% of the variance (adjusted R^2). However, only two variables have significant effects (5% threshold): the presence of at least one video and the estimated questionnaire length.

In the survey evaluation model, the whole model is not significant (p=.15) and the explained variance is very low (Adjusted R^2 =.0308). None of the variables has a significant effect, suggesting that the survey characteristics affect the break-off rate more than the survey evaluation conducted by those panelists who completed the survey.

3 Other approaches have been tested, namely: a) a Poisson regression with robust variance and excluding the two outliers (two surveys with much higher break-off rates than all others): this led to similar conclusions; b) a negative binomial regression also excluding the two outliers: in this case, besides the estimated length and presence of at least one video (which are significant in the results presented here), the presence of at least one slider and at least one grid also had significant impacts on the break-off.

4 We also ran an ordered logistic regression, and a regression with exponential or log transformation: the conclusions remain the same.

Table 11 Regressions of log(break-off rate) and of survey evaluation on survey characteristics

		Log(break-off rate) *N=184*			Survey evaluation *N=185*		
Explanatory variables		Coef.	S.E	P-value	Coef.	S.E	P-value
Survey Main Topic	Food / Beverages	.21	.17	.21	.05	.05	.26
	Society / Politics	-.06	.21	.76	-.04	.06	.46
	Health	-.02	.22	.92	-.01	.06	.93
	Insurance / Bank	-.18	.25	.46	<.01	.07	.98
	Media / Internet / New technologies	-.11	.26	-.42	-.03	.07	.69
Format of questions	Includes one or more slider	.63	.39	.11	.09	.11	.45
	Includes one or more ordering question	-.01	.15	.95	.01	.04	.83
	Includes one or more grid	.21	.16	.20	.02	.05	.69
	Includes one or more agree/ disagree question	.19	.14	.18	-.06	.04	.16
	Includes one or more multiple-response question	-.15	.19	.44	.09	.05	.09
	Includes one or more video	.65	.25	.01	.02	.07	.83
	Includes one or more open-text question	.18	.14	.19	-.06	.04	.12
	Includes one or more drop-down menu	.03	.17	.85	-.04	.05	.43
Survey length	Estimated length	.06	.01	<.01	<.01	<.01	.68
	Constant	1.05	.24	<.01	4.02	.07	<.01
Model fit	R^2		.3374			.1045	
	Adj. R^2		.2825			.0308	

5 Discussion and Conclusions

In this study, we have reviewed all of the surveys programmed by the online fieldwork company Netquest and implemented in their panel in Spain over a period of approximately six months. By reviewing several aspects of the 186 surveys, we are able to highlight the true practice of web surveys in a current opt-in online panel. Some of the main results that we wish to emphasize are presented below.

- Survey target populations: these are often very specific, and previous information about them is hard to come by, making it difficult to use quotas that ensure the sample is similar to the population of interest.
- Survey topic: 29% of the surveys are about food or drinks. This is the most common topic. The five most common topics cover 71.0% of all surveys.
- Question format: multiple-response questions are used very frequently, as are grids. Agree/disagree questions, open-text questions, ordering questions, and drop-down menus are also used quite often. On the other hand, videos and sliders are present in less than 10% of surveys.
- Estimated length: 56.8% of surveys have an estimated length between 5 and 14 minutes.
- Incentives: on average, the incentive for answering one survey is 14 points.
- Participation process: given the specificity of target populations and the use of quotas, it is sometimes necessary to invite a huge number of panelists in order to attain a small final number of completes. However, screen-out rates vary widely across surveys. The break-off rate also varies widely, but is much smaller overall. We are able to identify some variables that seem to be related with a higher break-off, including the presence of one or more videos and a longer estimated length. Thus, we would recommend avoiding videos and keeping questionnaires as short as possible.
- Devices used: although PCs are the main device used for participation, mobile participation is clearly non-negligible.
- Survey evaluation: the survey evaluation does not vary much across surveys, and is also very similar for PC and smartphone respondents. As variations are minimal, it is not surprising that we did not find any significant effect in the regression analysis.

Our overview suggests that opt-in panels are very distinct from other web surveys, in terms of the population they attempt to cover. This also has an effect on the participation process, and in particular on the screen-out rate. Opt-in panels can differ in other respects, too. Further research focusing on these opt-in panels is needed in order to better understand the specific challenges that they face, and the best approaches to overcome those challenges. This study was limited to only one panel, in one country, and we could only analyze variables at the survey level. In

order to further study the reasons for break-off, information about what takes place just before the break-off, rather than characteristics of the survey as a whole, is needed. This study was not able to account for the device used to complete the survey in many of the analyses presented, even if it would have been very interesting, for example, to study the break-off separately for PC and smartphone respondents. Future research in these directions would be helpful.

However, even more than a need for further research, this study suggests that there is a gap between research and practice, particularly in relation to question formats. Indeed, the guidelines from academic research recommend that question formats like multiple-response questions and grids should be avoided, but our analyses in this overview reveal that they are still used very often. Academic researchers may need to work harder when sharing their results and convincing practitioners to follow their recommendations. Researchers may also need to further adapt their research so it better meets practitioners' needs. To achieve these goals, they may need to look more closely at the reality of conducting surveys in the 21st century and focus further research on how to improve the most frequently used question formats.

References

AAPOR Standards Committee (2010). *AAPOR report on online panels*. AAPOR Executive Council. doi:10.1093/poq/nfq048

Bethlehem, J. & Biffignandi, S. (2011). *Handbook of Web Surveys*. NY: Wiley. ISBN: 978-0-470-60356-7

Brace, I. (2004). *Questionnaire design: how to plan, structure and write survey material for effective market research*. London 2004.

Callegaro, M. (2010). Do You Know Which Device Your Respondent Has Used to Take Your Online Survey? *Survey Practice*.

Callegaro, M., Baker, R.P., Bethlehem, J., Göritz, A.S., Krosnick, J.A., & Lavrakas, P.J. (2014). *Online panel research. A data quality perspective*. Chichester: Wiley.

Callegaro, M., Lozar Manfreda, K., & Vehovar, V. (2015). *Web Survey Methodology*. London: Sage.

Cape, P. (2012). Understanding Respondent Motivation. DGOF White paper, available at: http://www.dgof.de/wp3/wp-content/uploads/2012/07/WP_Understanding-Respondent-Motivation.pdf

Cape, P. & Phillips, K. (2015). Questionnaire Length and Fatigue Effects: The Latest Thinking and Practical Solutions. White paper, available at: https://www.surveysampling.com/site/assets/files/1586/questionnaire-length-and-fatiigue-effects-the-latest-thinking-and-practical-solutions.pdf

Chan, P., & Ambrose, D. (2011). Canadian online panels: Similar or different? *Vue Magazine*, 16-20. Retrieved from: http://www.mktginc.com/pdf/VUE%20JanFeb%202011021.pdf.

Couper, M.P. (2000). Web surveys – A review of issues and approaches. *Public Opinion Quarterly*, 64, 464-494.

Couper, M.P. (2008). *Designing effective web surveys*. NY: Cambridge University Press.

Couper, M.P, & Miller, P.V. (2008). Introduction to the special issue. *Public Opinion Quarterly*, 72(5), 831-835 doi:10.1093/poq/nfn066

Couper, M.P., Traugott, M., & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly*, 65(2), 230-253. doi:10.1086/322199

De Bruijne, M., & Wijnant, A. (2014). Improving Response Rates and Questionnaire Design for Mobile Web Surveys. *Public Opinion Quarterly*, 78(4): 951-962.

De Leeuw, E.D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, 21(2), 233-255.

Dillman, D.A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: John Wiley & Sons.

Dillman, D.A., & Bowker, D.K. (2001). The Web questionnaire challenge to survey methodologists. In U.-D. Reips, & M. Bosnjak (Eds.), *Dimensions of Internet science* (p. 159-178). Lengerich: Pabst Science Publishers.

Dillman, D.A., Smyth, J.D., & Christian, L.M. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York: Wiley.

Fowler, F.J. (1995). Improving Survey Questions: Design and Evaluation. *Applied Social Research Methods Series*, 38, 56-57.

Funke, F., Reips, U.-D., & Thomas, R.K. (2011). Sliders for the Smart: Type of Rating Scale on the Web Interacts with Educational Level. *Social Science Computer Review*, *29*(2), 221-231. doi:10.1177/0894439310376896

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

Göritz, A.S. (2006). Incentives in web surveys: Methodological issues and a review. *International Journal of Internet Science*, 1, 58-70.

Healey, B. (2007). Drop Downs and Scroll Mice: The Effect of Response Option Format and Input Mechanism Employed on Data Quality in Web Surveys. *Social Science Computer Review*, 25(1), 111-128. doi: 10.1177/0894439306293888

Herzog, A.R & Bachman, J.G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45, 549-559.

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Lambert, A.D., & Miller, A.L. (2015). Living with Smartphones: Does Completion Device Affect Survey Responses? *Research in Higher Education,* 56, 166-177

Lorch, J., & Mitchell, N. (2014). Why You Need to Make Your Surveys Mobile Friendly Now. Webinar organized by the American Marketing Association, May 2014, https://www.ama.org/multimedia/Webcasts/Pages/why-you-need-to-make-your-surveys-mobile-friendly-now-051514.aspx?tab=home

Lozar Manfreda, K.L, Batagelj, Z., & Vehovar, V. (2002). Design of Web survey questionnaires: Three basic experiments. *Journal of Computer-Mediated Communication*, 7(3), 0, doi:10.1111/j.1083-6101.2002.tb00149.x

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schütz, 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.

Matthijsse, S.M., de Leeuw, E.D., & Hox, J.J. (2015). Internet panels, professional respondents, and data quality. *Methodology*, 11(3), 81-88 Article doi:10.1027/1614-2241/a000094

Oppenheim, A.N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Pinter Publishers.

Ovadia, S. (2004). Ratings and rankings: Reconsidering the structure of values and their measurement. *International Journal of Social Research Methodology*, 7, 403-14.

Parsons, C. (2007). Web-Based Surveys: Best Practices Based on the Research Literature. Visitor Studies, 10(1), 13-33, doi:10.1080/10645570701263404.

Peytchev, A., & Hill, C.A. (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. *Social Science Computer Review*, 28, 19-335

Poynter, R. (2001). A Guide to Best Practice in Online Quantitative Research. In A. Westlake, W. Sykes, T. Manners, & M. Rigg (Eds.), *The Challenge of the Internet* (pp. 3-19); Proceedings of the ASC International Conference on Survey Research Methods. London: Association for Survey Computing.

Revilla, M. (2015). Effect of using different labels for the scales in a web survey. *International Journal of Market Research,* 57(2), 225-238. doi:10.2501/IJMR-2014-028

Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels really need to allow and adapt surveys to mobile devices? *Internet Research*, 26(5), 1209-1227

Saris, W. E., & Gallhofer, I. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. New York: JohnWiley & Sons, Inc. (Second Edition).

Saris, W.E., Revilla, M., Krosnick, J.A., & Shaeffer, E.M. (2010). Comparing Questions with Agree/Disagree Response Options to Questions with Construct-Specific Response Options. *Survey Research Methods*, 4(1), 61-79.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, wording, and context*. New York: Academic Press.

Smyth, J.D., Dillman, D.A., Christian, L.M., & Stern, M.J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly,* 70(1): 66–77.

Stenbjerre, M., & Laugesen, J.N. (2005). Conducting Representative Online Research: A Summary of Five Years of Learnings. Paper presented at ESOMAR Worldwide Panel Research Conference in Budapest, April 17-19 2005. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.485.1158&rep=rep1&type=pdf

Sudman, S., & Bradburn, N.M. (1982). *Asking questions*. San Francisco: Jossey-Bass Inc Pub.

Tourangeau, R., Conrad, F.G., & Couper, M.P. (2013). *The Science of Web Surveys*. Oxford University Press. ISBN: 978-0-19-974704-7

Tourangeau, R., Rips, L.J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Vonk, T., Van Ossenbruggen, R., & Willems, P. (2006). The effects of panel recruitment and management on research results: A study across 19 panels. ESOMAR: *Panel Research 2006*, 79-100

Walker, R., Pettit, R., & Rubinson, J. (2009). The foundations of quality initiative: a five-part

immersion into the quality of online research. Journal of Advertising Research, 49(4), 464-485.

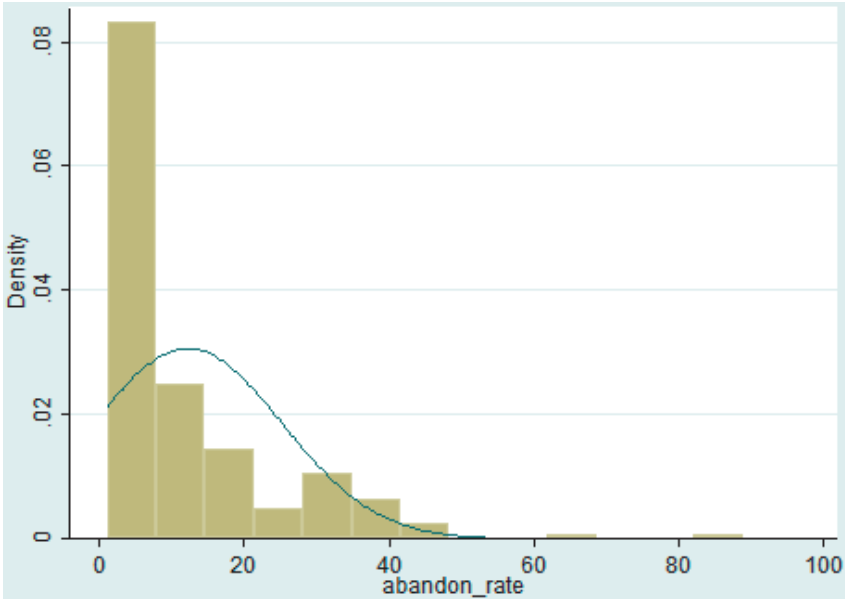
Wells, T., Bailey, J.T., & Link, M.W. (2013). Filling the void: Gaining a better understanding of tablet-based surveys. *Survey Practice*, 6

Yan, T., Conrad, F.G., Tourangeau, R., & Couper, M.P. (2010). Should I Stay or Should I Go: The Effects of Progress Feedback, Promised Task Duration, and Length of Questionnaire on Completing Web Surveys. *International Journal of Public Opinion Research*, 23(2), 131-147. doi:10.1093/ijpor/edq046.

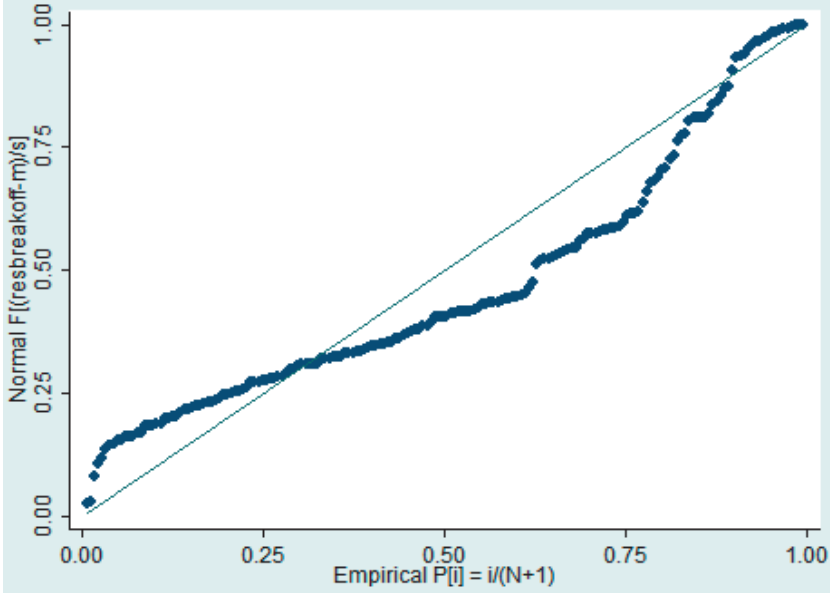
Zhang, C., & Conrad, F. (2014). Speeding in Web Surveys: The Tendency to Answer Very Fast and Its Association with Straightlining. *Survey Research Methods, 8*(2), 127-135.

Appendix 1

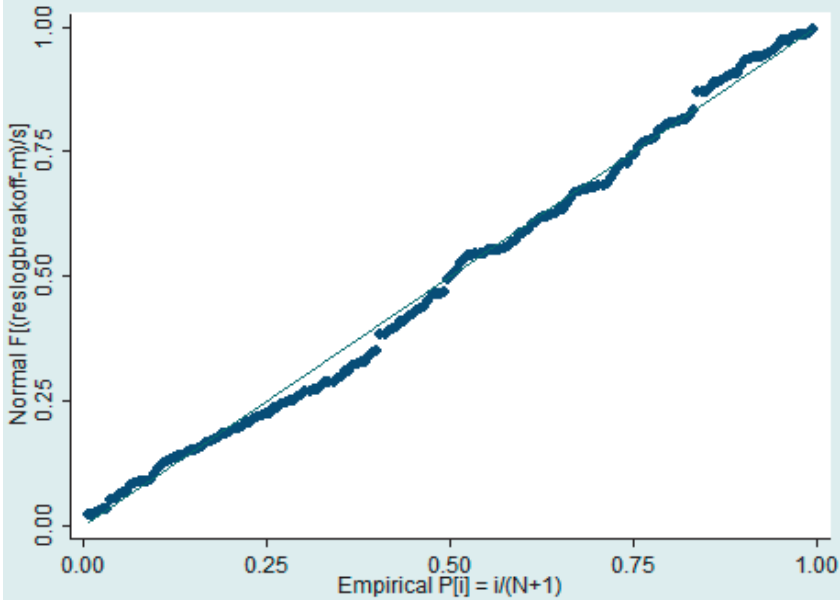
a) Histogram of the break-off rate



b) P-P plot of the residuals when using break-off rate



c) P-P plot of the residuals when using the logarithm of break-off rate



d) P-P plot of the residuals for average survey evaluation