



Universiteit
Leiden
The Netherlands

Public opinion without opinions? Item nonresponse and (the absence of) substantive opinions in public opinion surveys

Maat, J. van de

Citation

Maat, J. van de. (2019, September 17). *Public opinion without opinions? Item nonresponse and (the absence of) substantive opinions in public opinion surveys*. Retrieved from <https://hdl.handle.net/1887/76853>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/76853>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:
<http://hdl.handle.net/1887/76853>

Author: Maat, J. van de

Title: Public opinion without opinions? Item nonresponse and (the absence of) substantive opinions in public opinion surveys

Issue Date: 2019-09-17

CHAPTER 4

Data and Methods

4.1 Introduction

All three surveys were conducted via a web panel, but the panels and their composition differ (as will be detailed below). The large N of such web or internet panels made the random composition of subgroups possible to employ a between-subject-design, i.e. a design where ‘subjects in an experiment make choices in only one state of the world’ (Morton & Williams, 2010, p. 86). The advantage of a between-subject-design is that participants are not aware that they are part of an experiment, as might be the case when they are surveyed twice in a pre-test and a post-test format. The random assignment to subgroups enables the researcher to assume (and check) that the subgroups are similar; differences between the results of subgroups can be ascribed to the manipulation of the independent variable – question design.

The general instruction for all experiments read that the respondents were asked to give their opinion on issues in a survey. They were not aware of the experimental design of the survey; they did know they were participating in a survey to measure public opinion and that the results would be used for scientific research.

In this chapter, several general points are addressed regarding some methodological specifics of this study, including the experimental design and the use of internet panels, before turning to more specific aspects, i.e. the selection of issues for the questionnaire and the question design applied.

4.2 Experimental Research Designs

What are the main advantages of an experimental research design? The textbook answer is that such a design ‘engenders considerable confidence in the robustness and trustworthiness of causal findings’ (Bryman, 2012, p. 50). By manipulating one or more variables and holding other variables constant, differences in results can be ascribed to the manipulation. ‘The unique strength of experimentation is in describing the consequences attributable to deliberately varying a treatment. We call this causal description’ (Shadish, Cook, & Campbell, 2006, p. 9). Because of the researcher’s ability to manipulate variables and compare between groups, the internal validity of the findings is relatively strong (e.g. Manheim, Rich, Willnat, Brians, & Babb, 2012, pp. 103-104).

A true or full experiment is ‘a *randomized trial* in which the researcher randomly assigns units of observation to control and treatment groups’ (Druckman, Green, Kuklinski, & Lupia, 2006, p. 628). A crucial characteristic of any true experiment is random assignment. By randomly assigning subjects to either an experimental or a

control group, it can be assumed that as regards their composition the groups are similar in all aspects (Manheim et al., 2012, p. 106).

McDermott points to five major advantages of the experimental design: '1) [the] ability to derive causal inferences (...) 2) experimental control (...) 3) precise measurement (...) 4) ability to explore the details of process (...) 5) relative economy' (McDermott, 2002, pp. 38-39). Consequently, and in sync with other developments - technological and methodological innovations, an increasing interest in finding causal mechanisms, the possibility to test and refine theories and the emergence of new research questions (Arceneaux, 2010; Druckman et al., 2006; Morton & Williams, 2010) – experiments have become increasingly popular in political science. Jackson and Cox (2013, pp. 31-32) report an increase from 3 to 8.5 percent of social science articles using experiments between 1990 and 2010.

There are also some disadvantages to doing experiments: '1) artificial environment (...) 2) unrepresentative subject pools (...) 3) external validity (...) 4) experimenter bias' (McDermott, 2002, pp. 39-40). Not all disadvantages apply to each experiment; depending on the characteristics of an experiment some disadvantages may be absent. The main disadvantages are the artificiality of experiments and the limited external validity. The latter disadvantage refers to 'the generalizability of findings from a study, or the extent to which conclusions can be applied across different populations or situations' (McDermott, 2011, p. 34). The external validity is often deemed as lacking, because of the artificiality of the experimental setup (McDermott, 2011, p. 37) and the composition of the sample which is often non-random and unrepresentative of the targeted population (Jackson & Cox, 2013, p. 35). Three aspects are relevant when assessing the external validity of experimental findings: 'Support from theory or external information (...) 2) use of key characteristics of the studied individuals (...) 3) that the individuals studied are, if not a representative sample in a formal sense, at least broadly representative of the target population' (Jackson & Cox, 2013, p. 35).

Whether the external validity of an experiment is a serious concern, however, depends on the goal of the research (McDermott, 2011). If the research question essentially involves causal inference and internal validity, the experimental research design is the best choice. Internal validity is a prerequisite for drawing more general conclusions with respect to causality and herein lies the strength of an experimental design.

4.3 Survey Experiments

To examine the effects of question design in general and non-substantive response options in particular, doing a survey experiment is appropriate since a single

element can be isolated and studied. A survey experiment is an experiment in the sense that part of the data generating process is manipulated by the experimenter (Morton & Williams, 2010, pp. 30-31). The advantages of doing a survey experiment are summed up nicely by Arceneaux: '[The survey experiment design] possesses strong internal validity, as statistically significant differences in survey responses across question versions constitute strong evidence that differences in question wording are responsible for affecting people's expressed opinions. Yet because survey experiments typically draw on a broader (and sometimes representative) sample of the population of interest, they offer greater external validity than laboratory experiments, which often draw on convenience samples' (Arceneaux, 2010, p. 210). In this way survey experiments combine the best of two methods: causal inference and a realistic setting (Mutz, 2011, pp. 8-13). Survey experiments are artificial because they are set up by the researcher (Jackson & Cox, 2013, p. 43), but the general design corresponds to the way public opinion is usually gauged with polls and surveys. The internal validity of an experimental design can in this way be combined with to some extent externally valid results.

Survey experiments show causal relations. 'A survey experiment (...) is (...) a deliberate manipulation of the form or placement of items in a survey, for purposes of inferring how public opinion works in the real world. The word "experiment" (...) implies random assignment of respondents to control and treatment conditions. Comparing the decisions, judgments or behavior of the respondents in the treatment group to those in the control group reveals the causal effects under investigation' (Gaines et al., 2007, pp. 3-4). Survey experiments are becoming more popular in the social sciences because of the internal and external validity of the findings. Barabas and Jerit (2010, p. 226) warn that 'survey experiments generate effects that are observable among particular subgroups, not necessarily the entire population', but even they admit that survey experiments 'can be a valuable tool for studying public opinion' (Barabas & Jerit, 2010, pp. 226-227). If one wants to study the effect of certain question design options, as is the case here, a survey experiment is a most suitable design.

In this study, the internet survey experiment is applied. Examining the effects of question design stems from split-ballot designs, but a more elaborate design is used by, for example, using routing¹³ to guide respondents through a questionnaire when they answer filter questions in a certain way. Respondents were randomly assigned to subgroups. Differences between the outcome of the

13 Routing (in surveys) means that 'skipping and branching' occurs depending on the individual responses to certain survey questions (Caeyers, Chalmers, & De Weerd, 2012; De Leeuw, 2008). This process is relatively easy in computer-assisted and internet surveys, because after the questions are programmed correctly the interviewer or respondent is automatically redirected to the fitting next question.

surveys of the separate subgroups should therefore be attributed to the treatment variable: non-substantive response options.

4.4 Internet Surveys and Panels

Web or internet surveys have become popular because they ‘allow for simple, fast and easy access to large groups of potential respondents’ (Bethlehem & Biffignandi, 2011, p. 2), despite problems like undercoverage of the population, self-selection and nonresponse errors (Bethlehem, 2010; Bethlehem & Biffignandi, 2011; Couper, 2000). While all web surveys use the Internet for data collection, a variety of web surveys can be distinguished. For example, Couper (2000) has made a typology consisting of eight types of web surveys. Two methods are relevant for this study and both are panel-based: the non-probability based ‘volunteer panels of Internet users’ and the probability-based ‘pre-recruited panels of [the] full population’ (Couper, 2000, pp. 482-484; 488-490).

Proprietary online panels, also called online or access panels, are panels in which the respondents frequently answer survey questions on the internet (Callegaro et al., 2014, pp. 1-2). Internet panels can differ in the way respondents are recruited. The main distinction is between pre-recruitment and volunteer or convenience panels. Pre-recruitment or probability-based panels aim to include a representative sample of any specified population in the panel by recruiting them via random selection; in volunteer or convenience panels respondents register themselves via self-selection and self-registration (Stoop & Wittenberg, 2008, pp. 8-9). This distinction overlaps with Couper’s distinction (2000, p. 477) between non-probability and probability-based sampling which has consequences for the external validity of the findings.

There are several reasons why researchers work with volunteer opt-in panels. Volunteer panels usually have a large sample size; the response rates are often high; surveys are cheap to execute due to the absence of interviewers; and they can be executed very quickly (Couper, 2000; Couper & Miller, 2008; Dillman & Bowker, 2002; Stoop & Wittenberg, 2008). The main problems of volunteer or convenience panels are related to non-coverage and selection bias (Couper, 2000; Hoogendoorn & Daalman, 2009; Vonk, Ossenbruggen, & Willems, 2008).

The experiments in this study are internet survey experiments executed with panels. This has both negative and positive consequences: on the one hand the respondent may not be the person registered as panel member and the environment cannot be controlled, but on the other hand the response rate is high and interviewer effects are absent. Also, no or at least less social desirability bias occurs (Heerwegh, 2009; Kreuter et al., 2008).

4.4.1 Choice of Panels and External Validity

Three different panels were used for the three survey experiments in this study. For the first experiment the LISS panel was used, which is a pre-recruitment panel composed of a random sample of Dutch households; the panel includes 7,517 potential respondents of over 16 years old. The other two experiments were conducted with convenience or volunteer samples; the second experiment used the EenVandaag Opiniepanel (with about 45,800 potential respondents) and the third experiment the Team Vier internet panel (with about 16,000 potential respondents). The respondents of LISS and Team Vier's panel are paid for their participation, in points or in money, whereas the respondents of EenVandaag do not receive any monetary or other rewards. An overview of the three panels and their characteristics (at the point in time when the experiments were conducted) can be found in Table 4.1.

Even though in only one experiment a random sample is used which allows for generalization to the population – the LISS panel for the experiment with the DK option – any problems with external validity should not be exaggerated. In a report on online panels, the American Association for Public Opinion Research (AAPOR) stated that 'claims of "representativeness" should be avoided'. When generalizing to the population is not the goal of a study, however, '[a nonprobability online panel may be] an acceptable alternative to traditional probability-based methods' (AAPOR, 2010, p. 5). Hence why this study does not draw conclusions about 'the Dutch population' when using nonprobability online panels like the EenVandaag Opiniepanel and the Team Vier internet panel. Furthermore, the limitations of using such panels are discussed both in this chapter and in the concluding remarks in chapter 9. Being transparent about the use and limitations of nonprobability online panels strengthens the findings.

Another reason why external validity does not threaten this study is the focus is on internal validity and causal inference. All panels may suffer from selection bias, due to the recruitment of respondents and panel attrition. Fortunately, the aim of this study is to explore causal effects and not necessarily to generalize to the population, which is the strength of an experimental research design that is employed here (Arceneaux, 2010; Druckman et al., 2006; McDermott, 2002; Morton & Williams, 2010). The experimental design suggests internal validity, which is crucial for the ambition of this research project.

A final point that can be made about the limited generalisability of the findings is that the samples do correspond to the population of the respective internet panel, even if they are not representative of the Dutch population as a whole – depending on the specific panel used. This approach corresponds directly to how mass opinion polls normally would be executed. Internet panels are in practice often used to gauge public opinion and the results are more often than not presented as a representation

Table 4.1: (Internet) Panel Characteristics ^{a)}

| | LISS Panel | EenVandaag Opiniepanel | Team Vier Internet Panel |
|--|---|---|--|
| Number of respondents in panel (at the time of survey execution) | 7,517 | 45,780 | ± 16,000 |
| Sampling | Probability-based random & stratified Over 16 years old | Self-selected, convenience Over 18 years old | Self-selected, convenience Over 15 years old |
| Non-internet population | Included (via loaned equipment) | Excluded | Excluded |
| Sample representation | Comparable to high-quality RDD | Overrepresents hyper Internet users | Overrepresents hyper Internet users |
| Survey frequency | Once a month | Regularly | Regularly |
| (Financial) incentive | Cash (for each completed questionnaire) | None | Points for cash ^{b)} |
| Purpose of panel | 'Enabling researchers to benefit from existing data, to carry out their own survey or to design a special experiment' ^{c)} | 'translate the opinion of thousands of viewers directly to (among others) politics' ^{d)} | 'contribute through research to the successful corroboration of policy decisions of customers in the profit and not-for-profit sector' ^{e)} |
| Average unit response rate | 58 to 79 percent ^{f)} | 60-70 percent | n.a. |
| Unit response rate of survey experiment | 76.2 percent | 64.0 percent | n.a. ^{g)} |
| Type of issues | Empirical scientific research | Political and social issues; current affairs | A wide range, from retail to magazines, cars and governance ^{h)} |

a) The Internet Panel Characteristics table is inspired by the Knowledge Panel (<http://www.knowledgenetworks.com/knpanel/index.html>; visited on the 3rd of November 2011).

b) The respondent receives 20 points for every 10 minutes of research. After collecting 210 points, the respondent is paid 10 Euros (<http://www.teamvier.nl/nl/wie+zijn+we%3F/team+vier+panel>, visited on the 27th of May 2016).

c) (<http://www.lissdata.nl/lissdata/Home>, visited on the 27th of May 2016).

d) Original Dutch text: '*de mening van duizenden kijkers direct te vertalen naar onder meer de politiek*' (<http://opiniepanel.eenvandaag.nl/uitleg>, visited on the 27th of May 2016).

e) Original Dutch tekst: '*...door middel van onderzoek optimaal wil bijdragen aan het succesvol onderbouwen van beleidsbeslissingen van haar klanten in de profit en not-for-profit sector*'

f) (<https://www.lissdata.nl/lissdata/sites/default/files/bestanden/LISS%20panel%20statistics%202010.pdf>, visited on the 27th of May 2016).

g) The unit response rate is unknown, because the survey was closed after the target of 250 respondents completing a variant (for each subgroup) was reached.

h) The respondents can choose whether they want to participate in surveys about all issues or select some of them.

of what the general public wants, regardless of panel characteristics. In other words: this study is not a laboratory experiment where the translation of results to the 'real' world may be difficult (McDermott, 2011, pp. 34-35), but it is rather similar to how public opinion is usually gauged with internet panels.

4.5 Issue Selection

The survey questions for the experiments do not concern prognoses of the outcome of elections, but refer to substantive issues. Questions about facts and/or knowledge are not included. Surveys about substantive issues are done more often than pre-election polls and the results of such surveys are arguably more influential in the political decision-making process since they may provide very specific indications of what the public wants. The surveys deal with 'subjective phenomena' (Turner, 1981) and in line with Schuman and Presser (1996, p. 2), the focus is on 'questions dealing with attitudes, opinions, beliefs, values, preferences, and so on', i.e. 'attitude questions' in which respondents are asked to give their opinion about particular issues and respond to substantive statements. The questions in the surveys covered a range of topics to enable comparison between issues and to test whether the subject of the question matters with regard to the presence or absence of opinions, and the effect of design choices.

All three questionnaires consisted of two parts. The first part contained questions from existing long-term research like the Dutch Parliamentary Election Studies. The same eight questions were included in each questionnaire to enable comparison across experiments. The second part contained questions about current affairs, which were taken from polls and surveys at the time when the experiment was executed. Four general themes were included in the first part of the questionnaire and each theme included at least two questions from existing long-term research. The first three themes (socio-economic, ethical or moral, and multicultural) were included because these are assumed to be indicators of the main dimensions or cleavages in 21st century Dutch politics (Aarts & Thomassen, 2008; Pellikaan, 2010; Pellikaan et al., 2007). The fourth general topic is foreign affairs; this topic was selected because public opinion research suggests that the opinions on such foreign policy issues are often lacking and/or volatile (e.g. Alvarez & Brehm, 2002, p. 214; Everts, 2008, pp. 8-14). The public often does not have (enough) knowledge about foreign policy issues and a feeling of involvement may be missing since these issues usually do not affect the respondent personally; it is thus considered to be a cognitively relatively hard, technical and abstract topic.

The second part of the questionnaire included questions on current affairs

which were selected 'last minute', i.e. shortly before the survey was conducted, to make them as up to date as possible and include questions that were already part of other internet polls, for instance by Peil.nl or EenVandaag. These questions were replications. By replicating 'real'-life survey questions, the effect of manipulating question design could be compared to the original outcome to see whether a different picture of public opinion would be painted when other design choices were made.

The aim of this study was not only to compare between topics, but also between questions on the same topics. For socio-economic affairs, for example, there were differences expected between the general 'income differences' question versus the at the time hotly debated issue 'old-age pension'. The specific questions and design choices are discussed further in the Results chapters where the findings of each experiment are central.

4.6 Question Design

The main characteristic of each experiment is that between subgroups one question design element is manipulated. Everything else, including the content of the questions, the response categories offered, and the question order, is held constant. The experiments all build on the previous one(s) by adding another methodological element, i.e. a non-substantive response option, while maintaining some features of previous experiments. In the second experiment, for example, the DK option that was central in the first experiment is replicated while the filter question is added as a new element. In the third experiment the DK option and the filter question are repeated and the follow-up question is added. This approach of repeating parts of previous experiments has a number of advantages: 1) by reexamining a question design element, the findings can be validated; 2) by combining elements, e.g. the DK option and the filter question, a more detailed and nuanced analysis is possible than when only one element is analyzed in each experiment; and 3) by repeating an element in different panels, the panels can be compared.

Table 4.2 shows the way in which the three question design choices varied. These are a DK option, a filter question and a follow-up question. In the first experiment only the application of a DK option is varied, whereas in the third and final experiment all three elements are included. In each experiment a new element is added, while keeping at least some elements of the previous one(s). The explicit DK option is either offered as a 'Don't Know' or 'No Opinion' response category. These categories are replications of the original questionnaires.

Table 4.2: Question Design Characteristics

| | Experiment 1 <i>Don't Know</i> | Experiment 2 <i>Filter Question</i> | Experiment 3 <i>Follow-up Question</i> |
|---|-----------------------------------|--|---|
| DK explicitly mentioned in both question and answer | X | | |
| DK explicitly mentioned as response category | X | X | X |
| DK implicit as response category | X | X | X |
| Forced choice | X | X | X |
| Strongly worded filter question | | X | |
| Weakly worded filter question | | X | X |
| Follow-up Question | | | X |

To illustrate the differences between questions, an example of one question to which all design choices are applied is presented in Table 4.3; see Appendix A for the complete questionnaires in all question design variants.

Table 4.3: Question Design Applied to Example

| | Applied in Experiment... | Example Question |
|---|--------------------------|--|
| DK explicitly mentioned in both question and answer | 1 | Welfare benefits should be lowered in order to stimulate people to work ¹⁴ . Do you agree or disagree with this statement or don't you have an opinion? <i>Completely agree, agree, disagree, completely disagree, Don't Know.</i> |
| DK explicitly mentioned as answer category | 1, 2, 3 | Welfare benefits should be lowered in order to stimulate people to work. Do you agree or disagree with this statement? <i>Completely agree, agree, disagree, completely disagree, Don't Know.</i> |
| DK implicit as answer category | 1, 2, 3 | Welfare benefits should be lowered in order to stimulate people to work. Do you agree or disagree with this statement? <i>Completely agree, agree, disagree, completely disagree, (answer left blank)</i> |
| Forced choice | 1, 2 3 | Welfare benefits should be lowered in order to stimulate people to work. Do you agree or disagree with this statement? <i>Completely agree, agree, disagree, completely disagree.</i> |
| Strongly worded ¹⁵ filter question | 2 | [Introduction] Have you already heard or read enough about welfare benefits to have an opinion? <i>Yes, No.</i> |
| Weakly worded filter question | 2, 3 | [Introduction] Do you have an opinion on this or not? <i>Yes, No.</i> |
| Follow-up question | 3 | How upset would you be if the previously expressed opinion did not prevail when the issue was ultimately decided? <i>Very upset, Upset, A little bit upset, Not upset at all.</i> |

¹⁴ In Dutch: 'De bijstand moet verlaagd worden zodat mensen gestimuleerd worden om te werken'.

¹⁵ The distinction between the strongly worded filter question and the weakly worded filter question is based on previous research by Bishop (Bishop, 2005, pp. 22-23; Bishop et al., 1983, pp. 530-535), which is discussed more extensively in the chapter about filter questions.