Comparing Different Types of Web Surveys: Examining Drop-Outs, Non-Response and Social Desirability

Henning Silber[1], Julia Lischewski[2] and Jürgen Leibold[3]

Abstract

This paper aims to compare different types of web surveys in terms of response behaviour and data quality. To do so, the data of four online samples, two online access panels, a student sample, and a generated mail sample – randomly drawn from a systemically generated pool of email addresses – were contrasted. To investigate expected sample differences in drop-out rates, non-response, and data quality, closed and open-ended questions of varying levels of sensitiveness were employed. The main findings were that the two access panels lead to lower item non-response, but especially when sensitive questions were asked, data quality problems were revealed. Moreover, the access panelists showed a tendency to take short-cuts in the response process and to edit their answers in favour of social desirability.

1   Introduction

The number of online surveys has gradually exceeded mail surveys and computer assisted personal interviews, while even the commonly employed telephone surveys method of the past has, in the meantime, been less widely used than web surveys (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V., 2010:12; Bethlehem & Biffignandi, 2012:XI). Besides the ever increasing amount of people with internet access, the main reasons for this ascent are the lower costs and the comparative easiness of employing these kinds of surveys (Couper, 2000:476; Groves, 2004:502). But do survey researchers who employ web surveys do this at the expense of the quality of their data?

A large amount of studies on different survey types have addressed this issue and have applied mode comparisons by mainly contrasting online with various other modes (Bowling, 2005; Denscombe, 2009; Dillmann et al., 2009; Fricker et al., 2005; Kaplowitz, Hadlock & Levine, 2004; Kaplowitz et al., 2012; Kreuter, Presser & Tourangeau, 2008;

---

[1] Henning Silber, PhD Candidate, Center of Methods in Social Sciences, Göttingen University, 37073 Göttingen and Visiting Student Reseacher, Stanford University, Department of Communication, Stanford, CA 94305-2050; hsilber@stanford.edu

[2] Julia Lischewski, PhD Candidate, Center of Methods in Social Sciences, Göttingen University, 37073 Göttingen; julia.lischewski@sowi.uni-goettingen.de

[3] Jürgen Leibold, Postdoctoral Researcher, Center of Methods in Social Sciences, Göttingen University, 37073 Göttingen; juergen.leibold@sowi.uni-goettingen.de

Malhotra & Krosnick, 2007; de Leeuw, 2005; Shin, Johnson & Rao, 2012). While it was highlighted that telephone surveys are cost-intensive and need the facilities of a call centre, web surveys do not need to overcome this high barrier. On the other hand, it was shown that web surveys have serious drawbacks in representativeness. Notwithstanding the obvious advantages and disadvantages of web surveys, differences between the survey modes in terms of response behaviour to question formats, question order and scale effects are not that effortless to investigate. Contrasting web and mail survey modes, Shin, Johnson & Rao, (2012:228) concluded that mail surveys may entail higher unit non-response rates, whereas the amount of item non-response might be lower within web surveys. In addition, Knapp & Heidingsfelder (1999:3) showed considerable impacts of survey design on drop-out rates by comparing different web-surveys.

Besides all these findings, it would be too simplistic to divide the world of surveys exclusively in self-administered and externally administered or online and non-online surveys. Every survey mode, and especially the online mode itself, frequently employs an increasing number of subtypes (Bethlehem & Biffignandi, 2012:38; Couper & Coutts, 2006:228; Couper & Miller, 2008:831; Fricker, 2008:202; Koo & Skinner, 2005). Contrasting online survey subtypes, Yeager and colleagues showed in their groundbreaking article about the effects of different sampling strategies the impact of probability and non-probability sampling on accuracy (Yeager et al., 2011).

Academic online research polls various populations of respondents such as students, members of access panels or specific email samples. By using the rational choice approach (RC) as a heuristic, we assume that respondents will compare the benefits and the costs of taking part in a survey and similarly applying it while answering every single question (Dillman, 1978:12; Esser, 1986:38). Consequently, we start from the premise that different samples are connected with typical patterns of cost-effectiveness considerations. In line with this idea, non-professional respondents[4] should have a higher interest in the stated research aim than those hired with incentives for an online access panel. In contrast, the participants of such panels should be more interested in the fulfilment of their obligations. Considering that, on every stage of the answering process these costs and benefits of responding are taken into account by the respondents, our main research question is: What are the consequences of different costs-effectiveness considerations with respect to participation, item non-response and the quality of the answers. Each of these characteristics is closely related with the accuracy and the relevance of the retrieved information. In the following, we introduce our research hypotheses by focusing on participation (response rates and drop-outs), item non-response and the quality of answers (cognitive effort and social desirability). Accordingly to this, we outline our hypotheses, describe our research concept and present our findings in detail afterwards.

---

[4] We consider the participants of access panels as professionals and the volunteers of the other samples as non-professionals (Gittleman & Tirmarchie, 2009:2; Sparrow, 2007:180; Toepoel, Das & van Soest, 2008:986; 2009).

2   Hypotheses

Generally, we expect a distinction between "non-professional respondents" and "professional respondents" caused by the varying motivations for responding to the surveys.

Response rate

With regard to incentivisation and daily routine we assume a higher willingness to follow the invitation and to start answering the questionnaire by the professional respondents. The non-professionals visiting the first page may have more interest in the subject. Alternatively, professionals might have less interest, they, nevertheless, have greater motivation to fulfil their obligations (Görtiz 2004; 2006; Holbrook, Krosnick & Pfent 2007).

Additionally, it is likely that technical problems with the questionnaire appear rarely on the terminals of professionals, and therefore, it is often an easier task for the professional respondents to answer surveys than for the volunteers (Dillmann et al. 2009; Holbrook, Krosnick & Pfent 2007). Consequently, we expect a significant impact on the response rates.

Drop-outs

Secondly, we hypothesise differences in drop-outs between the surveys with professional and non-professional respondents. The combination of the given incentives and the panel conditioning, which arises especially for the professional sample, could generate a work-like professional relationship to the task of responding to the questionnaire. Because, the members of the access panels receive incentives only for answering the entire questionnaire, we expect a lower drop-out rate (Birnholtz et al. 2004; Bosjak & Tuten 2003; Göritz 2004; 2006; Hansen & Pedersen 2012; Herrwegh 2006; Sandez-Fernandez 2010). These effects should be particularly visible at the beginning of the questionnaire. We suppose that the participants of the non-professional samples should show, especially at this point, higher drop-outs, because of the insufficient financial motivation. However, the people from the access samples should not discontinue filling in the form, even if they have no relation to the topic (Frick, Bächtiger & Reips, 2001:195; Knapp & Heidingsfelder, 1999:6; Roose, Waege & Agneessens, 2003:416; Roose, Lievens & Waege, 2007:424).

Non-response

Moreover, the work-like relationship could cause the risk of providing incorrect answers to fit into the target sample and the motivation for answering the questionnaire in a shorter period of time occasionally scanning the given instructions (Göritz, 2007:481; Petric, Appel & de Leeuw, 2009:466; Tourangeau, 2007:189). Therefore, we also expect differences in the item non-response between the professional and non-professional samples. The following hypotheses are based on two main assumptions. First, we hypothesize less item non-response in the access data for closed-ended as well as for open-ended questions, because of the professional relationship. Second, the quantity of answers to open-ended questions should be higher for the professionals.

Besides, there is an on-going discussion about specifics of online access panelists in response behaviour towards different questions forms (Göritz, 2007; Dennis, 2001). While, little evidence was found for panel conditioning related to attitude questions, questions concerning the knowledge

of the respondents revealed remarkable effects (Das, Toepoel & van Soest, 2007; Kruse et al., 2009; Toepoel, Das & can Soest, 2009). Nevertheless, based on the RC approach, we expect panel conditioning effects for the target attitude questions.

Quality of answers
A possibility to test the differences concerning the response pattern of closed-ended questions is to check the distortion created by social desirability. Numerous studies have analysed the effect of sensitive questions for different modes. In comparison to CATI, PAPI and face-to-face interviews, most of these studies have established that web surveys show less social desirability bias (Chang & Krosnick, 2009:674; Kreuter, Presser & Tourangeau, 2008:861; de Leeuw, 2005:246). The determining argument, for the missing social desirability bias in web surveys is the absence of social presence (Taddicken, 2009:100). Due to this condition of more anonymity, the respondents provided responses with less social desirability.

Also various online survey subtypes can create differences in the perception of social presence and anonymity. The participants of the non-professional samples were invited to the survey via personal email. With regard to the recruitment method, it was considered that this personal email announcement by a university would create a form of social presence (Heerwegh, 2005:596; Heerwegh & Losseveldt, 2007:265). Additionally, the non-professional respondents may have more interest in the topic than the professional respondents (Groves, Presser & Dipko, 2004:26; Martin, 1994; Schnell, 1997:185).

When measuring social desirability, scales such as the Marlow-Crowne Scale (Crowne & Marlow, 1960; German version: Stöber, 1999) and the Balanced Inventory of Desirability Response (Paulhus, 1991; 2002) are used the most frequently. Besides these classical social desirability scales, it is also possible to measure the specific motivation to control prejudiced reaction (Dunton & Fazio, 1997; Banse & Gawronski, 2003). Another approach suggested by Bassili (1996) is the measurement of response latency. According to Mayerl (2005:1), the response latency could be "used to measure the chronic accessibility of attitudes".

In our context, the theory of response latency demands a longer response time from the fact that socially desired answers necessitate additional mental expenditure (Walczyk et al., 2009:35). According to this, we hypothesise that the professionals should express more prejudice in their answers, because of the lack of social presence, and their lesser interest in the topic. If our assumptions about cognitive effort prove true, we should also find a higher motivation to report unprejudiced behaviour in the non-professional samples.

Additionally, we will examine the quality of the open-ended answers. We expect them to be of a higher quality for the surveys answered by non-professional respondents, because of their greater interest in the topic based on self-selection (Groves, Presser & Dipko, 2004:26; Holland & Christian, 2009:198; Martin, 1994; Schnell, 1997:185).

3   Research concept

To test the postulated assumptions empirically, we employed data from different respondents groups surveyed with identical questionnaires. The subject of the questionnaires was "Multiculturalism in

Germany". In the academic research convenience samples of open online surveys with volunteers are typical; the most common are samples from students. Traditionally undergraduates are frequently asked to test new instruments for academic use, especially in social sciences and psychology (Druckmann & Kam, 2011:54; Flere & Lavric, 2008:399; Peterson, 2001:450; Wiecko, 2010:1186). Furthermore, in recent years the use of online access panels has become more popular among academic, and especially, market researchers. These non-probability web samples are selected via stratifications from panels of thousands of professionals, who do not fit the probability samples of any population. Commonly, quotas are used to create the samples of interest.

Our aim is to have a closer look at professional and non-professional web samples. Until now, little effort has been devoted to a systematic comparison across various web survey types in terms of unit and item non-response towards closed and open-ended questions and the effect of social desirability. Therefore, we combined previous findings on non-response (Biemer & Lyberg, 2003; Bosnjak & Tuten, 2001; Groves et al., 2002; Groves, 2004; Groves & Couper, 1998; Stoop, 2004; 2007) and sensitive questions (Lee, 1993; Lensvelt-Mulders, 2008; Reja et al., 2003; Tourangeau & Smith, 1996; Tourangeau & Yan, 2007) with the examination of different types of web surveys. For testing our hypotheses, we selected four samples, a student sample[5], and two online access panels of different professional providers. Moreover, the research includes another sample drawn randomly from a systematically generated pool of email addresses. Purpose of this generated mail sample was to test the assumptions on a second non-professional group without the typical drawbacks of homogeneous subjects (Druckman & Kam, 2011; Peterson, 2001).

4   Samples

Our four samples were collected between February and May 2011. A summary of the differences between the data collection methods appears in Table 1. For the first sample we invited 1066 individuals from a list of graduates and undergraduates of the social sciences. About 96% of those mails reached an email account. A total number of 139 individuals visited the first page a few days later and 73.4% of them completed the questionnaire. Using the panel partner programme of Globalpark, we contacted the members of two different online access panels[6]. In both, nearly 10% of the respondents completed the questionnaire in the short time period of 8 and 4 days respectively. The last sample was drawn from a list of automatically generated email addresses to ensure a wider spread in comparison to the student sample.[7] In 30% of the cases, the email could not be delivered to the recipient because of email errors. A closer look at the table reveals the inefficiency of the recruitment strategy of the third sample, indicated by the lowest contact rate (4.3%), the highest

---

[5]   Students in social sciences might be already more sensitised to the subject of our letter of announcement than respondents of the other samples.

[6]   The specification was a census-representative sample with respect to gender and region of Germans between 16 and 69 years.

[7]   We used 250 typical German forenames, the 100 commonest German surnames and four different email suffixes to generate a selection base with 357,084 email addresses. We selected 4,629 cases of this list at random and invited the owner personally.

drop-outs (28.1%) and, in particular, the long field time (80 days) for conducting this small sample (n=100).[8]

The marked differences between the socio-demographic distributions of the four samples reflect the various sampling strategies. However, there are also some similarities. In detail, the highly educated people are over-represented, the proportion of women is the highest in the student sample (54.3%) and females are slightly under-represented in the three web based samples. Moreover, with an average of 41 years, the participants of both access panels are the oldest respondent groups; as expected, the students are the youngest (25.2) and the average of those of the generated mail sample is 37.4 years.

Table 1: Sample description

	student sample[1]	access sample 1	access sample 2	GM sample
invitations	1066/10262	4500	4000	4629/32342
contact 1. page	139	305	225	139
invitations/ responses rate	0.135	0.068	0.056	0.043
completed	102	286	202	100
responses/ completation rate	0.734	0.938	0.898	0.719
fieldtime in days	15	4	8	80
incentivisation	no	yes	yes	no
age (mean)	25.2	41.1	41.2	37.4
proportion of women	0.543	0.497	0.488	0.490
proportion of qualification for university entrance	1	0.514	0.495	0.779

[1] Undergraduates and graduates
[2] Number of email addresses without mail delivery failure and returning message to sender.

Below we present our findings divided into three parts: drop-outs, non-response and answer quality towards open and closed-ended questions. Finally, we will discuss the implications for further web survey research.

5   Results

The analysis is presented in three parts.[9] First, we contrast the survey drop-out rates of the four web samples. Next, we compare quantitative and qualitative aspects of response behaviour towards open and closed-ended questions separately. Besides exploring item non-response, our set of analyses

---

[8]   Respondents were invited only once to participate in the surveys; reminder emails were not sent out to any of the four samples.
[9] The statistical analyses reported in this chapter were conducted with the statistical software packages SPSS Version 20.

investigates the qualitative aspect of data quality by examining whether the type of web survey has any influence on the quality of the answers in respect to different question formats and levels of sensitiveness. Most importantly, we are interested in the effects of different forms of social presence, response experience, topic interest and incentivisation in response behaviour.

Drop-outs

In web surveys, researchers have to cope with the same difficulties as in other surveys that a systematic drop-out rate will lead to biased results because of selective samples. Thus, the question arises whether the different data collection methods of our four samples have also caused differences in their drop-out rate. Further, which consequences arise from these drop-outs concerning the quality of answers? Examining the data, we expected - in line with earlier findings - fewer drop-outs from the professional respondents, as a result of the given incentives and the panel conditioning (Frick, Bächtiger & Reips, 2001:198; Heberlein & Baumgartner, 1978:455). Figure 1 shows the partial respondent drop-outs during the interview. Whereas the drop-out rate for the first access sample amounts to 10 and the second access sample comes to 6 percent, the non-professionals had significantly higher drop-out rates (student sample 27% and generated mail sample 28%).
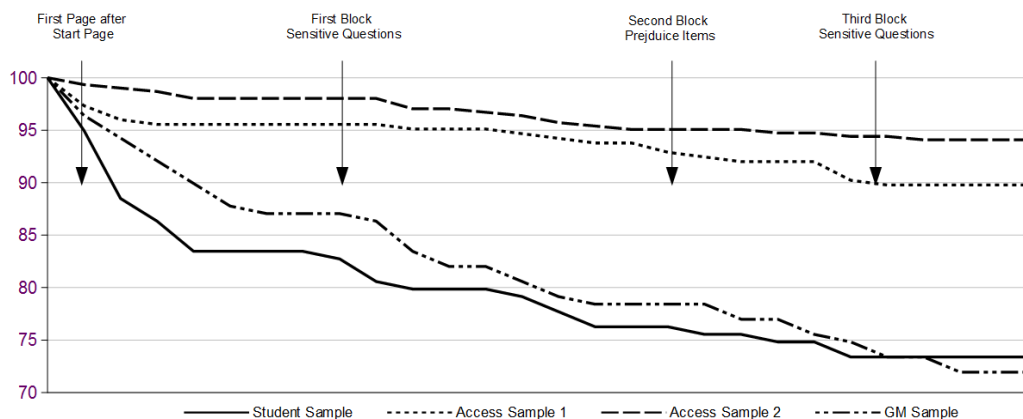

Figure 1: Drop-outs during the interview by sample (percent)

Only 3% of the respondents of the first and 1% of the second access panel abandoned the survey after the first page. In contrast, the breaking off rate for the first page was higher by the student sample (5%) and the generated mail sample (4%). Both samples also showed a similar pattern in relation to the content of the questions. We found the highest endurance in the professional access surveys. One reason for the lower drop-outs could be the work-like professional relationship, especially the effect of incentivisation. This could implicate that in the access survey, respondents, who have lower drop-out rates, answered the questionnaire without regard to the subject (this assumption is strengthened by the varying drop-out rates on the starting page). Moreover, if they were more likely to be disinterested, what could be the consequences for the quality of answers? Following this question, the investigated indicators for the quality of the answers are item non-response, response latency and response behaviour.

Open-ended questions

Non-response

Examining item non-response within open-ended questions, Table 2 and 3 show the number of answers to questions about stereotypes towards Muslims and towards Jews.[10][11] While merely around 25 percent of the respondents of the student sample and the generated mail sample answered the open-ended questions, twice as many of both access panel samples provided answers.

Table 2: Answers to open questions to stereotypes towards Jews by samples (in percent)

	student sample	access sample 1	access sample 2	GM sample
response	30	50	51	26
no response	70	50	49	74
total	100	100	100	100
	(N=102)	(N=287)	(N=202)	(N=135)

See appendix Table 2a for significance tests.

Table 3: Answers to open questions to stereotypes towards Muslims by samples (in percent)

	student sample	access sample 1	access sample 2	GM sample
response	25	49	52	21
no response	75	51	48	79
total	100	100	100	100
	(N=132)	(N=303)	(N=220)	(N=135)

See appendix Table 3a for significance tests.

Altogether, the number of responses towards open-ended questions on prejudice was much higher within the professional respondents of the access panels one and two. Besides that, the student sample and the email sample showed relatively similar response patterns, while the two access panels revealed very different results. These findings lead us back to our initial question: Does any survey type provide answers of higher data quality? Thus, we next examine whether the shown lower percentage of item non-response towards open-ended questions of the two access samples will have any negative impact on the quality of the given answers.

Quality of answers

Our former analyses of item non-response indicated that the access panelists answered the open-ended questions more often. In the following, we want to investigate the quality of those responses. Therefore, we divided the given answers in meaningful and meaningless responses. Meaningless

---

[10]   "Thinking about Jews, which positive or negative facts come to you mind?"
       "Thinking about Muslims, which positive or negative facts come to you mind?"
[11] Prejudice items used in this article addressed negative opinions toward Jews and Muslims, two different minorities. With regard to the concepts of latent communication and latent anti-Semitism, we expect a stronger connection of social desirability for prejudice statements toward Jews (Bergmann & Erb, 1986; Salzborn & Schwietring, 2003). The fundamental idea is, that while a large potential for anti-Semitism still lingers in Germany, the German social elite continues to suppress the public expression of anti-Semitic attitudes. Therefore, the expression of prejudices towards Jews is more taboo for Germans than negative attitudes toward Muslims.

refs to an answer which is hardly related to the requested answer.[12] Afterwards, we compare the length of the meaningful answers within the four web samples and exclude the meaningless.

Table 4: Substantive answers to open questions to stereotypes towards Jews by samples (in percent)

	student sample	access sample 1	access sample 2	GM sample
meaningless answer	3	39	46	23
substantive answer	97	61	54	77
total	100	100	100	100
	(N=31)	(N=145)	(N=104)	(N=26)

See appendix Table 4a for significance tests.

Table 5: Substantive answers to open questions to stereotypes towards Muslims by samples (in percent)

	student sample	access sample 1	access sample 2	GM sample
meaningless answer	4	34	31	5
substantive answer	96	66	69	95
total	100	100	100	100
	(N=25)	(N=124)	(N=107)	(N=21)

See appendix Table 5a for significance tests.

With regard to the quality of the responses towards the two open-ended questions on stereotypes, the Tables 4 and 5 show that the relative number of meaningful answers were much higher in the student and the generated mail sample. In detail, these samples provided as many meaningful answers for the question regarding Muslims, whereas on examining the questions about Jews alone, the percentage of meaningful answers was seen to be about 20 percent higher among the students.

An explanation for the differences in response quality might be a tendency of the access panelists to take short cuts in the response process. For instance, typing something meaningless into open-ended questions, which had hardly any connection to the desired answer. Throughout all survey types, the afore-mentioned differences in the open-ended questions regarding Muslims and Jews were probably a result of the extraordinary sensitiveness of Jewish topics in Germany.

---

[12] We rated responses as meaningless if they had no relation at all to the requested answer. For instance, some respondents answered the open questions only with numbers, random characters or smilies. Consequently, we rated every ambiguous statement with reference to the respective minority as meaningful.

Table 6: Length of response to open question towards Jewish stereotypes (mean: number of characters)

	student sample	access sample 1	access sample 2	GM sample
total	50.38	32.18	24.71	35.33
	(N=102)	(N=287)	(N=202)	(N=100)
without n.a.*	165.77	63.7	47.99	135.88
	(N=31)	(N=145)	(N=104)	(N=28)
without n.a. and	171.13	99.89	84.02	174.75
meaningless**	(N=30)	(N=89)	(N=56)	(N=20)

*Length without nonresponse
**Length only taking substantive answers into account

Table 7: Length of response to open question towards Muslim stereotypes (mean: number of characters)

	student sample	access sample 1	access sample 2	GM sample
Total	48.71	53.92	45.98	27.81
	(N=102)	(N=287)	(N=202)	(N=100)
without n.a.*	198.72	109.25	86.92	132.43
	(N=25)	(N=140)	(N=105)	(N=21)
without n.a. and	206.96	165.62	126.13	138.75
meaningless**	(N=24)	(N=92)	(N=72)	(N=20)

*Length without nonresponse
**Length only taking substantive answers into account

Moreover, with regard to the length of responses indicated in Tables 6 and 7, there were only small differences amongst the four web samples in the total amount of characters. By limiting to the meaningful responses only, the answers of the students and the respondents of the generated mail samples were significantly longer. Furthermore, the examination of the sensitive questions about Muslims showed that the length of responses of the non-professional samples were almost equal, whereas the length of responses of the graduates and undergraduates were significantly longer when answering the questions about Jews.

To sum up, while the access panelists typed more often characters, the response quality of non-professional respondents seemed to be higher. These findings suggest that, when the non-professionals answered the open-ended questions, they had taken the given task seriously and had answered meaningfully and extensively, while the access panelists had frequently been trying to abbreviate the response process by answering shorter and more meaninglessly. Furthermore, the students and the respondents of the generated mail sample showed roughly similar response behaviour towards open-ended questions.

Closed-ended questions

Non-response

To investigate the non-response, we selected three closed-ended blocks of questions (see Figure 1). For the purpose of controlling question placement effects, the first block was placed at the beginning of the questionnaire, the second in the middle and the third at the end. Moreover, to

compare differences within question sensitivity, we used prejudice items towards Jews and Muslims for the second block.

As the means in Table 8 indicate, in every single item block the respondents of access sample one and two decided more often on selecting item non-response options, whereas the respondents of the student sample and the generated mail sample selected this option less often. The respondents of access sample two showed the highest rate of non-response in the second item block with an average 1.24, whereas the 102 students answered nearly all the items (0.11) in the last block.

Table 8: Item missings[1] by sample for three items blocks[2]

First block (at the beginning of the questionnaire)

	mean of item missings[3]	number of missings (Percent)				
		N	0	1	2	3+
student sample	0.44	102	68.6	20.6	8.8	2.0
access sample 1	0.51	286	74.8	16.1	4.9	4.2
access sample 2	0.51	202	79.2	10.4	3.5	6.9
GM sample	0.41	100	79.0	13.0	3.0	5.0

Second block (prejudice items in the middle of the questionnaire)

	mean of item missings[4]	number of missings (Percent)				
		N	0	1	2	3+
student sample	0.39	102	77.5	10.8	6.9	4.8
access sample 1	1.17	286	62.9	13.3	6.3	17.5
access sample 2	1.24	202	58.9	13.4	8.9	18.8
GM sample	0.39	100	81.0	11.0	3.0	5.0

Third block (at the end of the questionnaire)

	mean of item missings[5]	number of missings (Percent)				
		N	0	1	2	3+
student sample	0.11	102	93.1	3.9	2.0	1.0
access sample 1	0.30	286	92.3	3.1	1.0	0.0
access sample 2	0.36	202	89.1	5.9	0.5	4.5
GM sample	0.27	100	88.0	7.0	3.0	2.0

1) Refusals and don´t know answers for seven items
2) The placement of the blocks in the questionnaire is shown in Figure 1.
3) No significant differences between the means.
4) The mean of the student and GM sample is significant lower than the two access samples.
5) The mean of the student sample is significant lower than the two access samples.

Moreover, with regard to the number of item non-responses in the second item block with prejudice items, professionals and non-professionals showed different response patterns, whereas

the amount of respondents with three or more non-responses was higher within the access panels. This is also reflected by the means.

The findings point out that in the case of closed-ended questions, the number of non-responses could have depended on sensitiveness. Particularly, the professional respondents seemed to have shown a tendency to select the non-response option when closed-ended sensitive questions were asked.

Quality of answers

Besides this complexity of item non-response, it is possible that the answers to closed-ended prejudice items are affected by social desirability. In this context, we examined the relationship of social desirability scales, response latency and prejudice against Muslims[13] and Jews[14] by sample. As indicated in Table 9, the respondents of the access panels frequently expressed significantly more negative attitudes against Jews as well as against Muslims, while the non-professional respondents showed a less prejudiced response pattern towards these religious groups.

Table 9: Prejudice* towards Jews and Muslims by sample (mean and standard deviation)

	prejudice against Jews			prejudice against Muslims		
	mean**	stand. deviation	N	mean***	stand. deviation	N
student sample	1.284	0.441	108	1.678	0.678	105
access sample 1	1.754	0.678	258	2.377	0.743	264
access sample 2	1.780	0.690	187	2.475	0.818	193
GM sample	1.433	0.598	107	2.033	0.890	107

*   Values range from one (no prejudices) to four (high prejudices)

**   The mean of the student sample is significant smaller compared to the means of the two access samples.

*** The means of the student and the GM sample are significant different compared to all other samples.

We expected that the answers to sensitive questions would be influenced by social desirability. On the other hand, a typical characteristic of web surveys is the lack of social presence, which could diminish the influence of social desirability. Our assumption was that the personal email announcement could invoke such social pressure for the non-professionals with effects on their perceived desirability. We used two social desirability measures, firstly a shortened version of the social desirability scale (GMC) by Stöber (1999) and secondly three items of the motivation to

---

[13] "Islamic and west-European values can be agreed with each other."

"Muslim culture fits well into our Western world."

"Immigration to Germany should be prohibited for Muslims."

"With so many Muslims living here in Germany, I sometimes feel like a stranger in my own country."

An index based on the individual mean score was computed for these items.

[14] "Judaism fits well to Europe."

"Jews have too much influence in Germany."

"As a result of their behaviour, Jews are not entirely without blame for being persecuted."

An index based on the individual mean score was computed for these items.

control prejudiced reactions scale (MUB) (Banse & Gawronski, 2003)[15]. In addition, we employed the response latency as a possible indicator for the measurement of social desirable response sets (Walczyk et al., 2005).[16] In the first step, we tested the assumed relationship of response latency and social desirability. Table 10 shows the correlation between response latency and the two social desirability scales for the block with the seven prejudice questions.

Table 10: Correlation of SD-scales with response latency for seven prejudice questions towards Jews and Muslims by sample[1]

		GMC	MUB
student sample	coefficient	0,221	0,147
	p-value	0,041	0,177
access sample 1	coefficient	0,120	0,168
	p-value	0,055	0,007
access sample 2	coefficient	0,159	0,146
	p-value	0,038	0,057
GM sample	coefficient	0,067	-0,103
	p-value	0,542	0,384

1) Under control of age, sex and education

The correlation analysis showed significant effects solely for the professional respondents. With decreasing GMC, it took the respondents of the first access sample less time to answer the attitude questions. We also found a significant positive effect for the scale of motivation to unprejudiced behaviour. Respondents who wanted to behave in an unprejudiced manner needed more time to answer the questions.

In a second step, we investigated the impact of these social desirability measurements on anti-Semitic and islamophobic attitudes. The results are shown in Table 11 and 12.[17]

---

[15] "I pay attention to the fact that my behaviour is free from prejudices." (MUB)
  "Going through life worrying about whether you might offend someone is just more trouble than it´s worth." (MUB)
  "I don't care if somebody thinks that I had prejudices towards minorities." (MUB)
  "I have kept before borrowed things." (GMC)
  "Sometimes I help only because I expect a consideration." (GMC)
  "Sometimes I throw garbage in the landscape or on the street." (GMC)
  An index based on the individual mean score was computed for the MUB- and GMC-items.
[16] Fazio suggests a correction of the response time with a logarithmic function for asymmetrical distributions. This form of transformation is frequently used in research on response time (Fazio, 1990; Huckfeldt et al., 1999; Devine et al., 2002; Kreuter, 2002; Mayerl & Urban, 2008) and is used in the following analysis.
[17] The coefficient of the control variables, age, sex and education were not significant.

Table 11: Linear regression (OLS) prejudices against Jews on SD scales and response latency by sample

		model 1	model 2[1]
		beta	beta
student sample	(constant)	0.894	1.651
	response latency	0.162	0.202
	GMC	-0.097	-0.023
	MUB	-0.299 **	-0.188 **
	R2	0.083	0.111
	N	94	89
access sample 1	(Constant)	3.701 ***	3.266 ***
	Response latency	-0.161 *	-0.199 *
	GMC	-0.121 +	-0.042
	MUB	-0.167 **	-0.165 *
	R2	0.080	0.099
	N	243	240
access sample 2	(Constant)	2.364 **	2.154 **
	Response latency	0.044	0.098
	GMC	-0.128 +	-0.072 +
	MUB	-0.154 *	-0.218 *
	R2	0.029	0.030
	N	170	161
GM sample	(Constant)	3.535 ***	4.307 ***
	Response latency	-0.093	-0.221
	GMC	0.027	0.023
	MUB	-0.486 ***	-0.469 ***
	R2	0.234	0.288
	N	94	88

Table 12: Linear regression (OLS) prejudices against Muslims on SD scales and response latency by sample

		model 1	model 2[1]
		beta	beta
student sample	(Constant)	3.541 ***	3.263 *
	Response latency	-0.039	-0.089
	GMC	0.116	0.071
	MUB	-0.569 ***	-0.551 ***
	R2	0.296	0.279
	N	91	86
access sample 1	(Constant)	3.492 ***	3.663 ***
	Response latency	0.027	0.016
	GMC	0.049	0.039
	MUB	-0.428 ***	-0.535 ***
	R2	0.167	0.166
	N	247	244
access sample 2	(Constant)	3.804 ***	4.399 ***
	Response latency	0.044	0.075
	GMC	0.035	0.019
	MUB	-0.436 ***	-0.706 ***
	R2	0.168	0.203
	N	175	165
GM sample	(Constant)	3.176 *	4.964 **
	Response latency	0.044	-0.051
	GMC	0.023	-0.012
	MUB	-0.413 ***	-0.570 ***
	R2	0.149	0.257
	N	94	88

1) Model 2: Linear regression under control of age, sex and education

Significance: +p<=0.10, *p<=0.05, **p<=0.01, ***p<=0.001

Firstly, there are some differences between the attitudes towards Jews and Muslims. Whereas, we found a stable influence on the MUB scale throughout all samples on prejudice towards the various groups, a barely noticeable effect on GMC and the response latency could be noticed solely in the explanations of prejudices against Jews. Furthermore, under control of the two social desirability scales, response latency had only a significant effect on the first access panel. In short, the longer it took the respondents to answer, the less prejudices seem to have been expressed. In line with our expectations, we found that the lower the MUB value was, the higher the conveyed prejudices against Muslims and Jews were. The respondents of the access panels with higher GMC, expressed in line with the hypothesis, less prejudice against Jews than those with a lower tendency to social desirability. These findings support the view that social desirability works throughout the editing process, in which the respondents evaluate their responses in accordance to less prejudiced attitudes (Holtgraves, 2004:171).

Secondly, and in contrast to our hypothesis, we found an effect of social desirability on anti-Semitism in the data drawn from the professional samples. The professional respondents answered in a less socially desirable manner and needed less time to answer, as well. It seems that the response time only had an unexpected negative effect with reference to anti-Semitism, whereas negative attitudes towards Muslims were apparently not influenced by the response time. However, the scope of this finding is very limited. Due to time measurement was solely possible for each page and refers, therefore, to the anti-Semitism as well as the islamophobia items.

While MUB reduced prejudice throughout all groups and samples, GMC showed, exclusively in the access panels, a rather small unexpected negative effect on prejudices against Jews. A possible explanation for this difference in answer behaviour could be that professional respondents are more trained to fill out web-questionnaires, but they also tend to disguise their opinion in favour of social desirability. It seems that the work-like relationship could abolish the anonymity of the web panels and generate a social impact. This could have influenced the quality of responses, whereas personal sentiment, however, seems to have not influenced the answers of the non-professionals.

6   Limitations

The present study has certain limitations. First, the time measurement represents each page of the survey and refers, therefore, to the respondents' attitudes toward both religious minority groups, Jews and Muslims. Second, we used only short versions of the social desirability measures. The short versions are well tested, but results obtained with the long versions might, however, have differed from our findings. Third, our study does not include an online survey based on a probability sampling method. Including a probability online survey would have significantly strengthened the ability to generalise our results. Fourth, a fifth anonymous group, without a personalized email invitation, would have enabled us to control for social impact on the results. Finally, the four online samples are of small sample sizes. Ideally, future studies should meet these limitations by using the complete measures of social desirability, asking only one question per page, including an online sample based on a probability sampling method from the general population, and increasing the sample sizes.

7   Conclusion

With reference to earlier findings (Gittleman & Tirmarchie, 2009:2; Sparrow, 2007:180; Toepoel, Das & van Soest, 2008:986; 2009), we assumed that there would be differences in answer behaviour between the "professional" and "non-professional" respondents.

For unit non-response, we did not find differences between these two groups of respondents. In contrast, the drop-out rates of the non-professionals were somewhat higher than those of the access panel participants. This was not only consistent for the first page drop-out, but throughout the whole questionnaire. Furthermore, we assumed that the work-like relationship might cause a lower item non-response rate for the professionals. Contrary to this expectation, the findings did not reveal differences for non-sensitive items with predefined answers amongst the samples. However, for

sensitive questions, the two groups showed entirely different results. The professional respondents tended to non-response more often than the volunteers. Since Shoemaker (2002:198) showed that non-response is related to item sensitivity and cognitive effort, the average sensitivity and effort could have been higher for professional respondents who did not anticipate prejudice items in the questionnaire because of their lack of interest in the topic. Moreover, we observed more answers for open-ended questions, albeit of lower quality within the access panels.

We also hypothesised a lower influence of social desirability for the access panels with regard to the absence of social presence. For prejudice against Muslims, we did not find differences among our four samples. Only the motivation to control prejudiced behaviour showed a significant effect across all samples. This was also revealed as true regarding the attitudes towards Jews in the non-professional samples. In contrast to this, the answers of the access panel participants towards these items were also influenced by the scale of motivation to unprejudiced behaviour and, in one case, by the response latency. The obvious reason behind this finding might be the historically based suppression of anti-Semitic statements in Germany. Surprisingly, the combination of anonymity and selectivity led to a stronger tendency for social desirability in the access panel data. Therefore, the results suggest that the work-like relationship of the professional respondents caused a form of social presence and social pressure, which caused confirmative response behaviour that undermined the anonymity of the access panels.

With regard to the discussion of panel conditioning effects related to attitude questions (Das, Toepoel & van Soest, 2007; Kruse et al., 2009; Toepoel, Das & van Soest 2009), our findings supported the previous findings that attitude questions are not influenced by panel conditioning for questions without sensitive content. In addition, however, we observed fairly strong effects within item non-response and answer behaviour for attitude questions with a higher level of sensitiveness.

All in all, the findings suggest that the type of online survey used has a significant impact on data quality, an observation which was especially noticed in the distinction between the student sample and the generated email sample with "non-professional respondents" on the one hand and the access panel types with "professional respondents" who used trained heuristics and tended to take shortcuts in the response process on the other hand. We believe that this could be of special importance to researchers who would like to ask sensitive questions. A particular survey type might produce less non-response, but a closer investigation might reveal substantial data quality problems.

The shown differences between the non-professional samples and the access panels remained steady over many aspects of data quality and challenge the accuracy of online access panels. However, these differences were far more significant when sensitive questions such as closed and open-ended questions on stereotypes and prejudice were asked. Former research had already pointed out that access panels might be problematic due their lack of representativeness (Yeager et al., 2011). Our findings add another drawback to the use of online access panels, namely that item non-response and social desirability could be similarly problematic. Alongside these, we suggest the use of multiple samples of different web survey types for online researchers who may not be able to use probability samples of the target population. Thus the lack of representativeness still remains, but stable results within a multiple sample type approach could ensure the findings to a certain degree.

Declaration of conflicting interests

Funding

References

[1]   Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (2010): *Jahresbericht 2010.*

[2]   Banse, R. and Gawronski, B. (2003): Die Skala Motivation zu vorurteilsfreiem Verhalten: Psychometrische Eigenschaften und Validität. *Diagnostica*, **49(1)**, 4–13.

[3]   Bassili, J. N. (1996): Meta-Judgmental Versus Operative Indexes of Psychological Attributes: The Case of Measures of Attitude Strength. *Journal of Personality and Social Psychology*, **71(4)**, 637–653.

[4]   Bergmann, W. and Erb, R. (1986): Kommunikationslatenz, Moral und Öffentliche Meinung: Theoretische Überlegungen zum Antisemitismus in der Bundesrepublik Deutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, **38(4)**, 223–246.

[5]   Bethlehem, J. G. and Biffignandi, S. (2012): *Handbook of Web Surveys*. Hoboken, NJ: Wiley.

[6]   Biemer, P. P. and Lyberg, L. E. (2003): *Introduction to Survey Quality*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

[7]   Bosnjak, M. and Tuten, T. L. (2001): Classifying Response Behaviors in Web-based Surveys. *Journal of Computer-Mediated Communication*, **6(3)**, Retrieved August 27th, 2013, from http://www.ascusc.org/jcmc/vol6/issue3/boznjak.html

[8]   Bosnjak, M. and Tuten, T.L. (2003): Prepaid and promised incentives in web surveys: an experiment. Social Science Computer Review, **21(2)**, 208–217.

[9]   Bowling, A. (2005): Mode of Questionnaire Administration Can Have Serious Effects on Data Quality. *Journal of Public Health*, **27(3)**, 281–291.

[10]   Birnholtz, J.P., Horn, D.B., Finholt, T.A. and Bae, S.J.(2004): The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents. *Social Science Computer Review* **22(3)**, 355–362.

[11]   Chang, L. and Krosnick, J. A. (2009): National Surveys Via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly*, **73(4)**, 641–678.

[12]   Couper, M. P. (2000): Web Surveys: A Review of Issues and Approaches. *American Association for Public Opinion Research*, **64(4)**, 464–494.

[13]   Couper, M. P. and Coutts, E. (2006): Online-Befragung: Probleme und Chancen verschiedener Arten von Online-Erhebungen. In A. Diekmann (Ed.), *Methoden der Sozialforschung*, 217-243. Wiesbaden: VS Verl. für Sozialwiss.

[14]   Couper, M. P. and Miller, P. V. (2008): Web Survey Methods: Introduction. *Public Opinion Quarterly*, **72(5)**, 831–835.

[15]   Crowne, D. P. and Marlowe, D. (1960): A New Scale of Social Desirability Independent of Psychopathology. *Journal of Consulting Psychology*, **24(4)**, 349–354.

[16]   Das, M., Toepoel, V. and van Soest, A. (2007): Can I Use a Panel? Panel Conditioning and Attrition Bias in Panel Surveys. *Center Discussion Paper,* **56**, 1–26.

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

[18]   Dennis, J. M. (2001): Are Internet Panels Creating Professional Respondents? A Study of Panel Effects. *Marketing Research*, **13(2)**, 34–38.

[19]   Denscombe, M. (2009): Item Non-Response Rates: A Comparison of Online and Paper Questionnaires. *International Journal of Social Research Methodology*, **12(4)**, 281–291.

[20]   Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E. and Vance, S. L. (2002): The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice. *Journal of Personality and Social Psychology*, **82(5)**, 835–848.

[21]   Dillmann, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J. and Messer, B. L. (2009): Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, **38(1)**, 1–18.

[22]   Dillman, D. A. (1978): *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.

[23]   Druckman, J. N. and Kam, C. D. (2011): Students as Experimental Participants: A Defense of the "Narrow Data Base". In J. N. Druckman (Ed.), *Cambridge Handbook of Experimental Political Science*, 41-57. Cambridge: Cambridge University Press.

[24]   Dunton, B. C. and Fazio, R. H. (1997): An Individual Difference Measure of Motivation to Control Prejudiced Reactions. *Personality and Social Psychology Bulletin*, **23(3)**, 316–326.

[25]   Esser, H. (1986): Über die Teilnahme an Befragungen. *ZUMA-Nachrichten*, **18**, 38–47.

[26]   Fazio, R. H. (1990): A Practical Guide to the Use of Response Latency in Social Psychological Research. In C. Hendrick & M. S. Clark (Eds.), *Research Methods in Personality and Social Psychology*, 74–97. Newbury Park: Sage Publications.

[27]   Flere, S. and Lavric, M. (2008): On the Validity of Cross-Cultural Social Studies Using Student Samples. *Field Methods*, **20(4)**, 399–412.

[28] Frick, A., Bächtiger, M. T. and Reips, U.-D. (2001): Financial Incentives, Personal Information and Drop-Out in Online Studies. In U.-D.Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science*, 209-219. Lengerich: Pabst.

[29] Fricker, R. D. (2008): Sampling Methods for Web and E-mail Surveys. In N. Fielding, R. M. Lee & G. Blank (Ed.), The Sage Handbook of Online Research Methods, 195-216. Los Angeles; London; New Delhi; Singapore: Sage.

[30] Fricker, S., Galesic, M., Tournageau, R. and Yan, T. (2005): An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly*, **69(3)**, 370–392.

[31] Gittelman, S. and Trimarchi, E. (2009): *On the road to clarity: Differences between sample sources.* Retrieved August 27th, 2013, from

http://www.mktginc.com/pdf/Casro-The%20road%20to%20clarity%201-30-09.pdf

[32] Göritz, A. S.(2004): The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Market Research*, **46(3)**, 327–345.

[33] Göritz, A. S. (2006): Incentives in web studies: methodological issues and review. *International Journal of Internet Science* **1(1)**, 58–70.

[34] Göritz, A. S. (2007). Using online panels in psychological research. In A. N. Joinson (Ed.), *The Oxford handbook of Internet psychology,* 473–485. Oxford; New York: Oxford University Press.

[35] Groves, R. M., Presser, S. and Dipko, S. (2004): The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, **68(1)**, 2–31.

[36] Groves, R. M. (2004): *Survey Errors and Survey Costs*. Hoboken, N.J: Wiley-Interscience.

[37] Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. (2002): *Survey Nonresponse*. New York: Wiley.

[38] Groves, R. M. and Couper, M. (1998): *Nonresponse in Household Interview Surveys*. New York: Wiley.

[39] Hansen, K. M. and Pedersen, R. T. (2011): Efficiency of different recruitment strategies for web panels, *International Journal of Public Opinion Research*, **24(2)**, 238-249.

[40] Heerwegh, D. (2005): Effects of Personal Salutations in E-mail Invitations to Participate in a Web Survey. *Public Opinion Quarterly*, **69(4)**, 588–598.

[41] Heerwegh, D. (2006): An investigation of the effect of lotteries on web survey response rates. *Field Methods*, **18(2)**, 205–220.

[42] Heerwegh, D. and Loosveldt, G. (2007): Personalizing E-mail Contacts: Its Influence on Web Survey Response Rate and Social Desirability Response Bias. *International Journal of Public Opinion Research*, **19(2)**, 258–268.

[43] Holbrook, A. L., Krosnick, J. A. and Pfent, A. M. (2007): Response rates in surveys by the news media and government contractor survey research firms. In J. Lepkowski, B. Harris-Kojetin, P. J. Lavrakas, C. Tucker, E. de Leeuw, M. Link, M. Brick, L. Japec, & R. Sangster (Eds.), *Advances in Telephone Survey Methodology*, 499-528. New York: Wiley.

[44]   Holland, J. L. and Christian, L. M. (2009): The Influence of Topic Interest and Interactive Probing on Responses to Open-Ended Questions in Web Surveys. *Social Science Computer Review*, **27(2)**, 196–212.

[45]   Holtgraves, T. (2004): Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin*, **30(2)**, 161–172.

[46]   Huckfeldt, R., Levine, J., Morgan, W. and Sprague, J. (1999): Accessibility and the Political Utility of Partisan and Ideological Orientations. *American Journal of Political Science*, **43(3)**, 888–911.

[47]   Kaplowitz, M. D., Hadlock, T. D. and Levine, R. (2004): A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly*, **68(1)**, 94–101.

[48]   Kaplowitz, M. D., Lupi, F., Couper, M. P. and Thorp, L. (2012): The Effect of Invitation Design on Web Survey Response Rates. *Social Science Computer Review*, **30(3)**, 339–349.

[49]   Knapp, F. and Heidingsfelder, M. (1999): Drop-out-Analyse: Wirkung des Untersuchungs-designs. In U. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, & K. W. A. Moser (Eds.), *Current Internet science - trends, techniques, results.* Zürich: Online Press.

[50]   Koo, M. and Skinner, H. (2005): Challenges of Internet Recruitment: A Case Study with Disappointing Results. *Journal of Medical Internet Research*, **7(1)**, e6.

[51]   Kreuter, F. (2002): *Kriminalitätsfurcht: Messung und methodische Probleme*. Opladen: Leske + Budrich.

[52]   Kreuter, F., Presser, S. and Tourangeau, R. (2008): Social Desirability Bias in CATI, IVR, and Web Surveys: The Effect of Mode and Question Sensitivity. *Public Opinion Quarterly*, **72(5)**, 847–865.

[53]   Kruse, Y., Callegaro, M., Dennis, J. M., DiSorga, C., Subias, S., Lawrence, M. and Tompson, T. (2009): Panel Conditioning and Attrition in the AP-Yahoo! News Election Panel Study. *American Association for Public Opinion Research*, 5742–5754.

[54]   Lee, R. M. (1993): *Doing Research on Sensitive Topics*. London: Sage Publications.

[55]   Lensvelt-Mulders, G. (2008): Surveying Sensitive Topics. In E. D. de Leeuw, D. A. Dillman, & J. J. Hox (Eds.), *International Handbook of Survey Methodology*, 461–478. New York: Lawrence Erlbaum Ass.

[56]   Malhotra, N. and Krosnick, J. A. (2007): The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples. *Political Analysis*, **15(3)**, 286–323.

[57]   Martin, C. L. (1994): The Impact of Topic Interest on Mail Survey Response Behaviour. *Journal of the Market Research Society*, **36(4)**, 327–338.

[58]   Mayerl, J. (2005): Controlling the Baseline Speed of Respondents: An Empirical Evaluation of Data Treatment Methods of Response Latencies. In C. van Dijkum, J. Blasius, & B. van Hilton (Eds.), *Recent Developments and Applications in Social Research Methodology. Proceedings of the Sixth International Conference on Logic and Methodology* (2nd ed.), 1-20. Leverkusen-Opladen: Barbara Budrich.

[59]   Mayerl, J. and Urban Dieter. (2008): *Antwortreaktionszeiten in Survey-Analysen: Messung, Auswertung und Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

[60]   Paulhus, D. (1991): Measures of Personality and Social Psychological Attitudes. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of Social Psychological Attitudes*, 17–59. Academic Press.

[61]   Paulhus, D. L. (2002): Socially Desirable Responding: The Evolution of a Construct. In H. I. Braun, D. N. Jackson, D. E. Wiley, & S. Messick (Eds.), *Role of Constructs in Psychological and Educational Measurement*, 49–69. Mahwah, NJ: L. Erlbaum.

[62]   Peterson, R. A. (2001): On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research*, **28(3)**, 450–461.

[63]   Petric, I., Appel, M. and de Leeuw, E. D. (2009): *Online Interviewing Through Access Panel: Quantity and Quality Assurance.* Conference Paper. Worldwide Readership Research Symposia.

[64]   Reja, U., Manfreda, K. L., Hlebec, V. and Vehovar, V. (2003): *Open-ended vs. Close-ended Questions in Web Questionnaires.* Metodoski Zvezki, **19**, 159-177.

[65]   Roose, H., Lievens, J. and Waege, H. (2007): The Joint Effect of Topic Interest and Follow-Up Procedures on the Response in a Mail Questionnaire: An Empirical Test of the Leverage-Saliency Theory in Audience Research. *Sociological Methods & Research*, **35(3)**, 410–428.

[66]   Roose, H., Waege, H. and Agneessens, F. (2003): Respondent Related Correlates of Response Behaviour in Audience Research. *Quality and Quantity*, **37(4)**, 411–434.

[67]   Salzborn, S. and Schwietring, M. (2003): Antizivilisatorische Affektmobilisierung. Zur Normalisierung des sekundären Antisemitismus. In M. Klundt, S. Salzborn, M. Schwietring, & G. Wiegel (Eds.), Erinnern, verdrängen, vergessen. Geschichtspolitische Wege ins 21. Jahrhundert, 43-76. Giessen: Netzwerk für politische Bildung, Kultur und Kommunikation e.V.

[68]   Sanchez-Fernandez, J., Munoz-Leiva, F., Montoro-Rios, F. J. and Ibanez-Zapata, J. A. (2010): An analysis of the effect of pre-incentives and post-incentives based on draws on response to web surveys. *Quality & Quantity*, **44(2)**, 357-373.

[69]   Schnell, R. (1997): *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen: Leske+Budrich.

[70]   Shin, E., Johnson, T. P. and Rao, K. (2012): Survey Mode Effects on Data Quality: Comparison of Web and Mail Modes in a U.S. National Panel Survey. *Social Science Computer Review,* **30(2)**, 212–228.

[71]   Shoemaker, P. J. (2002): Item Nonresponse: Distinguishing between don't Know and Refuse. *International Journal of Public Opinion Research*, **14(2)**, 193–201.

[72]   Sparrow, N. (2007): Quality Issues in Online Research. *Journal of Advertising Research*, **47(2)**, 179–182.

[73]   Stöber, J. (1999): Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität. *Diagnostica*, **45(4)**, 173–177.

[74]   Stoop, I. A. L. (2004): Surveying Nonrespondents. *Field Methods*, **16(1)**, 23–54.

[75]   Stoop, I. A. L. (2007): The Hunt for the Last Respondent: Nonresponse in Sample Surveys. *Public Opinion Quarterly*, **71(1)**, 167–169.

[76]   Taddicken, M. (2009): Die Bedeutung von Methodeneffekten der Online-Befragung in der empirischen Sozialforschung: Zusammenhänge zwischen computervermittelter Kommunikation und erreichbarer Datengüte. In N. Jackob, H. Schoen, & T. Zerback (Eds.), *Sozialforschung im Internet. Methodologie und Praxis der Online-Befragung*, 91–107. Wiesbaden: VS Verlag für Sozialwissenschaften.

[77]   Toepoel, V., Das, M. and van Soest, A. (2008): Effects of Design in Web Surveys: Comparing Trained and Fresh Respondents. *Public Opinion Quarterly*, **72(5)**, 985–1007.

[78]   Toepoel, V., Das, M. and van Soest, A. (2009): Relating Question Type to Panel Conditioning: Comparing Trained and Fresh Respondents. *Survey Research Methods*, **3(2)**, 73-80.

[79]   Tourangeau, R. and Smith, T. W. (1996): Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, **60(2)**, 275–304.

[80]   Tourangeau, R. and Yan, T. (2007): Sensitive Questions in Surveys. *Psychological Bulletin*, **133(5)**, 859–883.

[81]   Tourangeau, R. (2007): Incentives, Falling Response Rates, and the Respondent-Researcher Relationship. Paper presented at the *Proceedings of the Ninth Conference on Health Survey Research Methods,* 244–253. Hyattsville.

[82]   Walczyk, J. J., Mahoney, K. T., Doverspike, D. and Griffith-Ross, D. A. (2009): Cognitive Lie Detection: Response Time and Consistency of Answers as Cues to Deception. *Journal of Business and Psychology*, **24(1),** 33–49.

[83]   Walczyk, J. J., Schwartz, J. P., Cliftone, R., Barett, A., Wei, M. and Zha, P. (2005): Lying Person-To-Person About Life Events: A Cognitive Framework for Lie Detection. *Personnel Psychology*, **58(1)**, 141–170.

[84]   Wiecko, F. M. (2010): Research Note: Assessing the Validity of College Samples: Are Students Really That Different? *Journal of Criminal Justice*, **38(6)**, 1186–1190.

[85]   Yeager, D. S., Krosnick, J. A., Chang, L., Javitz, H. S., Levendusky, M. S., Simpser, A. and Wang, R. (2011): Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly*, **75(4)**, 709–747.

Appendix

Table 2a: Significances for the multiple comparison for Table 2

	student sample	access sample 1	access sample 2	GM sample
student sample		0.00	0.00	0.52
access sample 1			0.83	0.00
access sample 2				0.00

Table 3a: Significances for the multiple comparison for Table 3

	student sample	access sample 1	access sample 2	GM sample
student sample		0.00	0.00	0.60
access sample 1			0.47	0.00
access sample 2				0.00

Table 4a: Significances for the multiple comparison for Table 4

	student sample	access sample 1	access sample 2	GM sample
student sample		0.00	0.00	0.11
access sample 1			0.21	0.12
access sample 2				0.03

Table 5a: Significances for the multiple comparison for Table 5

	student sample	access sample 1	access sample 2	GM sample
student sample		0.00	0.01	0.95
access sample 1			0.62	0.01
access sample 2				0.01