Erschienen in:
Journal of Official Statistics 22 (2), 2006, 177-19.

Putting a Questionnaire on the Web is not Enough

A Comparison of Online and Offline Surveys Conducted in the Context of the German Federal Election 2002

Thorsten Faas[1] *and Harald Schoen*[2]

The article analyses how different methods of recruiting Internet users for online surveys affect survey results in terms of marginal distributions and associations between variables. The general hypothesis is that self-selection processes in some kinds of online surveys, especially in open online surveys bias marginal distributions: those more interested and involved in the topic (in our case: politics) are expected to be over-represented. This bias furthermore causes (political) attitudes to be much more structured among participants of (open) online surveys. The hypothesis is tested using data from face-to-face interviews of Internet users, an access panel of online users and an open online poll, all collected in the context of the 2002 German federal election. Empirical analyses show that both marginal distributions and relationships of variables stemming from the open online survey differ clearly from findings based on the other surveys. Additionally, the analysis reveals that standard weighting procedures do not reduce these biases substantially.

Key words: Online Surveys; survey methodology; Internet research; election studies; Germany.

[1] Institut für Politikwissenschaft, Universität Duisburg-Essen, Campus Duisburg, Lotharstraße 65, 47048 Duisburg, Germany. Email: Thorsten.Faas@uni-duisburg.de
[2] Institut für Politikwissenschaft, Johannes Gutenberg-Universität, Saarstraße 21, 55099 Mainz Germany. Email: harald.schoen@politik.uni-mainz.de

1 Introduction

Like previous innovations in the communication sector, the advent of the Internet added a new technique to social scientists' toolbox for collecting (survey) data. Only a few years after its invention, it is already frequently used to conduct online surveys, as it provides numerous advantages to users: Using Internet technology allows for interviewing huge numbers of respondents within a rather short period of time (Weible and Wallace 1998; Schaefer and Dillman 1998). In 2004 in Germany, e.g., more than 510,000 Germans took part in an online survey called “Perspektive Deutschland” within a short period of time, which – according to the initiators – is the largest socio-political survey in the world (Perspektive Deutschland 2005a). Online surveys dramatically lower the marginal costs of data collection compared to traditional personal, mail or telephone interviews (e.g. Mehta and Sivadas 1995; Schuldt and Totten 1999; Sheehan and McMillan 1999). They do not require the presence of interviewers, who, moreover, could potentially cause measurement error. They also offer a chance to combine different kinds of stimuli, e.g. sounds, pictures, and movies. Finally, due to the novelty of the instrument, results of online surveys quite easily can reach the public’s attention (Taylor 2000: p. 53, p. 57; Batinic 2001, pp. 12-14; Alvarez et al. 2003, p. 23).

Notwithstanding these advantages, online surveys are also criticized. Critiques point to technical problems; e.g. due to specific browser settings and consequently systemic incompatibilities (Couper et al. 2001). It is also argued that the more visually oriented, possibly multimedia presentation of stimuli may cause systematic distortions of results (Dillman 2000, pp. 352-433). More severely, critiques question whether results obtained from online surveys can be generalized to the public as a whole (e.g. Ray et al. 2001). After all, respondents in online surveys are obviously not a random sample of the population, but tend

to be younger and better educated (e.g. Bandilla et al. 2001, pp. 8-11; Vehovar et al. 2002, p. 239).

But even generalizing findings from online surveys to the population of Internet users is everything but trivial. One first of all has to acknowledge that the classification of “Internet users” is not self-evident. More important, though, it has to be recognized that "online survey" is a label used for very different kinds of surveys (Couper 2000; Schonlau et al. 2002). Sometimes, the term “online survey” refers to an offline-recruited random sample that has been equipped with online technology, which is, e.g., what Knowledge Networks in the US or forsa (with its omni.net) in Germany do. More commonly, access panels of randomly selected, previously in offline surveys recruited Internet users are interviewed online and constitute “online surveys.”[3] Probably the most popular form of online surveys, however, is the open, unrestricted web survey, where a questionnaire is put on the web and everybody can join in voluntarily. Though all of these applications use the same technology (namely the Internet), they nonetheless are quite different instruments in terms of methodological standards and rigour. Hence, assessments of advantages and drawbacks of online surveys must not be formulated without specifying which kind of online survey is addressed.

Our aim here is to shed more light on the data quality that different types of surveys of Internet users render. We shall compare Internet users from a traditional offline survey based on face-to-face interviews (which we use as a yardstick) to two types of online surveys: an access panel, whose members were previously recruited from offline surveys, and an open, unrestricted online poll. All of the surveys were conducted in the context of the 2002 German federal election. The outline is as follows: First, we will present our theoretical argument and

---

[3] This must not be confused with – to use Couper’s (2000) terminology – “volunteer opt-in panels”. Throughout the article, the term “access panel” refers to panels of Internet users who were randomly selected from standard offline surveys.

specify our hypotheses. Then, we will investigate whether the results obtained from the three surveys differ with respect to marginal distributions. After that, we will analyze whether associations between variables differ between the three surveys, as – after all – the aim of social science is to look at distributions *and* associations. Finally, we will conclude. Our general hypothesis is that results based on (the Internet users from) the offline survey and the access panel should not differ significantly, since they both constitute a random sample of Internet users. However, due to the self-selection processes in open online surveys, we expect marginal distributions and associations to be different there: those more deeply involved in politics are expected to be over-represented, since an online *election* survey is more attractive to them. This bias of marginals should furthermore cause political attitudes to be much more structured among these participants, which should affect (in fact: strengthen) the associations among variables as well.

2 Theoretical Expectations

Advocates of online surveys claim that these constitute a serious alternative to traditional forms of gathering interview data when appropriately weighted, even for analysis of the population as a whole (see, e.g., Harris Interactive 2000; Perspektive Deutschland 2005b). However, this claim has not gone without criticism (e.g., Ray et al. 2001). Even restricting oneself to Internet users does not necessarily solve all problems. Depending on the type of online survey, it cannot be taken for granted that online surveys yield results that are representative for *online users* only. Depending on the way respondents are recruited one can expect different degrees of validity: if respondents are recruited using probability sampling procedures results will be valid; in contrast, using procedures that allow for self-selection, results can be expected to be biased both in terms of marginal distributions and associations

among variables. One can look at different types of recruiting Internet users to make this point clear.

If we first of all look at established procedures of recruiting offline samples, the population as a whole is used as a sampling frame from which a random sample is drawn using multi-staged sampling techniques. Of course, there are problems of defining the sampling frame appropriately, of sampling error and of non-response (Groves 1987, S159-S166, 1989, pp. 1-37). Despite these problems, however, using this strategy allows one to approximate a random sample of the population as a whole quite well. But if that is the case, such a sample should also include a representative sub-sample of Internet users.

When using an access panel of online users, the situation is quite similar. In the case of online access panels, a sample is drawn from a panel of Internet users who previously have agreed to participate in online surveys. The recruiting of members of access panels usually takes place continuously during offline surveys based on random samples. Ideally, this represents a multi-staged sampling process that also – in the end – yields a representative sample of Internet users. Of course, as Internet users are more likely to be male, younger and more highly educated than the public as a whole, this procedure does *not* yield a representative sample of the general public (see, e.g., Bandilla et al. 2001, p. 17; Batinic 2001, pp. 48-51).[4]

Concerning open online surveys, the situation is different: In this case there is no clear-cut sampling frame at all, since potential respondents decide on an entirely voluntary (and idiosyncratic) basis to participate. In total, there are three thresholds of participation in this case: First, a person has to become aware of the survey; second, the person must have access

[4] There are even doubts whether access panels are representative for Internet users since it appears that heavy users are over-represented (see Faas 2003).

to the Internet; third, (s)he has to decide to participate. While the second and third thresholds parallel the mechanisms at work when pre-recruited access panels are used, especially the first threshold is a unique and far-reaching feature of open online surveys, as no effort is made to actively *and systematically* recruit respondents. As far as online surveys are advertised in the Internet, persons without an access to the Web have no chance at all to notice them; in contrast, heavy Internet users are very likely to become aware of them (see Bandilla 1999, p. 12; Hauptmanns 1999, pp. 24-29). Additionally, advertisements for online surveys are not distributed equally among web sites; rather, it depends upon the subject of the survey on which web sites ads will be placed: hints to political surveys will be found more frequently on web sites with political content than on sports sites (see Vehovar 2002, p. 235; Bosnjak 2002). Thus, persons who are interested in politics are disproportionately more likely to become aware of an open political Internet survey than other people (and then also to take part). As a result, the composition of such an open online survey should be heavily biased in terms of sex, age, education, and political involvement, e.g., in terms of interest in politics, likelihood of voting and party membership. Taken together, the sample selection process should considerably influence marginal distributions of socio-demographic and substantial variables.

Turning to associations among variables, they are generally said to be more robust to selection biases than marginal distributions (e.g., Schnell 1991, p. 133). Still, it cannot be precluded that the hypothesized composition effects stemming from different strategies to recruit respondents cause differences in the associations of variables, too. Participants in open online surveys are expected to be (more) heavily involved in politics. With respect to associations among variables, this is likely to have consequences: The political attitudes of persons deeply involved in politics are more crystallized, more stable and more structured than political orientations of persons who lack any political involvement (see Bartle 2000; Converse 1964; Zaller 1990). Thus, one can expect correlations of political attitudes to be strongest among

respondents of open online surveys (see for composition effects that may distort associations between variables Berk 1983; Groves 1989, 90-95; Schoen 2004). In addition, since people involved in politics are usually more polarized in terms of their partisan affiliations, their partisan attitudes can be expected to be more structured. Specifically, attitudes towards political objects, e.g., candidates, of the same party should thus correlate more strongly while attitudes towards objects of different parties should be correlated more negatively (see, e.g., Campbell et al. 1960, pp. 128-136; Falter et al. 2000, pp. 251-255).

It appears to be even more reasonable to expect stronger correlations among participants in the open online survey when measurement error is taken into account. So-called nonattitudes (Converse 1970; see also Schuman and Presser 1996, pp. 147-160) are less frequently found among better-educated and politically involved persons. Consequently, measurement error can be expected to be less common there. Since measurement error regularly lowers correlations (see Achen 1983; 1985), correlations will be stronger among the better educated and politically involved than among other persons (see, e.g., Norpoth and Lodge 1985; Zaller 1992, 21; Bartle 2000).

As the preceding discussion shows, self-selection bias limits the validity of survey results. One way to deal with this problem may be weighting the data. It is argued that appropriate weighting procedures will make survey results representative even if self-selection processes are at work (e.g., Terhanian and Bremer 2002; Perspektive Deutschland 2005b). One might be sceptical whether weighting procedures will improve the results because weights have to fit some criteria for this purpose; most importantly, the variables that the weights are based on must be correlated very strongly with the variables of interest. In addition, one can expect that the self-selection bias depend on the overall subject of the survey. In an election survey, the bias is likely to connected to interest in politics, in a survey about football, it is probably

linked to interest in football. As a consequence, it should be very hard to find universally applicable weights to control for self-selection biases.

Still, we will include weighting in our analysis. In order to test the effect of weighting empirically we will analyze the data in two steps, first, without any weighting procedures and then, using socio-demographically defined weights. These weights are of course, not as complex as the weights used by advocates of online surveys. Their propensity weights go far beyond socio-demographic variables and also include substantial variables. Still, even simple socio-demographic weights should improve results to some extent, if weighting is the key to solving the self-selection processes.

Before we turn to our empirical analysis, a final word of caution: We are, of course, aware of that fact that surveys can (and often do) differ in other important respects – field times, response rates and most prominently the survey mode. In a face-to-face interview, an interviewer sits in front of the respondent, reads out the questions and writes down the answers, while the stimuli are presented visually in online surveys. No interviewer is involved in data gathering since respondents answer questions by themselves. However, in the present article we will focus our attention on self-selection effects.[5]

In summary, we argue that recruitment procedures influence the validity of results of Internet users. Using a face-to-face survey of Internet users as a yardstick, we do not expect a bias in sample composition for the access panels of online users, but a strong bias for an open online

---

[5] In a general perspective, the effect of the online mode on survey answers is not agreed upon. On the one hand, it is argued that the nonexistence of an interviewer makes measurement error less likely that is caused by effects of social desirability (see e.g. Nicholson et al. 1998; King and Miles 1995; Stanton 1998; Ilieva et al. 2002, p. 368, p. 374). Additionally, it is pointed to the fact that respondents in online surveys are not forced to answer immediately and therefore may give more reasonable answers. On the other hand, it is noticed that the anonymity of an online survey may lead respondents to give more nonsense answers (see Batinic 2001, p. 57). As a result, it cannot be concluded whether measurement error in online surveys is more or less common than in face-to-face surveys.

survey due to self-selection processes. The distortion of marginal distributions can bias correlations among variables, too: especially participants of unrestricted online surveys are expected to have political attitudes that are better structured than those of offline respondents. Additionally, we test how standard weighting procedures affect the results.

3 Data

In order to test these hypotheses, data similar in terms of content, but collected in different ways are needed. Our empirical analysis will be based on three different data sets:

- An offline recruited sample of the whole public being interviewed personally, from which only Internet users are used for the following comparison.
- A sample based on an access panel of previously in offline surveys recruited online users.
- An open, unrestricted online survey.

All of them were conducted in the context of the 2002 German federal election (see Table 1 for details). In terms of questions, the surveys each contained numerous (identical) items coming from the standard toolbox of electoral researchers, like candidate and issue orientations, trust in political institutions, voting intentions (see for an overview for Germany Klein et al. 2000; Falter/Schoen 2005).

To give some more details about the surveys, the first survey is a representative sample of Germany's population aged 16 and over. From August 12th to September 21st, respondents were randomly selected using sample points, random routes and last birthdays and interviewed personally. In sum, the survey comprised 1,665 respondents, 507 of them

identified themselves as Internet users.[6] They should arguably comprise a representative survey of German Internet users at the time. The second survey is also a representative sample of German Internet users, albeit selected in a very different way. A total of 598 users – again aged 16 and over – were randomly selected from an access panel which comprised about 4,000 respondents in 2002; its members had been previously recruited in standard offline surveys. This survey was fielded from September 13th to September 21st; the survey mode was a web survey. Finally, a third survey was conducted as an open, unrestricted online poll. Everybody could logon to www.wahlumfrage2002.de ("wahlumfrage2002" stands for electionsurvey2002) and fill out the online questionnaire from August 20th to September 22nd. In other words, participants recruited themselves; a total of 34.098 did so. In the following, we will restrict this to 29,583 respondents: They answered at least five of the questions asked (see Faas 2002 for further details about this survey).

Table 1: Details concerning the three surveys

	Face-to-face survey	Survey based on access panel	Open online survey
Field time	August 12th to September 21st, 2002	September 13th to September 21st, 2002	August 20th to September 22nd, 2002
Respondents	1.665	598	34.098/29.583[a]
Recruiting of respondents	Random selection based on sample points, random routes and last birthdays	Random selection from a previously offline recruited access panel	Self-recruiting without restrictions
Mode	Paper-and-pencil	Web survey	Web survey

[a] The total number of respondents amounts to 34,098, 29,583 of them gave at least five valid answers to the questions asked.

[6] The question used was: "Do you use the Internet?", answering options were "no", "yes, at home", "yes, at work", "yes, primarily at home, but also at work" and "yes, primarily at work, but also at home". We counted all those as Internet users that checked one of the "yes" options.

In a nutshell, then, we have three surveys each capable of reaching German Internet users in 2002. Two of them should arguably be representative for the target group of Internet users, the third one – an open, online survey – is "the odd one out", as it is entirely based on self-selection. To make comparisons between the three surveys more feasible, we will – along with comparing the three surveys' raw data – adjust the socio-demographic composition of the open, online survey in terms of sex, age and education to the other two. However, to do so, we will have to look at the socio-demographic composition first, so we will come back to the questions of appropriate weights after that.

4 Empirical Findings

4.1 Marginal distributions of socio-demographic variables

Concerning marginals, we first of all find no differences in terms of sex and education between Internet users from the offline survey and those coming from the access panel (see Table 2): In both cases, the percentage of male Internet users is in the high fifties, the percentage of those having tertiary education (i.e., are eligible to go to university) is in the high forties. We do find statistically significant differences in terms of age, although their extent is not too high: the access panellists are about two years younger with an average age of 37, even though the same sampling frame is used in terms of age (16 and over).

However, differences instead of similarities dominate the picture once we look at the open, unrestricted online survey. Participants are significantly and substantially younger, better educated and more often male than female. The strength of the detected biases concerning this open poll is truly remarkable: 78 per cent of the respondents are male, 76 per cent have

completed tertiary education and on average, they are only 33 years old. Compared to the other distributions, these socio-demographic scores are considerably skewed.

What do these results imply in terms of weighting? Obviously, there is no need for socio-demographic weighting concerning the comparison between the offline survey and the access panel, as their respective composition is more or less the same. However, we do have to weight the data stemming from the open online survey. We do so by adjusting its combined distribution of sex, age and education to the respective distribution in the access panel; for this purpose, age and education were classified into three groups (age: up to 29, 30 to 59, 60 and over; education: still in school, up to secondary education, tertiary education). Thus, in the following, there will always be two results coming from the open online survey: an unweighted and a weighted one.

Table 2: Socio-demographic characteristics of the respondents of the three surveys

	Face-to-face Survey	Survey based on access panel	Open online survey	Significance[a] of difference between		
	(1)	(2)	(3)	1 and 2	1 and 3	2 and 3
Sex - % male	57	59	78	n.s.	***	***
Age – mean	39	37	33	**	***	***
Education - % tertiary (respondents still in school are excluded)	48	49	76	n.s.	***	***

[a] Significance levels: n.s. – not significant, *: $p<.05$, **: $p< .01$, ***: $p<.001$, based on ANOVA/Scheffe-test.

4.2 Marginal distributions of substantial variables

The pattern that emerges when we look at substantial variables of political involvement is basically the same as it was before for the socio-demographic variables. This time, we do not observe *any* significant differences between the offline survey and the access panel. The level

of general interest in politics, of specific interest in the election campaign, the likelihood of voting and the share of party members is generally rather modest within the two (with the exception of voting), but virtually identical between the two (see Table 3).

Quite the reverse is true for the open online survey. The respective scores show a considerably higher level of political involvement and consequently significant differences when compared to the two other surveys. The self-selected respondents are highly interested in politics, highly interested in the campaign, are even more certain to vote and – most strikingly – almost one in four of them is a party member! Given the size of these differences, it is, of course, not surprising to see that all of these differences are statistically significant. A survey labelled “electionsurvey2002” apparently attracts mainly politically involved persons – even their considerable number of about 30,000 does not alter that tendency. Weighting the data hardly affects the results at all.

Table 3: Political involvement among the participants of the three surveys (results for the open online survey are shown unweighted as well as weighted)

	Face-to-face Survey	Survey based on access panel	Open online survey	Significance[a] of difference between		
	(1)	(2)	(3)	1 and 2	1 and 3	2 and 3
Interest in politics[b] – Mean	0,4	0,3	1,1 / 1,0	n.s.	*** / ***	*** / ***
Interest in campaign[c] – Mean	0,3	0,4	1,2 / 1,2	n.s.	*** / ***	*** / ***
Likelihood of voting[d] – mean	1,8	1,8	1,9 / 1,9	n.s.	*** / ***	*** / ***
Party membership[e] – in %	7	7	23 / 23	n.s.	*** / ***	*** / ***

[a] Significance levels: n.s. – not significant, *: $p<.05$, **: $p<.01$, ***: $p<.001$, based on ANOVA/Scheffe-test.

[b] The wording of the question was: "How strong is your interest in politics? Is it very strong, rather strong, medium, rather weak or very weak?" Coding was from –2 (very weak) to +2 (very strong).

[c] The wording of the question was: "Another question about the federal election 2002: How closely do you follow the campaign? Very closely, closely, medium, not very closely, not at all?" Coding was from -2 (not at all) to +2 (very closely).

[d] The wording of the question was: "In the upcoming federal election, will you definitely vote, probably vote, maybe vote, probably not vote, definitely not vote?" Coding was from -2 (definitely not vote) to +2 (definitely vote)

[e] The question wording was: "Are you a member of party? And if so, do you hold an office?" Coding was 1 (for members regardless of holding an office or not) and 0 (for nonmembers).

4.3 Associations

The analysis of marginal distributions has revealed two important results that should also affect the findings in terms of associations: Participants of the open online survey are more sophisticated, i.e., they have – on average – a higher level of formal education, and are more politically involved. Both findings support the expectation that we outlined above: Their attitudes should be more structured and hence, associations among variables should be stronger. Our surveys include several items that should theoretically be connected; we will look at left–right positions of respondents and feeling thermometers towards parties and candidates.

As Germany's party system is structured along the left–right dimension, party-related feeling thermometers should be closely connected to left–right positions of respondents. However, as the left–right dimension requires some political sophistication (see on the liberal–conservative dimension in the U.S. Kuklinski et al. 1982; Sniderman et al. 1991: 176; Jacoby 2004), the association between the two can be used as a first reasonable test of our hypothesis. Table 4 shows the respective scores: First of all, the basic assumption underlying our argument (i.e., the connection between left–right and the feeling thermometers) is supported: The further to the right respondents place themselves on the left–right-dimension, the better they evaluate the Christian Democrats and the worse they evaluate the Social Democrats, the Greens and the Party of Democratic Socialism (PDS). It is also reasonable that the scores for the liberals are lowest, as the party is usually seen around the middle of the scale. However, that is not our main concern here.

The key point here is that associations in the open online survey take on the highest value for *all* party feeling thermometers. Thus, our expectations are clearly met. While there are again no differences between the access panel and the offline survey, the open online survey really is the odd one out – even after controlling for the differing socio-demographic composition, which again hardly affects the results at all.

Table 4: Associations between respondents' left-right-positions and feeling thermometers for political parties (Pearson's r, results for the open online survey are shown unweighted as well as weighted)

Correlation of respondent's left–right position[b] with feeling thermometer[c]…	Face-to-face survey	Survey based on access panel	Open online survey	Significance[a] of difference between		
	(1)	(2)	(3)	1 and 2	1 and 3	2 and 3
CDU	0,49	0,43	0,62 / 0,59	n.s.	*** / ***	*** / ***
SPD	-0,31	-0,38	-0,50 / -0,51	n.s.	*** / ***	*** / ***
FDP	0,33	0,29	0,49 / 0,45	n.s.	*** / ***	*** / ***
Greens	-0,47	-0,49	-0,60 / -0,58	n.s.	*** / ***	*** / ***
PDS	-0,28	-0,31	-0,51 / -0,47	n.s.	*** / ***	*** / ***

[a] Significance levels: n.s. – not significant, *: $p<.05$, **: $p<.01$, ***: $p<.001$, based on *z*-test of differences between correlation coefficients.

[b] The wording of the question was: "People often use the terms 'left' and 'right' in politics. Using this scale from 1 to 11, where would you place yourself, if 1 stands for left and 11 stands for right?"

[c] The wording of the question was "Generally speaking, what is your opinion about the political parties? Please use this scale from –5 to +5. –5 means that you have a very poor opinion about a party, +5 means that you have a very good opinion about a party."

Another opportunity to test our propositions is to look at associations *among* feeling thermometers, since these should also follow specific patterns. Chancellor Gerhard Schröder was the Social Democratic Party's (SPD) candidate, his challenger Edmund Stoiber ran for the Christian Democrats (CDU/CSU). Hence, one would expect a strong (positive) association between the feeling thermometers for the candidate and his respective party and strong *negative* associations between opposing candidates and/or parties. However, the level of strength can again be expected to vary between the three surveys, this time mainly due to the higher involvement in politics that characterizes the participates of the open online survey. They should be more familiar, but also more polarized when it comes to evaluating candidates and parties.

Table 5: Associations among feeling thermometers for parties and chancellor candidates (Pearson's r, results for the open online survey are shown unweighted as well as weighted))

Correlated feeling thermometers[b, c]	Face-to-face survey	Survey based on access panel	Open online survey	Significance[a] of difference between		
	(1)	(2)	(3)	1 and 2	1 and 3	2 and 3
Schröder – Stoiber	-0,25	-0,52	-0,60 / -0,62	***	*** / ***	*** / ***
CDU – SPD	-0,10	-0,40	-0,53 / -0,55	***	*** / ***	*** / ***
Schröder – CDU	-0,28	-0,44	-0,56 / -0,58	***	*** / ***	*** / ***
Stoiber – SPD	-0,22	-0,49	-0,60 / -0,61	***	*** / ***	*** / ***
Stoiber – CDU	0,77	0,78	0,86 / 0,87	n.s.	*** / ***	*** / ***
Schröder – SPD	0,70	0,68	0,86 / 0,87	n.s.	*** / ***	*** / ***

[a] Significance levels: n.s. – not significant, *: $p<.05$, **: $p<.01$, ***: $p<.001$, based on *z*-test of differences between correlation coefficients.

[b] The wording of the party question was "Generally speaking, what is your opinion about the political parties? Please use this scale from –5 to +5. –5 means that you have a very poor opinion about a party, +5 means that you have a very good opinion about a party."

[c] The wording of the candidate question was "Generally speaking, what is your opinion about Gerhard Schröder and Edmund Stoiber?" Please use this scale from –5 to +5. –5 means that you have a very poor opinion about him, +5 means that you have a very good opinion about him."

Empirically, this is again confirmed. The correlations are significantly stronger for the open online survey for all cases (see Table 5) – with or without weighting. Apart from that, we also do find a puzzling result this time when it comes to comparing the offline survey and the access panel. This is *not* case for the consonant comparisons (i.e., the correlations between the two main contenders and their respective parties), where the correlations are virtually the same. However, when it comes to dissonant comparisons (i.e., correlations between opposing candidates and/or parties), the correlations are much stronger in the access panel than in the offline survey. There is a solution to this apparent puzzle, though: Both samples have a different field time. The access panel was fielded in the very final week of the campaign, while the offline survey had a much longer field time. It is known in electoral research that attitudes become more polarized in the course of an election campaign (see Schoen 2005). Empirically, this is supported, as the differences between the two surveys decrease

considerably, if we restrict the offline to those users that took part in the final week of the campaign.

5 Conclusion

As Couper (2000, p. 464) rightly pointed out: “We stand at the threshold of a new era of survey research, but how this will play is not yet clear.” What is also already very clear is that online surveys are and will be frequently used. However, their usage must be evaluated critically. Within this article, our aim was to shed some light on the question of the data quality that two types of online surveys (one based on an access panel, the other an open, unrestricted web survey) yield.

Looking at marginal distributions and associations among variables, our results provide a mixed picture. The marginal distributions stemming from an access panel, whose members were previously recruited from offline surveys, do not differ from the results obtained from traditional offline surveys of Internet users. In contrast, the results based on the open online survey are dramatically skewed: Socio-demographic, but also substantial variables are clearly biased. The picture that emerges with respect to associations among variables is also far from perfect. Due to the higher sophistication and the higher involvement in politics among Internet users, we expected associations among variables to be higher when based on open online surveys. Again, this was confirmed, although it has to be pointed out that the distortions are not as large as they are in the case of marginal distributions. And as with distortions of marginal distributions, bias in associations is not reduced substantially when data are weighted socio-demographically.

Hence, the conclusion must be that open online surveys do not yield results representative for online users (neither in terms of marginal distributions nor in terms of associations). If one wants survey results representative for Internet users s/he must carefully select a sample of online users. It is worth the effort! It does not suffice to put a questionnaire on the web and to call upon online users to participate in the survey. Using standard weighting procedures does not help either. And it is by no means clear that more sophisticated weighting mechanisms help, since the direction and the extent of self-selection bias probably depend upon the topic of the survey. Hence, general weighting procedures may be not very useful. As a result, one has to conclude that the advent of the Internet will not ease conducting serious surveys. A good survey requires good, tedious work. Simply putting a questionnaire on the web is not enough.

6. References

Achen, C.H. (1983). Towards Theories of Data. In Political Science: The State of the Discipline, A.W. Finifter (ed.). Washington, DC: American Political Science Association.

Achen, C.H. (1985). Proxy Variables and Incorrect Signs on Regression Coefficients. Political Methodology, 11, 299-316.

Alvarez, R.M., Sherman, R.P., and Van Beselaere, C. (2003). Subject Acquisition for Web-Based Surveys. Political Analysis, 11, 23-43.

Bandilla, W. (1999). WWW-Umfragen – Eine alternative Datenerhebungstechnik für die empirische Sozialforschung? In Online Research, B. Batinic, A. Werner, L. Gräf, and W. Bandilla (eds). Göttingen: Hogrefe. [In German]

Bandilla, W., Bosnjak, M., and Altdorfer, P. (2001). Effekte der Erhebungsverfahren? Ein Vergleich zwischen einer Web-basierten und einer schriftlichen Befragung zum ISSP-Modul Umwelt. ZUMA-Nachrichten, 49, 7-28. [In German]

Bartles, J. (2000). Political Awareness, Opinion Constraint and the Stability of Ideological Positions. Political Studies, 48, 467-484.

Batinic, B. (2001). Fragebogenuntersuchungen im Internet. Aachen: Shaker. [In German]

Berk, R.A. (1983). An Introduction to Sample Selection Bias in Sociological Data. American Sociological Review, 48, 386-398.

Berrens, R.P., Bohara, A.K., Jenkins-Smith, H., Silva, C., and Weimer, D. (2003). The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples. Political Analysis, 11, 1-22.

Bosnjak, M. (2002). (Non)Response bei Web-Befragungen. Aachen: Shaker. [In German]

Campbell, A., Converse, P.E., Miller, W.E., and Stokes, D.E. (1960). The American Voter. New York: Wiley.

Converse, P.E. (1964). The Nature of Belief Systems in Mass Publics. In Ideology and Discontent, D.E. Apter (ed). New York: Free Press.

Converse, P.E. (1970). Attitudes and Non-Attitudes. In: The Quantitative Analysis of Social Problems, E.R. Tufte (ed.). Reading: Addison-Wesley.

Couper, M.P. (2000). Web Surveys. Public Opinion Quarterly, 64, 464-494.

Couper, M.P., Traugott, M.W., and Lamias, M.J. (2001). Web Survey Design and Administration. Public Opinion Quarterly, 65, 230-253.

Dillman, D.A. (2000). Mail and Internet Surveys. 2nd ed. New York: Wiley.

Faas, T. (2002). www.wahlumfrage2002.de: Analysen und Ergebnisse. Bamberg: University of Bamberg (http://www.uni-bamberg.de/sowi/politik/bbp/BBP-II-11.pdf). [In German]

Faas, T. (2003). Offline rekrutierte Access Panels: Königsweg der Online-Forschung? ZUMA-Nachrichten, 53, 58-76. [In German]

Falter, J.W. and Schoen, H. (eds.) (2005). Handbuch Wahlforschung. Wiesbaden: VS Verlag für Sozialwissenschaften.

Falter, J.W., Schoen, H., and Caballero, C. (2000). Dreißig Jahre danach: Zur Validierung des Konzepts 'Parteiidentifikation' in der Bundesrepublik. In 50 Jahre Empirische Wahlforschung in Deutschland. Entwicklung, Befunde, Perspektiven, Daten,. M. Klein, W. Jagodzinski, E. Mochmann, and D. Ohr, (eds). Wiesbaden: Westdeutscher Verlag. [In German]

Groves, R.M. (1987). Research on Survey Data Quality. Public Opinion Quarterly, 51, S156-S172.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Harris Interactive (2000). 2000 Election Winners: George W. Bush … and Online Polling: Harris Interactive scores unprecedented 99% accuracy in predicting 2000 election outcome. Press release available at http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=208.

Hauptmanns, P. (1999). Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. In Online Research, B. Batinic, A. Werner, L. Gräf, and W. Bandilla (eds). Göttingen: Hogrefe. [In German]

Ilieva, J., Baron, S., and Healey, N.M. (2002). Online Surveys in Marketing Research: Pros and Cons. International Journal of Market Research, 44, 361-376.

Jacoby, W.G. (2004). Ideology in the 2000 Election: A Study in Ambivalence. In Models of Voting in Presidential Elections, H.F. Weisberg and C. Wilcox (eds). Stanford: Stanford University Press.

King, W.C. and Miles, E.W. (1995). A Quasi-Experimental Assessment of the Effect of Computerizing Noncognitive Paper-and-Pencil Measurements: A Test of Measurement Equivalence. Journal of Applied Psychology, 80, 643-651.

Klein, M., Jagodzinski, W., Mochmann, E., and Ohr, D. (eds.) (2000). 50 Jahre empirische Wahlforschung in Deutschland, Wiesbaden: Westdeutscher Verlag. [In German]

Kuklinski, J.H., Metlay, D.S., and Kay, W.D. (1982). Citizen Knowledge and Choices on the Complex Issue of Nuclear Energy. American Journal of Political Science, 26, 615-642.

Mehta, R. and E. Sivadas (1995). Comparing Response Rates and Response Content in Mail Versus Electronic Surveys. Journal of Market Research Society, 37, 429-440.

Nicholson, T., White, J.B., and Duncan, D. (1998). Drugnet: A Pilot Study of Adult Recreational Drug Use via the WWW. Substance Abuse, 19, 109-121.

Norpoth, H. and Lodge, M. (1985). The Difference between Attitudes and Nonattitudes in the Mass Public: Just Measurements. American Journal of Political Science, 29, 291-307.

Perspektive Deutschland (2005a). Projektbericht Perspektive-Deutschland 2004/2005 – Die größte gesellschaftspolitische Umfrage. Düsseldorf: McKinsey (http://perspektive-deutschland.de/files/presse_2005/Projektbericht_Perspektive_Deutschland_2004_05.pdf). [In German]

Perspektive Deutschland (2005b). Die Methodik der Umfrage: Wie repräsentativ sind die Ergebnisse von Perspektive-Deutschland? http://perspektive-deutschland.de/files/presse_2005/pd4-Die_Methodik_der_Umfrage.pdf. [In German]

Ray, N., Griggs, K., and Tabor, S. (2001), Web Based Survey Research Workshop. WDSI, April (http://telecomm.boisestate.edu/research/)

Schaefer, R. and Dillman, D.A. (1998). Development of A Standard E-mail Methodology: Results of An Experiment. Public Opinion Quarterly, 62, 378-397.

Schnell, R. (1991). Wer ist das Volk? Kölner Zeitschrift für Soziologie und Sozialpsychologie, 43, 106-137. [In German]

Schoen, H. (2004). Online-Umfragen – schnell, billig, aber auch valide? Ein Vergleich zweier Internetbefragungen mit persönlichen Interviews zur Bundestagswahl 2002. ZA-Information, 54, 27-52. [In German]

Schoen, H. (2005). Candidate Orientations in Election Campaigns. An Analysis of the German Federal Election Campaigns from 1980 to 1998. Encounters: Political Science in Translation 1, 20-39 (http://www.sfu.ca/igs/APR/encounters/V1N1/german.pdf).

Schonlau, M., Fricker, R.D., Jr., and Elliott, M.N. (2002). Conducting Research Surveys via E-mail and the Web. Santa Monica, CA: RAND.

Schuldt, B.A. and Totten, J.W. (1999). E-mail Surveys: What We've Learned Thus Far. Quirk's Marketing Research Review, July (www.quirks.com)

Schuman, H. and Presser, S. (1996). Questions and Answers in Attitude Surveys. 2nd ed. Thousand Oaks: Sage.

Sheehan, K.B. and McMillan, S.J. (1999). Response Variation in E-Mail Surveys: An Exploration. Journal of Advertising Research, 39, 45-54.

Sniderman, P.M., Brody, R.A., and Tetlock, P.E. (1991). Reasoning and Choice. Explorations in Political Psychology. Cambridge: Cambridge University Press.

Stanton, J.M. (1998). An Empirical Assessment of Data Collection Using the Internet. Personnel Psychology, 51, 709-726.

Taylor, H. (2000). Does Internet Research Work? International Journal of Market Research, 42, 51-63.

Terhanian, G. and Bremer, J. (2002). A Multi-Method Approach for Reducing Error in Internet-Based Surveys of Non-Probability Samples. Paper presented at the 98th Annual Meeting of the American Political Science Association, Boston August 26 – September 1.

Vehovar, V., Batagelj, Z., Manfreda, K.L., and Zaletel, M. (2002). Nonresponse in Web Surveys. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: Wiley.

Weible, R. and Wallace, J. (1998). The Impact of Internet on Data Collection. Market Research, 10, 19-23.

Zaller, J.R. (1990). Political Awareness, Elite Opinion Leadership, and the Mass Survey Response. Social Cognition, 8, 125-153.

Zaller, J.R. (1992). The Nature and Origins of Mass Opinion. Cambridge: Cambridge University Press.

Received March 2004

Revised July 2005