

Dutch Parliamentary Election Study 2017

A comparison of three different survey modes

Roderik Rekker
Tom van der Meer
Wouter van der Brug



UNIVERSITEIT VAN AMSTERDAM

Amsterdam, February 2020

University of Amsterdam (UvA)
Dutch Electoral Research Foundation (SKON)

Table of contents

1: Introduction	p. 4
2: Research description	p. 6
3. Unit non-response	p. 7
4: Representativeness	p. 9
5: Item non-response	p. 12
6: Means	p. 14
7: Variances	p. 17
8: Time trends	p. 18
9: Test-retest reliability	p. 20
10: Criterion validity	p. 22
11: Multiple regression	p.24
12: Conclusion and recommendations	p. 26

1. Introduction

Ever since 1971, the Dutch Parliamentary Election Study (DPES) has been conducted using *face-to-face interviews* and a *fresh probability sample*. However, survey practices have changed substantially since then. In most modern surveys, home visits by interviewers have been replaced by online questionnaires (i.e., *web-based interviewing*). In addition, many surveys nowadays rely on *ongoing internet panels* rather than the recruitment of a fresh sample of respondents for each new wave. Web-based interviewing and internet panels can offer substantial cost reductions compared to face-to-face interviews with fresh probability samples. In addition, they can also provide scientific advantages. For example, web-based surveys provide new opportunities to conduct survey experiments and ongoing internet panels allow researchers to follow the same respondents over time. One of the crucial advantages of panel studies in electoral research is that they enable researchers to observe over time changes in political attitudes, behavior and perceptions at the level of individual voters.

Despite the clear merits of web-based interviewing and internet panels, face-to-face interviews with fresh samples can however still be considered a ‘gold standard’ with regard to representativeness and data quality. Web-based surveys and internet panels are for example often less representative for the population that they examine due to lower response rates and panel attrition. To examine how a switch to web-based interviewing and an internet panel would affect representativeness and data quality in the case of the Dutch Parliamentary Election Study, the DPES round of 2017 combined three different survey modes:

- **CAPI (Computer Assisted Personal Interviewing):** *Face-to-face interviews of a fresh probability sample of the Dutch adult population.*
- **CAWI (Computer Assisted Web-based Interviewing):** *Web-based interviews of a fresh probability sample of the Dutch adult population.*
- **Panel:** *Web-based interviews of an ongoing internet panel that is originally based on a random probability sample of Dutch households.*

This report will compare if and how these different interview modes and different ways of drawing samples have produced differential outcomes. More specifically, this report aims to answer the following questions:

- **Q1a.** *Does web-based interviewing produce different results compared to face-to-face interviewing?*
- **Q1b.** *If so, does web-based interviewing improve or deteriorate the quality of the data compared to face-to-face interviewing?*
- **Q2a.** *Does an ongoing internet panel produce different results compared to a fresh probability sample?*
- **Q2b.** *If so, does an ongoing internet panel improve or deteriorate the quality of the data compared to a fresh probability sample?*

The second chapter of this report will first describe the data collection procedures in more detail. The subsequent chapters will then compare unit non-response (chapter 3), representativeness (chapter 4),

item non-response (chapter 5), means (chapter 6), variances (chapter 7), time trends (chapter 8) test-retest reliability (chapter 9), criterion validity (chapter 10), and estimates from multiple regression models (chapter 11) across the three survey modes. The final chapter will provide conclusions and recommendations.

2. Research description

The three different survey modes of the DPES 2017 were conducted as depicted in Figure 1. A first group of respondents completed the survey using ‘computer-assisted personal interviewing’ (CAPI). In this survey method, fieldworkers brought home visits to the respondents to read the questions to them and to record their answers on a tablet computer. Respondents were selected using a random probability sample of all eligible Dutch voters that was provided by Statistics Netherlands (CBS). The CAPI fieldwork was executed by research agency Kantar Public.

A second group of respondents was interviewed using ‘computer-assisted web interviewing’ (CAWI). No interviewer was present in this survey mode, as respondents completed the questionnaire online. As for the CAPI, the CAWI-respondents were selected by Statistics Netherlands (CBS) as a random probability sample of Dutch voters and the data collection was conducted by research agency Kantar Public.

A third group of participants consisted of members of the ongoing ‘LISS-panel’ (Langlopende Internet Studies voor de Sociale Wetenschappen). The LISS-panel is managed by research agency CentERdata and consists of 5,000 households. These households were selected on the basis of probability sampling by Statistics Netherlands (CBS) to obtain a nationally representative sample. The members of the LISS-panel participate in regular online questionnaire over an extended period of time. The DPES 2017 was likewise administered in the LISS-panel using a web-based survey (CAWI).

After completing the *main questionnaire*, respondents were invited to complete a *supplementary questionnaire*. For the CAPI-respondents, the interviewer left a paper drop-off questionnaire after every interview. Respondents were asked to complete this questionnaire at their own convenience and return it by post. In addition, a very small number ($N = 8$) of the CAPI respondents completed the supplementary questionnaire online. The LISS-respondents were also administered this supplementary questionnaire, but in the form of a second online form. The CAWI-respondents who were not part of the LISS-panel were not invited for the supplementary questionnaire.

Alongside the main questionnaire and supplementary questionnaire, Statistics Netherlands (CBS) contributed a third source of data in the form of demographic information on respondents, such as their municipality’s degree of urbanization.

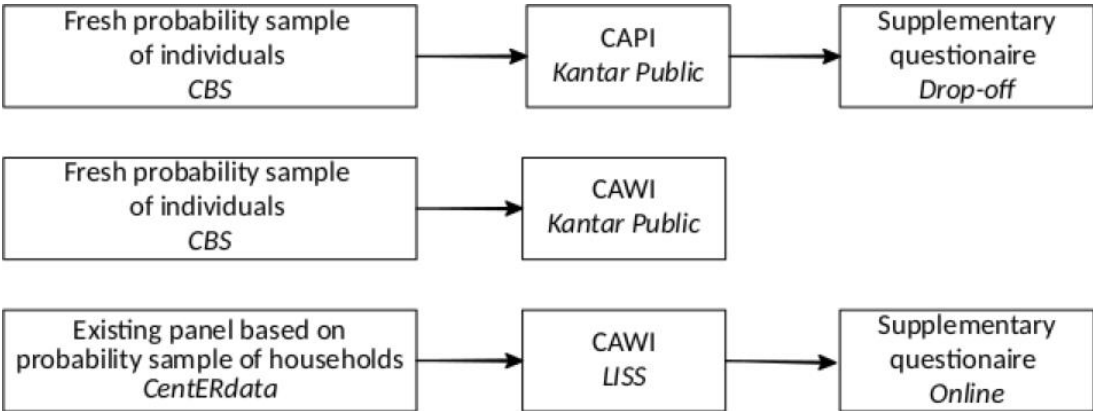


Figure 1: The three parts of the DPES 2017.

3. Unit non-response

Key findings

- *Face-to-face interviewing yielded a better response rate than web-based interviewing, but the differences are modest.*

A known drawback of web-based interviewing is that response rates tend to be lower than those of face-to-face interviews. When people are contacted by an interviewer for a home visit, they often seem more inclined to participate in a survey than when they are asked to complete an online questionnaire. This chapter will therefore examine to what extent response rates differed between survey modes in the DPES 2017.

To better understand these differences, it is important to keep in mind that the strategy to approach respondents differed slightly between the three survey modes. The CAPI-respondents were approached in four stages. In the first stage, 1900 respondents received an introduction letter and were subsequently approached at least three times by an interviewer. Respondents received a voucher of 15 euro after a completed interview. In the second round, over 700 respondents received a card that again notified them that they would be contacted by an interviewer and that the incentive was raised to a 20-euro voucher. In the third round, about 80 respondents were called and the incentive was raised to a voucher of 25 euro. In the fourth round, about 385 respondents received a card that notified them of a final opportunity to participate and receive a 25-euro voucher.

The CAWI-respondents were also contacted in four subsequent stages. In the first stage, respondents received a letter (i.e., by post) that explained how they could participate in the online survey. The incentive in this round was a voucher of 10 euro. In the second stage, respondents received a card with a second invitation. The incentive was unchanged in this round. In the third stage, respondents received a reminder letter and the incentive was raised to 15 euro. In the fourth round, all respondents for whom the telephone number was known were called. They were now given the opportunity to receive a direct link to the questionnaire by e-mail and the incentive was raised to 20 euro. In sum, there are slight differences in the procedure and incentives between the CAPI and the CAWI mode that may have affected response rates. Therefore, some caution is warranted in attributing differences in response rates to the survey modes themselves. The panel-respondents were invited to participate through an e-mail invitation, which is the usual method of approaching participants of the LISS-panel.

Table 1 displays the response rates in each of the three survey modes. The share of respondents that agreed to participate was 49.1% in the CAPI-mode and a somewhat lower 45.7% in the CAWI-mode. After removing respondents who could not be identified correctly (i.e., the individual who agreed to participate was not the person who was selected for the sample), this number dropped to 48.8% in the CAPI-mode and 44.4% in the CAWI-mode. In CAWI-surveys, there are however usually some respondents who close the questionnaire before finishing it. As such, the number of respondents that completed the questionnaire until the end was 40.3% in the CAWI-mode, against a substantially higher 48.8% in the CAPI-mode. In the panel-mode, a much higher share of 78.1% of the approached respondents completed the entire questionnaire. However, this number is not directly comparable to those of the CAPI- and CAWI-mode because respondents who either refused to participate in the LISS-panel (i.e., non-response) or dropped out after a while (i.e., panel attrition) were already excluded from selection for the DPES 2017.

In conclusion, the CAPI-mode yielded a somewhat better response rate than the CAPI-mode, particularly when we look at the number of respondents who completed the entire questionnaire. This pattern is in line with the experience from many other surveys. However, the difference in response rate between the survey mode was of a relatively modest magnitude (8.5 percentage points; 21.1 percent). Potentially, this difference can be further reduced in the future by intensifying the strategy to approach respondents. The next chapter will examine to what extent these differential response rates have affected the representativeness of the sample in all three survey modes.

Table 1: Response rate.

	CAPI	CAWI	Panel
Selected respondents	1900	1600	2243
Positive response received	932 (49,1%)	731 (45,7%)	1790 (79,8%)
Correctly identified respondents	927 (48,8%)	711 (44,4%)	1790 (79,8%)
Completed main questionnaire until end	927 (48,8%)	645 (40,3%)	1751 (78,1%)
Completed supplementary questionnaire	723 (38,1%)	NA	1180 (52,6%)

4. Representativeness

Key findings

- *Web-based interviewing yielded an overall representativeness that was at least as good as that of face-to-face interviewing, but additional measures are advisable to reach the oldest age cohort.*
- *Recruiting respondents from the LISS-panel resulted in a slightly less representative sample compared to using a fresh probability sample.*

A core aim of the DPES has always been to provide a sample that is representative of the Dutch electorate. To this end, the DPES has used a fresh probability sample that was drawn by Statistics Netherlands (CBS) in most of its rounds. This means that every Dutch voter has an equal chance of being selected and that disparities between the population and the sample can only arise due to *selective non-response*, which is the phenomenon that people who refuse participation in a survey usually differ from those who participate on key characteristics. For example, people who are not interested in politics and do not vote also tend to show less interest to participate in political surveys.

In the DPES 2017, the CAPI-mode yielded a somewhat better response rate than the CAWI-mode (see chapter 3). This also means that there is a stronger potential for selective non-response in the CAWI-mode. Furthermore, it often proves harder to reach older voters with online surveys because they are still less likely to use the Internet. Using an online internet panel furthermore introduces an additional source of sample bias in the form of *selective panel attrition*, which is the process through which respondents with certain characteristics are more likely to quit their participation in an ongoing panel after a while. Although the sample of the LISS-panel was originally recruited with a probability sample from Statistics Netherlands, respondents who dropped out of the panel over the years had to be replaced. To reach people without a computer or Internet access, the LISS-panel gives respondents the possibility to lend an easy-to-use computer with free Internet-access.

Table 2 compares the distribution of demographic variables and vote choice across the three survey modes with the population figures as provided by Statistics Netherlands (CBS). It offers information on the relative and the absolute distortion. Surprisingly, we can see that the CAWI-mode featured a slightly better average representativeness on both the full set of categories (1.8%) as well as on those measuring vote choice (2.1%) than the CAPI-mode (2.1% and 2.3% distortion, respectively). As such, the lower response rate of the CAWI-mode (see chapter 3) did not result in a lesser overall representativeness. The distortion of the CAWI-mode was strongly driven by an underrepresentation of voters over age 75; the distortion of the CAPI-mode more strongly by the underrepresentation of urban voters. Regarding vote choice we find various differences, and a consistent underrepresentation of (NB: reported) non-voters in all modes. This underrepresentation does not solely indicate a sampling problem, but also respondents' likelihood to overreport their turnout (Dahlggaard et al. 2019).

The results in Table 2 also show that the sample that was recruited from an ongoing internet panel revealed a substantially lesser representativeness (average distortion of 3.1% on all traits, and 2.6% on vote choice) than the CAPI and CAWI-samples that were recruited from a fresh probability sample. Especially young voters, single voters, and voters from urban areas were much more underrepresented in this survey mode. This is very likely a result of the selective panel attrition that inevitably occurs in ongoing internet panels. Interestingly, older voters were not underrepresented in the panel-mode.

This indicates that the efforts of the LISS-panel to include groups that are usually harder to reach with online surveys (e.g., by providing easy-to-use-computers) have borne fruit.

In sum, these results indicate that web-based interviewing is a suitable alternative for face-to-face interviewing when it comes to representativeness. It is however advisable to include measures to counter the strong underrepresentation of older (75+) voters in the CAWI-mode. The better representation of this group in the LISS-panel indicates that such measures can be effective. Speculatively, it also seems conceivable that the representation of older voters in online surveys will improve with time as the penetration of Internet-access among older citizens increases further (e.g., because of generational replacement). The representativeness of the sample in the panel-mode was however somewhat less satisfactory. As such, it appears that recruiting respondents from an ongoing internet panel resulted in a slightly less representative sample than one would obtain by using a fresh probability sample.

Table 2: Representativeness.

	Population	CAPI			CAWI			Panel		
		Response	Relative Distortion	Absolute Distortion	Response	Relative Distortion	Absolute Distortion	Response	Relative Distortion	Absolute Distortion
Vote choice										
VVD	17.4%	19.3%	110.9%	1.9%	20.0%	114.9%	2.6%	17.7%	101.7%	0.3%
PVV	10.6%	8.1%	76.4%	-2.5%	10.8%	101.9%	0.2%	8.9%	84.0%	-1.7%
CDA	10.1%	12.7%	125.7%	2.6%	11.6%	114.9%	1.5%	15.2%	150.5%	5.1%
D66	10.0%	15.1%	151.0%	5.1%	15.5%	155.0%	5.5%	13.2%	132.0%	3.2%
GroenLinks	7.4%	10.5%	141.9%	3.1%	10.1%	136.5%	2.7%	11.1%	150.0%	3.7%
SP	7.4%	8.8%	118.9%	1.4%	8.3%	112.2%	0.9%	10.7%	144.6%	3.3%
PvdA	4.7%	8.2%	174.5%	3.5%	6.4%	136.2%	1.7%	8.0%	170.2%	3.3%
ChristenUnie	2.8%	5.2%	185.7%	2.4%	4.9%	175.0%	2.1%	4.7%	167.9%	1.9%
Partij voor de Dieren	2.6%	3.5%	134.6%	0.9%	3.5%	134.6%	0.9%	4.0%	153.8%	1.4%
50Plus	2.5%	2.2%	88.0%	-0.3%	3.0%	120.0%	0.5%	2.9%	116.0%	0.4%
SGP	1.7%	2.2%	129.4%	0.5%	1.2%	70.6%	-0.5%	1.4%	82.4%	-0.3%
DENK	1.7%	0.7%	41.2%	-1.0%	1.0%	58.8%	-0.7%	0.3%	17.6%	-1.4%
Forum voor Democratie	1.5%	2.2%	146.7%	0.7%	1.9%	126.7%	0.4%	1.3%	86.7%	-0.2%
Other party or blank	1.6%	1.5%	93.8%	-0.1%	1.4%	87.5%	-0.2%	1.8%	112.5%	0.2%
Did not vote	18.1%	9.0%	49.7%	-9.1%	7.3%	40.3%	-10.8%	7.5%	41.4%	-10.6%
Age										
18-24	10.6%	10.0%	94.3%	-0.6%	9.4%	88.7%	-1.2%	6.5%	61.3%	-4.1%
25-34	14.6%	11.9%	81.5%	-2.7%	12.8%	87.7%	-1.8%	10.8%	74.0%	-3.8%
35-44	14.8%	14.2%	95.9%	-0.6%	17.7%	119.6%	2.9%	13.0%	87.8%	-1.8%
45-54	19.1%	18.9%	99.0%	-0.2%	18.0%	94.2%	-1.1%	16.3%	85.3%	-2.8%
55-64	17.2%	21.3%	123.8%	4.1%	20.1%	116.9%	2.9%	21.5%	125.0%	4.3%
65-74	14.1%	14.8%	105.0%	0.7%	16.7%	118.4%	2.6%	22.5%	159.6%	8.4%
75+	9.6%	9.3%	96.9%	-0.3%	5.2%	54.2%	-4.4%	9.4%	97.9%	-0.2%
Gender										
Male	49.3%	51.6%	104.7%	2.3%	49.2%	99.8%	-0.1%	47.8%	97.0%	-1.5%
Female	50.7%	48.4%	95.5%	-2.3%	50.8%	100.2%	0.1%	52.2%	103.0%	1.5%
Urbanization										
Very high	22.4%	17.4%	77.7%	-5.0%	22.2%	99.1%	-0.2%	14.3%	63.8%	-8.1%
High	30.5%	31.6%	103.6%	1.1%	31.9%	104.6%	1.4%	25.8%	84.6%	-4.7%
Medium	16.9%	19.0%	112.4%	2.1%	16.3%	96.4%	-0.6%	22.9%	135.5%	6.0%
Low	21.3%	22.7%	106.6%	1.4%	20.8%	97.7%	-0.5%	21.5%	100.9%	0.2%
Very low	8.8%	9.4%	106.8%	0.6%	8.8%	100.0%	0.0%	15.5%	176.1%	6.7%
Region										
North	10.4%	12.5%	120.2%	2.1%	8.5%	81.7%	-1.9%	11.1%	106.7%	0.7%
East	21.3%	21.8%	102.3%	0.5%	22.8%	107.0%	1.5%	22.2%	104.2%	0.9%
West	46.6%	43.4%	93.1%	-3.2%	47.4%	101.7%	0.8%	42.7%	91.6%	-3.9%
South	21.7%	22.3%	102.8%	0.6%	21.2%	97.7%	-0.5%	24.1%	111.1%	2.4%
Marital state										
Married	50.2%	53.5%	106.6%	3.3%	55.8%	111.2%	5.6%	56.0%	111.6%	5.8%
Divorced	9.8%	9.7%	99.0%	-0.1%	7.0%	71.4%	-2.8%	11.2%	114.3%	1.4%
Widowed	6.1%	5.2%	85.2%	-0.9%	4.5%	73.8%	-1.6%	6.2%	101.6%	0.1%
Single	34.0%	31.6%	92.9%	-2.4%	32.7%	96.2%	-1.3%	26.6%	78.2%	-7.4%
Country of origin										
Dutch origin	82.9%	88.1%	106.3%	5.2%	84.7%	102.2%	1.8%	85.2%	102.8%	2.3%
Western origin	7.4%	6.8%	91.9%	-0.6%	8.7%	117.6%	1.3%	9.4%	127.0%	2.0%
Non-western origin	9.7%	5.1%	52.6%	-4.6%	6.7%	69.1%	-3.0%	5.5%	56.7%	-4.2%
Average distortion:			2.1%			1.8%			3.1%	
Avg. distort. vote choice:			2.3%			2.1%			2.5%	

Note:

Green: Less than 2.5%.

Orange: 2.5% - 5%.

Red: More than 5%.

5. Item non-response

Key findings

- *Web-based interviewing produced a higher number of ‘Don’t know’ and ‘Won’t say’ answers compared to face-to-face interviewing.*
- *Face-to-face interviewing generated more responses in the center categories of the scales compared to web-based interviewing.*

Whereas some respondents refuse to participate in the entire survey (i.e., unit non-response, see chapter 3), others are unable or unwilling to answer specific questions. This is known as *item non-response*. In the DPES 2017, nearly all questions included ‘don’t know’ and ‘won’t say’ as response categories. However, there is always a risk that some respondents answer ‘don’t know’ as a fast and easy way to reach the end of the survey, even if they would actually be able to answer the question. Reversely, some respondents who have no idea what the question is about may artificially try to produce an answer to make a better impression. Whereas the former seems more likely in web-based interviewing in which some respondents may want to rush to the end of the survey, the latter may occur more often in face-to-face interviews because some respondents may want to make a favorable impression on the interviewer.

Table 3 displays the amount of ‘don’t know’ and ‘won’t say’ answers on key variables in each of the three survey modes. The number of responses in the center category (e.g., 3 on a scale from 1 through 5) are also displayed because this may be the most likely response for respondents who do not know how to answer a question, but are afraid to admit it.

As expected, the results indicate that respondents in the CAWI-mode were more likely than respondents in the CAPI-mode to answer a question with ‘don’t know’ (i.e., 6.7% versus 3.0%) or ‘won’t say’ (1.3% versus 0.4%). Reversely, the CAPI-mode (24.0%) generated somewhat more responses in the center categories of the scales compared to the CAWI-mode (21.8%). The panel-mode yielded very similar results to the CAWI-mode. The more frequent use of the center category in web-based surveys is not consistently and strongly related to a lesser use of item non-response. The more frequent use of the center category could potentially suppress the variation in scores, which we examine in chapter 7. Fortunately, however, the vast majority of respondents provided substantive answers to the questions regardless of survey method. As such, item non-response does not seem to be a reason for great concern in any of the sampling and interview modes.

Table 3: Item non-response.

Core Variables	Don't Know			Won't Say			Center Category		
	CAPI	CAWI	Panel	CAPI	CAWI	Panel	CAPI	CAWI	Panel
V024: Interested in politics	0.1%	1.1%	0.5%	0.0%	0.0%	0.1%	68.4%	67.9%	63.2%
V083: Satisfaction with govern.	0.4%	2.0%	3.2%	0.0%	0.3%	0.2%	35.6%	37.8%	36.9%
V098: Income differences - p. resp.	1.6%	5.2%	12.5%	0.1%	1.7%	1.6%	23.9%	23.3%	20.6%
V108: European unification: - p. resp.	3.9%	9.3%	15.0%	0.0%	1.5%	1.4%	21.7%	18.1%	18.6%
V118: Foreigners - p. resp.	0.6%	4.2%	8.5%	0.3%	1.1%	1.5%	22.2%	17.7%	17.9%
V133: Left-right self-rating	4.1%	6.8%	6.7%	0.4%	2.8%	1.3%	21.0%	14.2%	13.7%
V255: External efficacy (A)	4.1%	10.8%	18.8%	0.5%	0.4%	0.8%	NA	NA	NA
V258 and V259: Internal efficacy (A)	1.0%	6.0%	5.4%	0.1%	0.9%	0.8%	NA	NA	NA
V260 until V263: Political cynicism (A)	1.1%	5.4%	5.8%	0.5%	0.8%	1.3%	28.2%	38.8%	32.6%
Sympathy Scores									
V200: Sympathy score: VVD	3.3%	7.2%	7.0%	0.8%	1.7%	1.4%	14.7%	11.7%	8.7%
V201: Sympathy score: PvdA	4.0%	7.6%	7.3%	0.8%	1.7%	1.6%	19.8%	17.3%	11.7%
V202: Sympathy score: PVV	2.7%	6.2%	6.6%	0.6%	1.7%	1.5%	9.8%	4.8%	4.1%
V203: Sympathy score: CDA	4.5%	7.9%	7.8%	0.8%	1.8%	1.6%	22.2%	17.3%	13.9%
V204: Sympathy score: SP	6.6%	10.3%	8.9%	0.6%	1.8%	1.5%	18.2%	15.0%	10.1%
V205: Sympathy score: D66	5.4%	9.4%	8.0%	0.6%	1.5%	1.6%	16.2%	12.8%	14.0%
V207: Sympathy score: GroenLinks	5.0%	7.9%	8.5%	0.6%	1.7%	1.5%	13.6%	9.0%	8.8%
Average:	3.0%	6.7%	8.2%	0.4%	1.3%	1.2%	24.0%	21.8%	19.6%

Note:

Don't know and won't say:

Green: Less than 5%.

Orange: Between 5% and 10%.

Red: More than 10%.

A: Average across items.

NA: Response scale without center category.

Don't know includes 'does not know party' for sympathy scale.

6. Means

Key findings

- *There were small to modest differences in mean scores and variances between web-based interviewing and face-to-face interviewing.*
- *Although both modes yielded rather similar mean scores, there were some small differences between respondents that were recruited from an ongoing internet panel and respondents from a fresh probability sample.*

The differences between the sampling methods have the potential to affect substantive outcomes. In addition, the answers that respondents give may also depend on the interview mode itself. When interacting with an interviewer, some respondents may for example answer questions in a more socially desirable manner than they would in an online survey. Therefore, the mean scores on variables may differ between the three components of the DPES.

Table 4 displays differences in mean scores between the interview modes. The “fresh sample CAWI-mode” serves as a reference point, to which both the CAPI-interview mode and other sampling method are compared in a regression analysis. The dependent variables are the standardized scores (i.e., z-scores) on key variables. As such, an effect of +0.10 for example means that the z-score of the variable was 0.10 higher in the CAPI-mode or panel-mode compared to the CAWI mode. Four subsequent regression models were specified for each key variable. The first shows the raw differences between the survey modes without any control variables. The second model shows differences after controlling for demographic characteristics: gender, age, educational level, urbanization, part of country, country of origin, and marital status. The third model then displays differences between the survey modes after controlling for both demographic characteristics and vote choice. Importantly, this means that the second and the third model tell us if differences between the survey modes can be remedied by using survey weights. The fourth model finally controlled for demographic characteristics and vote choice, as well as respondents’ scores on a 5-item *social desirability scale*. This scale includes items that respondents tend to give socially desirable answers to, even if those do not reflect reality. An example of an item is “I am always courteous, even to people who are disagreeable.” The other four items can be found in the DPES 2017 codebook. An F-Test revealed that CAPI-respondents indeed scored higher on this scale (+0.18; $p < .001$) than respondents in the CAPI-mode and the panel-mode. This confirms that this scale can capture at least some of respondents’ tendency to provide socially desirable answers in face-to-face interviews.

The results show that there were small to modest differences in mean scores between respondents in the CAPI and CAWI-mode. As expected, respondents in the CAPI mode gave more optimistic and socially desirable answers than CAWI-respondents on most key variables. The second and third regression model furthermore indicate that these differences were almost completely unaltered by controlling for demographic characteristics and vote choice. Importantly, this implies that differences between the survey modes cannot be remedied by using survey weights. This also suggests that differences between CAPI and CAWI-respondents are unlikely to be caused entirely by differences in the composition of the samples (i.e., see chapter 4). Instead, differences between the survey modes may likely be attributed to interviewer mode effects, in particular the natural tendency to give more socially desirable answers in CAPI-interviews. Nonetheless, the fourth regression model revealed that the differences between respondents in the CAPI and the CAWI-mode could not be explained by their scores on the social desirability scale. This should however not be taken as final evidence that social

desirability did not play a role in these differences. Although CAPI-respondents scored higher on the social desirability scale than CAWI-respondents, it is very possible that this scale did not capture all differences in socially desirable response tendencies between the survey modes.

Although the CAWI and panel-mode yielded rather similar mean scores, the results in Table 4 also show some small differences. Because both groups were interviewed using web surveys, these differences can likely be attributed to the differential sample composition in both modes (see chapter 4). Alternatively, panel members may have scored differently due to their greater experience with filling out surveys. The third and fourth regression model indicate that the differences in mean scores between the CAWI and panel-mode can be reduced by controlling for demographic variables and vote choice, but only to a very limited extent. As such, survey weights may do some (but not much) good in making scores from both modes comparable.

To conclude, there were small to modest differences in mean scores between web-based interviewing and face-to-face interviewing. Respondents who were interviewed by an interviewer quite consistently gave more optimistic and socially desirable answers than respondents who filled out an online questionnaire. A plausible explanation for these differences is the well-known tendency to give more socially desirable answers in face-to-face interviews (e.g., Tourangeau c.s. 2000). Differences in mean scores between respondents who were recruited from a fresh probability sample and respondents who were recruited from an ongoing internet panel were smaller and less frequent. However, those differences are likely in part a result from the lesser representativeness of the panel sample, which implies that the mean scores from fresh the sample should be considered more trustworthy. The differences in mean scores between survey modes could not be reduced by controlling for demographic characteristics and vote choice, which indicates that survey weights cannot be used to make scores comparable.

Table 4: Means and variances.

Core Variables	CAPI Mean Effect of face-to-face interviewing on standardized scores				CAPI Variance Effect of face-to-face interviewing on absolute standardized scores				Panel Mean Effect of using panel on standardized scores				Panel Variance Effect of using panel on absolute standardized scores			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
V024: Interested in politics (R)	+0.05	+0.06	+0.06	+0.06	+0.01	+0.01	-0.02	-0.02	-0.03	-0.05	-0.04	-0.04	+0.10**	+0.09*	+0.08	+0.08
V083: Satisfaction with govern. (R)	+0.18***	+0.20***	+0.15**	+0.15**	-0.01	-0.00	+0.01	+0.01	+0.01	+0.01	+0.02	+0.02	-0.02	+0.00	-0.01	-0.01
V098: Income differences - p. resp.	+0.04	-0.00	-0.00	-0.01	-0.08**	-0.09**	-0.09**	-0.09**	+0.17***	+0.11*	+0.07	+0.08	+0.02	+0.03	+0.03	+0.04
V108: European unification: - p. resp.	-0.06	-0.11*	-0.06	-0.06	-0.05	-0.05	-0.03	-0.03	+0.15**	+0.13*	+0.12**	+0.13**	+0.00	-0.00	+0.00	+0.00
V118: Foreigners - p. resp.	-0.17***	-0.21***	-0.17***	-0.18***	-0.04	-0.04	-0.04	-0.04	+0.06	+0.02	+0.02	+0.02	-0.01	-0.02	-0.02	-0.02
V133: Left-right self-rating	+0.00	-0.02	+0.01	+0.01	-0.15***	-0.15***	-0.15***	-0.15***	-0.03	-0.04	+0.00	+0.01	+0.03	+0.01	+0.02	+0.02
V255: External efficacy	+0.06	+0.11*	+0.08	+0.07	-0.07***	-0.07***	-0.07**	-0.07**	-0.17**	-0.16**	-0.14*	-0.14*	+0.04*	+0.03	+0.03	+0.03
V258 and V259: Internal efficacy	-0.10	-0.04	-0.04	-0.04	+0.10**	+0.09**	+0.09**	+0.08*	-0.07	-0.05	-0.04	-0.03	+0.07*	+0.10**	+0.10**	+0.10**
V260 until V263: Political cynicism	-0.22***	-0.25***	-0.23***	-0.22***	.01	.02	.03	.02	-0.04	-0.03	-0.04	-0.04	+0.01	+0.05	+0.05	+0.05
Sympathy Scores	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
V200: Sympathy score: VVD	+0.12*	+0.12*	+0.11*	+0.11*	-0.16***	-0.16***	-0.15***	-0.16***	+0.12*	+0.11*	+0.15*	+0.14**	+0.02	+0.01	+0.01	+0.01
V201: Sympathy score: PvdA	+0.18***	+0.21***	+0.14***	+0.14***	-0.17***	-0.17***	-0.17***	-0.16***	+0.03	+0.03	+0.01	+0.01	+0.08**	+0.09**	+0.09**	+0.10**
V202: Sympathy score: PVV	+0.02	-0.01	+0.06	+0.06	-0.11***	-0.12***	-0.09***	-0.10***	+0.00	+0.00	+0.04	+0.04	+0.07**	+0.07*	+0.09***	+0.09***
V203: Sympathy score: CDA	+0.09	+0.08	+0.07	+0.07	-0.17***	-0.18***	-0.18***	-0.18***	+0.17***	+0.13*	+0.12*	+0.12*	+0.09**	+0.09**	+0.07*	+0.08*
V204: Sympathy score: SP	+0.13**	+0.11*	+0.10*	+0.09*	-0.16***	-0.16***	-0.16***	-0.16***	+0.05	+0.04	-0.01	-0.01	+0.10**	+0.10**	+0.11***	+0.10**
V205: Sympathy score: D66	+0.07	+0.10*	+0.06	+0.05	-0.19***	-0.20***	-0.19***	-0.19***	+0.12*	+0.12*	+0.14**	+0.13**	+0.04	+0.04	+0.05	+0.05
V207: Sympathy score: GroenLinks	+0.09	+0.11*	+0.07	+0.07	-0.19***	-0.20***	-0.18***	-0.18***	+0.13**	+0.14*	+0.10*	+0.11*	+0.07*	+0.07*	+0.10**	+0.10**

Note:

Reference category: CAWI with fresh probability sample.

1: Difference.

2: Difference after controlling for gender, age, educational level, urbanization, part of country, country of origin, and marital status.

3: Difference after controlling for gender, age, educational level, urbanization, part of country, country of origin, marital status, and vote choice.

4: Difference after controlling for gender, age, educational level, urbanization, part of country, country of origin, marital status, vote choice, and social desirability score.

*: p < .05

** : p < .01

***: p < .001

Green: 0.00 through 0.07; -0.00 through -0.07

Orange: 0.08 through 0.14; -0.08 through -0.14

Red: 0.15 or higher; -0.15 or lower

(R): Reversed scored to facilitate interpretation.

7. Variances

Key findings

- *There were small to modest differences in variances between web-based interviewing and face-to-face interviewing. Web-based interviewing produced larger variances than face-to-face interviews, which (all else being equal) can be considered an advantage.*
- *The variances of respondents that were recruited from an ongoing internet panel were mostly similar to those of respondent from a fresh probability sample.*

Interview and sampling modes may not only affect the mean score of all respondents in general, but also specifically the scores of some individuals or groups. As a result, the overall amount of variation in scores (i.e., the variance) may also differ between survey modes. Unlike differential mean scores, differences in variances can alter the magnitude of differences between groups and the strength of associations. Because researchers are usually interested in such associations, differences in variances between survey modes are arguably at least as important as differences in mean scores.

There are at least two reasons to expect that web-based interviewing may yield larger variances than face-to-face interviews. First, chapter 5 revealed that respondents are more likely to choose the center category of a scale in a face-to-face interview, potentially as a more socially desirable alternative to admitting to the interviewer that they don't know the answer. All else being equal, this implies that face-to-face interviews should produce somewhat smaller variances. Second, respondents with extreme views on both ends of a scale may moderate their views somewhat in a face-to-face interview to appear more socially desirable.

Table 4 displays differences in variances between the survey modes. The only difference with the analyses for the mean scores (see chapter 6) is that the dependent variable here was an absolute z-score, rather than a regular z-score. For example, a score of 1 on this variable indicates that the respondent scored either one standard deviation below or above the mean. The results reveal small to moderate differences in variances on most variables. As expected, the variances of most key variables were larger in the CAWI-mode than in the CAPI-mode. The differences are particularly apparent on the sympathy scores for political parties, which suggest that respondents are relatively hesitant to express either strong sympathy or strong antagonism towards a party to an interviewer. As was the case for the mean levels in the previous chapter, the differences in variances between the CAPI and the CAWI-mode could not be reduced by controlling for demographic characteristics and vote choice, or by controlling for scores on a social desirability scale. This indicates that survey weights will be ineffective in making variances comparable between both survey modes. The panel-mode contrarily yielded mostly similar variances to the CAWI-mode.

In conclusion, there were small to modest differences in variances between web-based interviewing and face-to-face interviewing. All else being equal, the larger variances of web-based interviewing can be considered an advantage of this survey mode. The more variation there is in scores, the more possibilities researchers have to compare (groups of) respondents and make substantive inferences. The variances of respondents that were recruited from an ongoing internet panel were mostly similar to those of respondent from a fresh probability sample.

8. Time trends

Key findings

- *Web-based interviewing mostly yielded substantively similar time trends from earlier rounds of the DPES compared to face-to-face interviewing, but clear discontinuities were visible for a limited number of groups on a limited number of variables.*

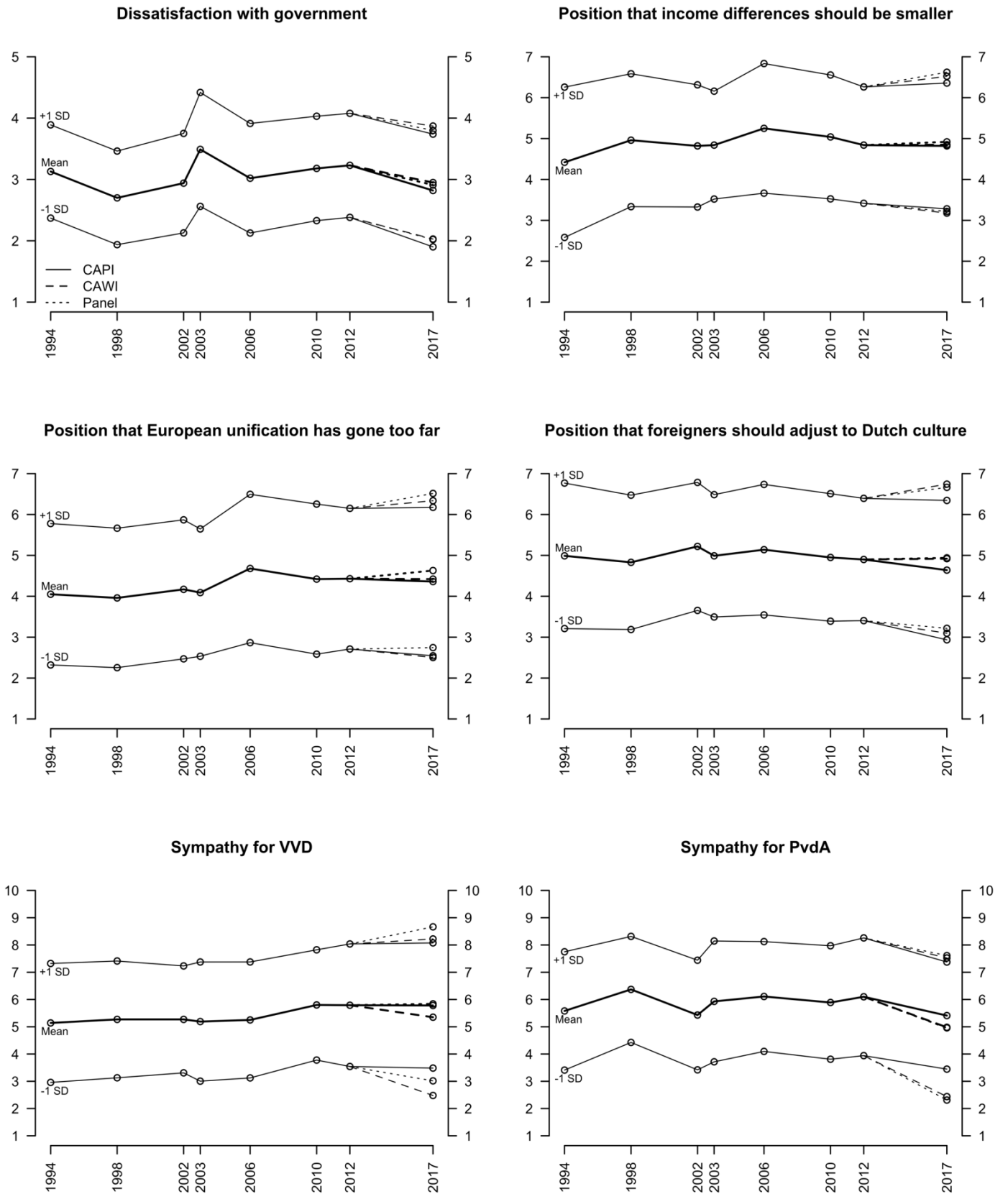
Dating back to 1971, the Dutch Parliamentary Election Study is the longest running political survey in the Netherlands. Arguably the most important risk of changing the survey mode is therefore that scores may become incomparable to previous rounds and that the DPES could consequently lose its unique ability to examine how public opinion has evolved over time. Discontinuities in average scores could be caused by the differential mean levels in each survey mode that were revealed in chapter 6. In addition, discontinuities in the time trends for specific groups of voters may also be introduced by the differential variances that were found in chapter 7. However, if and to what extent altering the survey mode leads to substantively different inferences about time trends depends on the relative magnitude of differences between survey modes compared to the strength of over-time changes.

Figure 2 depicts time trends since 1994 for six key variables. To show the impact of changing variances, this graph shows time trends not only for the mean of each variable, but also for scores of one standard deviation above or below this mean. All analyses were weighted for both demographic characteristics and vote choice. The results show that most time trends were substantively similar across the three survey modes. In those instances when the direction of the trends differs, this is due to substantively very small divergences (i.e., between a very small decrease or a modest increase) that are unlikely to affect the long term trend. We cannot, however, ascertain potential differences in middle- to long-term trends.

There are however some noticeable exceptions in which time trends differ meaningfully between the three components. This only happened in cases where differential means and differential variances affected scores in the same direction. Based on the CAPI-mode, we would for example conclude that sympathy for the PvdA was reasonably constant between 2012 and 2017 among voters who were unsympathetic towards this party. In other words, we would conclude that the amount of explicit dislike for the PvdA among Dutch voters was unchanged during this period. However, we would draw a substantively different conclusion when looking at scores from the CAWI-mode or panel-mode. In this case, we would contrarily conclude that there was indeed an increase in explicit dislike for the PvdA between 2012 and 2017.

To conclude, web-based interviewing mostly yielded substantively similar time trends in relation to earlier rounds of the DPES to face-to-face interviewing, but clear discontinuities were visible for a limited number of groups on a limited number of variables. The DPES can therefore still be used to examine changes in public opinion since 1971 after the introduction of a new survey mode, but researchers should be cautious and use the 2017 data to check if specific over-time trends were altered by the introduction of new survey modes.

Figure 2: Time trends for key variables.



9. Test-retest reliability

Key findings

- *Despite some methodological reservations, web-based interviewing appears to have yielded a better test-retest reliability than face-to-face interviewing.*

A core component of data quality is that respondents' scores are determined by their genuine orientations and characteristics, rather than by random variations. If respondents carefully consider their response, they will probably give the same answer when they are asked the same question again at a later moment. Contrarily, respondents will likely give a very different response the second time if they had randomly selected their answer on the first occasion. Strong associations between scores on the first and second occasion that a question was asked can therefore be taken as evidence for measurement reliability in a survey mode. This type of reliability is known as 'test-retest reliability'.

The test-retest reliability in the DPES 2017 could be assessed by comparing respondents' answers in the initial main questionnaire to their responses in the supplementary questionnaire that they completed at a later moment. Although the supplementary questionnaire did not ask questions that were literally identical to those in the main questionnaire, it featured many items that were highly similar (see the codebook for the exact wording of the items). Before turning to the results, it should be emphasized that this method to determine test-retest reliability has some shortcomings. First, differences between scores on the main questionnaire and the supplementary questionnaire may be due to the slightly different question working. Second, the supplementary questionnaire was administered only in the CAPI-mode and the panel-mode, but not in the CAWI-mode. Although differences in test-retest reliability between both modes seem more likely to be driven by the interview mode (i.e., web-based or face-to-face), they may alternatively be caused by the differential way in which respondents were recruited (i.e., fresh probability sample or ongoing internet panel with more 'professional' respondents). Third, the supplementary questionnaire was administered as a paper and pencil questionnaire (i.e., PAPI) in the CAPI-mode and as a web-survey in the panel-mode. As such, differences in test-retest reliability between both modes may reflect strengths and weaknesses of paper and pencil interviewing as well as face-to-face interviewing. Fourth, the average amount of time that passed between the completion of the main questionnaire and the supplementary questionnaire differed between both survey modes. In the panel-mode, respondents completed the main questionnaire in March and the supplementary questionnaire in July of 2017. In the CAPI-mode, the interviewer left a paper and pencil questionnaire that respondents could fill out at their own convenience. Unfortunately, it is impossible to determine exactly how much time passed for most CAPI-respondents because they were not asked to provide the date on which they filled out the supplementary questionnaire. However, it seems likely that the average period between both questionnaires was (much) longer for the panel-respondents.

Despite these limitations, the results in Table 5 give an impression of the test-retest reliability in both survey modes. Surprisingly, this reliability was substantially higher in the panel-mode than in the CAPI-mode. Although this finding should be interpreted with caution in the light of methodological limitations, this indicates at the very least that web-based interviewing did not produce more random or less deliberate answers than face-to-face interviewing. If anything, the opposite might have been the case.

Table 5: Test-retest reliability.

Core Variables	Criterion Variable	CAPI	Panel
V024: Interested in politics	S043: Interested in politics	0.61	0.71
V083: Satisfaction with govern.	S171: Good job govern.	0.69	0.69
V098: Income differences - p. resp.	S132: Income differences (R)	0.54	0.61
V108: European unification: - p. resp.	S053: Trust - European Union	0.44	0.54
V118: Foreigners - p. resp.	S156: Cultured harmed by immigr. (R)	0.42	0.57
V133: Left-right self-rating	NA	NA	NA
V255: External efficacy	S141: Vote makes difference	0.27	0.38
V258 and V259: Internal efficacy	NA	NA	NA
V260 until V263: Political cynicism	NA	NA	NA
Sympathy Scores			
V200: Sympathy score: VVD	S090: Probability vote for VVD	0.56	0.68
V201: Sympathy score: PvdA	S091: Probability vote for PvdA	0.38	0.63
V202: Sympathy score: PVV	S092: Probability vote for PVV	0.59	0.77
V203: Sympathy score: CDA	S094: Probability vote for CDA	0.48	0.62
V204: Sympathy score: SP	S093: Probability vote for SP	0.46	0.63
V205: Sympathy score: D66	S095: Probability vote for D66	0.53	0.64
V207: Sympathy score: GroenLinks	S097: Probability vote for GroenLinks	0.50	0.64
Average:		0.50	0.62

Note:

Pearson's R correlation coefficients.

(R): Reversed scored to facilitate interpretation.

Green: Higher than 0.50

Orange: Between 0.30 and 0.50

Red: Lower than 0.30

10. Criterion validity

Key findings

- *Web-based interviewing yielded a better criterion validity than face-to-face interviewing.*
- *Recruiting respondents from an ongoing internet panel resulted in an identical criterion validity compared to using a fresh probability sample.*

Another key indicator of a survey's data quality lies in the criterion validity of its measurements. A measurement is said to possess criterion validity if it can be used to predict key outcomes. These outcomes may either be in the future (i.e., predictive criterion validity) or in the present (i.e., concurrent criterion validity). Because the main purpose of any election study is to explain why voters vote the way they do, respondents' vote choice can be seen as a key criterion in this type of surveys. Associations between respondents' vote choice and their scores on key variables can therefore be taken as an indicator of concurrent criterion validity.

Table 6 displays the criterion validity in each survey mode. Because vote choice is a categorical construct, its correlation with the key variables was determined using regression analyses in which the key variables featured as the dependent variables and vote choice was specified as a categorical (i.e., dummy-recoded) independent variable. The correlations between the key variables and vote choice were then calculated as the square root of the explained variances of these regression analyses. Survey modes could subsequently be compared by calculating the differences between the correlations in each survey mode. The statistical significance of these differences was finally determined by using an F-test for the joint significance of all interactions between survey mode and vote choice in predicting the key variables.

Surprisingly, the results in Table 6 show that the CAWI-mode featured a better criterion validity than the CAPI-mode. This finding may partially be explained by the fact that web-based interviewing produced a larger variance for most variables (see chapter 7). If the variation in a variable is larger, it can be used more effectively to distinguish the voters of different parties. Hence, the variable will have a greater ability to predict outcomes like vote choice. Respondents' tendency to give more optimistic and socially desirable answers in face-to-face interviewing may also have played a role. For example, non-voters or PVV-voters may have more openly expressed their political cynicism and dissatisfaction with the government in the web-based interviews than in the face-to-face interviews. This may explain why web-based interviews were also more effective in distinguishing these voters from other respondents on these variables. Comparisons between the CAWI and the panel-mode did not reveal any significant differences. This indicates that recruiting respondents from an ongoing internet panel resulted in an identical criterion validity compared to using a fresh probability sample.

Table 6: Concurrent criterion validity.

Core Variables	CAPI	CAWI	Panel	CAPI - CAWI	Panel - CAWI
V024: Interested in politics	0.29	0.30	0.23	-0.01	-0.07
V083: Satisfaction with govern.	0.44	0.55	0.50	-0.12**	-0.06
V098: Income differences - p. resp.	0.42	0.55	0.50	-0.14*	-0.05
V108: European unification: - p. resp.	0.44	0.44	0.48	-0.01	+0.04
V118: Foreigners - p. resp.	0.43	0.44	0.47	-0.02	+0.03
V133: Left-right self-rating	0.62	0.62	0.65	+0.00	+0.03
V255: External efficacy	0.39	0.47	0.45	-0.08	-0.02
V258 and V259: Internal efficacy	0.31	0.27	0.21	0.04	-0.05
V260 until V263: Political cynicism	0.34	0.44	0.39	-0.10	-0.05
Sympathy Scores					
V200: Sympathy score: VVD	0.51	0.58	0.55	-0.07**	-0.04
V201: Sympathy score: PvdA	0.42	0.48	0.50	-0.06*	+0.01
V202: Sympathy score: PVV	0.61	0.60	0.64	+0.00	+0.04
V203: Sympathy score: CDA	0.49	0.52	0.51	-0.04***	-0.01
V204: Sympathy score: SP	0.42	0.48	0.49	-0.06	+0.01
V205: Sympathy score: D66	0.52	0.59	0.55	-0.07**	-0.04
V207: Sympathy score: GroenLinks	0.51	0.53	0.52	-0.02	-0.01
Average:	0.45	0.49	0.48	-0.04	-0.01

Note:

Multiple correlation voting vote choice (i.e., square root of the explained variance of vote choice). Statistical significance has been determined by using an F-test for the joint significance of all interactions between survey mode and vote choice in predicting the variables in the table. As such, the p-value does not always correspond with the magnitude of the standardized effect size.

(R): Reversed scored to facilitate interpretation. The averages were not tested for significance.

*: p < .05

** : p < .01

***: p < .001

Green: Difference of 0.05 or smaller

Orange: Difference larger than 0.05 and smaller than 0.10

Red: Difference larger than 0.10

11. Multiple regression

Key findings

- *In all but a few cases, web-based interviewing and face-to-face interviewing yielded identical estimates in multiple regression models.*
- *The multiple regression estimates from an ongoing internet panel were mostly similar to those from a fresh probability sample, but significant differences were observed in a sizable minority of cases.*

In scientific studies, election surveys are typically used to analyze how a number of explanatory variables predict individual differences in an outcome variable. This chapter therefore examines if the estimates from such models are affected by the choice of survey mode. Chapter 7 revealed that web-based interviewing yielded somewhat larger variances than face-to-face interviewing, which (*ceteris paribus*) could result in slightly larger parameter estimates in regression models.

Table 7 displays the standardized estimates from multiple regression models with respondents' sympathy scores for parties as the dependent variables and five common predictors of these scores as the independent variables. All models were estimated using survey weights (based on both demographic characteristics and vote choice) and controlled for core demographic characteristics. Interaction effects between the core predictors and survey mode were used to examine if identical estimates were obtained in all three survey modes.

When comparing the CAWI and the CAPI mode, only 3 out of the 35 interaction effects reached statistical significance. This is only slightly more than what may be expected based on random chance in the absence of any systematic differences ($35 * 0.05 = 1.75$). Moreover, the three significant differences all occurred in the same model (i.e., the one predicting sympathy for the VVD), which suggests that the differential estimates may be interrelated. In other words, the CAPI and the CAWI mode did not produce different estimates in all but a few cases and even these exceptions could still be attributable to random chance. These analyses therefore suggest that systematic differences between web-based and face-to-face interviewing in estimates from multiple regression models are either rare or absent.

The comparisons between the fresh sample and the ongoing internet panel however revealed more differences. Even though most estimates were similar between the CAWI and the panel mode, significant differences were observed in a sizable minority of 9 out of 35 cases. This is clearly more than what may be expected based on random chance. It therefore appears that the recruitment of respondents from an ongoing internet panel has at least some potential to alter the estimates of multiple regression analyses. One speculative explanation for this finding is that panel respondents have learned to better connect or distinguish certain questions by completing previous waves of the panel. Selective panel attrition could also play a role.

Table 7: Multiple regression models predicting sympathy for parties.

	CAWI	CAPI	Panel	CAPI - CAWI	Panel - CAWI
Sympathy for VVD					
Left-right rating: Party-respondent distance	-0.44 (0.06)***	-0.27 (0.05)***	-0.28 (0.04)***	+0.17 (0.07)*	+0.16 (0.07)*
Income differences: Party-respondent distance	-0.09 (0.06)	-0.29 (0.04)***	-0.31 (0.04)***	-0.20 (0.07)**	-0.22 (0.07)**
Foreigners: Party-respondent distance	-0.16 (0.05)**	-0.16 (0.04)***	-0.11 (0.03)**	+0.01 (0.07)	+0.05 (0.07)
European unification: Party-respondent distance	-0.04 (0.05)	-0.03 (0.04)	-0.12 (0.04)**	+0.01 (0.06)	-0.07 (0.07)
Political cynicism	-0.37 (0.07)***	-0.14 (0.06)*	-0.17 (0.03)***	+0.22 (0.09)*	+0.20 (0.07)**
Sympathy for PvdA					
Left-right rating: Party-respondent distance	-0.29 (0.06)***	-0.18 (0.05)***	-0.22 (0.04)***	+0.11 (0.07)	+0.07 (0.07)
Income differences: Party-respondent distance	-0.05 (0.06)	-0.10 (0.04)*	-0.22 (0.04)***	-0.05 (0.08)	-0.17 (0.08)*
Foreigners: Party-respondent distance	-0.07 (0.07)	-0.09 (0.04)*	-0.21 (0.04)***	-0.02 (0.08)	-0.13 (0.08)
European unification: Party-respondent distance	-0.12 (0.07)	-0.15 (0.04)***	-0.05 (0.05)	-0.03 (0.08)	+0.07 (0.08)
Political cynicism	-0.22 (0.07)**	-0.18 (0.06)**	-0.17 (0.04)***	+0.05 (0.08)	+0.04 (0.09)
Sympathy for PVV					
Left-right rating: Party-respondent distance	-0.17 (0.06)**	-0.25 (0.05)***	-0.23 (0.04)***	-0.09 (0.08)	-0.06 (0.08)
Income differences: Party-respondent distance	0.02 (0.05)	-0.04 (0.04)	-0.06 (0.04)	-0.06 (0.07)	-0.08 (0.07)
Foreigners: Party-respondent distance	-0.20 (0.06)**	-0.23 (0.04)***	-0.29 (0.05)***	-0.04 (0.07)	-0.09 (0.08)
European unification: Party-respondent distance	-0.26 (0.06)***	-0.19 (0.05)***	0.24 (0.05)***	+0.07 (0.08)	+0.02 (0.08)
Political cynicism	0.10 (0.07)	0.10 (0.05)	0.16 (0.05)***	+0.00 (0.09)	+0.06 (0.08)
Sympathy for CDA					
Left-right rating: Party-respondent distance	-0.35 (0.06)***	-0.22 (0.06)***	-0.20 (0.05)***	+0.12 (0.08)	+0.15 (0.07)*
Income differences: Party-respondent distance	-0.10 (0.08)	-0.15 (0.04)***	-0.18 (0.05)***	-0.05 (0.09)	-0.08 (0.09)
Foreigners: Party-respondent distance	-0.08 (0.06)	-0.04 (0.04)	-0.12 (0.04)**	+0.03 (0.07)	-0.04 (0.07)
European unification: Party-respondent distance	0.02 (0.06)	0.05 (0.04)	-0.17 (0.04)***	+0.03 (0.07)	-0.19 (0.07)**
Political cynicism	-0.06 (0.08)	-0.07 (0.05)	-0.06 (0.04)	-0.00 (0.09)	-0.01 (0.10)
Sympathy for SP					
Left-right rating: Party-respondent distance	-0.39 (0.06)***	-0.26 (0.05)***	-0.29 (0.05)***	+0.13 (0.08)	+0.10 (0.08)
Income differences: Party-respondent distance	-0.13 (0.08)	-0.15 (0.04)***	-0.26 (0.05)***	-0.02 (0.09)	-0.13 (0.09)
Foreigners: Party-respondent distance	0.02 (0.06)	-0.03 (0.05)	-0.03 (0.04)	-0.05 (0.08)	-0.06 (0.07)
European unification: Party-respondent distance	-0.15 (0.06)*	-0.08 (0.04)	-0.15 (0.04)***	+0.07 (0.07)	-0.00 (0.07)
Political cynicism	0.07 (0.08)	-0.06 (0.06)	0.13 (0.04)**	-0.12 (0.10)	+0.06 (0.09)
Sympathy for D66					
Left-right rating: Party-respondent distance	-0.30 (0.07)***	-0.22 (0.04)***	-0.11 (0.05)*	+0.08 (0.08)	+0.19 (0.08)*
Income differences: Party-respondent distance	0.04 (0.07)	-0.05 (0.03)	-0.19 (0.05)***	-0.09 (0.08)	-0.23 (0.08)**
Foreigners: Party-respondent distance	-0.15 (0.06)*	-0.17 (0.04)***	-0.20 (0.05)***	-0.01 (0.07)	-0.05 (0.08)
European unification: Party-respondent distance	-0.11 (0.07)	-0.15 (0.04)***	-0.25 (0.05)***	-0.03 (0.08)	-0.14 (0.08)
Political cynicism	-0.21 (0.07)**	-0.07 (0.06)	-0.11 (0.04)**	+0.14 (0.09)	+0.10 (0.08)
Sympathy for GroenLinks					
Left-right rating: Party-respondent distance	-0.41 (0.06)***	-0.35 (0.04)***	-0.33 (0.05)***	+0.06 (0.07)	+0.08 (0.08)
Income differences: Party-respondent distance	-0.03 (0.08)	-0.06 (0.04)	-0.29 (0.05)***	-0.04 (0.08)	-0.26 (0.09)**
Foreigners: Party-respondent distance	-0.11 (0.07)	-0.18 (0.04)***	-0.19 (0.04)***	-0.06 (0.08)	-0.07 (0.08)
European unification: Party-respondent distance	-0.19 (0.07)**	-0.08 (0.04)*	-0.06 (0.05)	+0.10 (0.07)	+0.12 (0.08)
Political cynicism	-0.11 (0.07)	0.01 (0.05)	-0.03 (0.04)	+0.12 (0.09)	+0.08 (0.08)
Average absolute value:	0.16	0.14	0.18	0.07	0.10

Note:

The columns 'CAWI', 'CAPI' and 'Panel' display standardized results from a set of multiple regression models with respondents' sympathy scores for parties as the dependent variables. The independent variables are the listed variables, as well as the following variables that are omitted from the table: age (linear), gender, educational level (dummy recoded), social class self-image (dummy recoded), and religious denomination (dummy recoded). The columns 'CAPI - CAWI' and 'Panel - CAWI' display interaction effects from a second set of regression models that additionally included interaction terms between all independent variables and survey mode.

*: p < .05

** : p < .01

***: p < .001

Green: No significant difference

Orange: Significant difference p < .05

Red: Significant difference p < .01

12. Conclusions and recommendations

Key recommendations

- *Face-to-face interviewing can be replaced entirely by self-completion in future DPES-rounds.*
- *Almost all respondents can be invited to complete their survey online. Additional measures should be implemented to raise the response rate among older voters (e.g., by giving this specific group the opportunity to fill out a paper-and-pencil questionnaire).*
- *Time trends can still be examined after a switch to self-completion, but researchers are advised to use the 2017 data to check and correct for potential discontinuities.*
- *A substantial proportion of the respondents should be recruited from an ongoing internet panel (LISS). This has the advantage that it (re)introduces a dynamic element in the design of the DPES.*
- *While the substantive differences between fresh samples and the LISS-panel were small, the fresh sample was more representative of the Dutch population. As long as this remains to be the case, it is advisable to also recruit a sizable number of respondents from a fresh probability sample so that the DPES can maintain its status as a benchmark for representativeness.*

Ever since 1971, the Dutch Parliamentary Election Study has been conducted using face-to-face interviews and a fresh probability sample. This report examined if and to what extent the representativeness and data quality of the DPES would be affected by a switch to web-based interviewing or by recruiting respondents from an ongoing internet panel. To this end, the three survey modes were compared on key indicators. For each indicator, it was determined if the results differed between the survey modes and, if so, which survey mode yielded the best data quality.

Can face-to-face interviewing be entirely replaced by web-based interviewing in the DPES?

Face-to-face interviews are considerably more expensive than web-based interviewing. Moreover, web-based surveys offer additional advantages such as greater possibilities to conduct survey experiments. Many face-to-face surveys have therefore switched to web-based interviewing in recent years. The analyses in this report clearly indicated that this is also a feasible option for the Dutch Parliamentary Election Study. Compared to face-to-face interviewing, web-based interviewing yielded a slightly better overall representativeness, a better variability in scores, a better test-retest reliability, and a better criterion validity. Both survey modes were about tied with regard to item non-response: Whereas web-based interviewing yielded more 'don't know' answers, face-to-face interviewing produced more answers in the center category of the scale. Moreover, web-based interviewing produced highly similar estimates in multiple regression models.

The only benefits of face-to-face interviewing were therefore that this survey mode featured a somewhat better response rate and a better reach of older voters. However, it seems quite possible to overcome these limitations of web-based interviewing in future DPES-rounds. For example, voters over age 75 could be sent a paper-and-pencil questionnaire along with the invitation letter for the online survey. By giving older voters the choice to either complete the questionnaire online or on paper, the representation of this group in the DPES can be ensured. In addition, this measure may also

raise the overall response rate of the survey. Moreover, additional measures can be considered to raise the response rate of future web-based DPES-rounds, such as raising the monetary incentives for respondents or the intensity of contact attempts.

Because the DPES has a unique ability to examine how public opinion has evolved since 1971, arguably the biggest drawback of changing its survey mode is that doing so may create discontinuities in time trends. The analyses in this report indicated that this problem is manageable, but not to be neglected. For all reasonable purposes, short-term time trends were substantively unaltered for most groups of voters on most variables. However, some clear discontinuities were visible for some groups (e.g., voters with exceptionally high or low scores) on a limited number of variables. As such, time trends can still be examined after a switch to web-based interviewing, but researchers are advised to use the 2017 data to check and correct for potential discontinuities that may otherwise alter substantive inferences in some instances. Provided sufficient funding is available, these corrections could be further improved by including face-to-face interviews in the DPES round of 2021 for a fairly limited subsample, to ascertain the validity of the estimated time trends in the longer run.

Can the recruitment of respondents from a fresh probability sample be supplemented or replaced by the recruitment from an ongoing internet panel?

Fresh probability samples are the gold standard in survey research when it comes to achieving a representative sample. Even though the LISS-panels was initially recruited through probability sampling of households, its representativeness was lower than the one that was recruited from a fresh probability sample of individuals. For a study like the DPES, which aims to be a benchmark for representativeness, this can be considered a crucial drawback. So, as long as the LISS-panel remains to be less representative than fresh samples in terms of background characteristics and voting behavior, it is advisable to continue recruiting a sizable number of respondents from a fresh probability sample in future DPES-rounds.

The results in this report therefore suggest that the most promising way forward for the DPES is to move to a mixed-mode design, which has self-completion as the mode, and which combines respondents who were recruited from a fresh probability sample with others who were recruited from an ongoing internet panel. Respondents from the LISS-panel obtained only slightly different mean scores on key variables than other respondents. In some cases, the LISS-panel also produced somewhat different estimates in multiple regression models. On all other indicators (e.g., variances, item non-respondents, and criterion validity), the panel sample however produced (almost) identical results. This means that it is possible to add panel respondents to a fresh probability sample without causing major comparability problems. For most purposes, panel respondents and respondents from a fresh probability sample can reasonably be analyzed together in a single dataset. As such, the DPES can achieve the best of both worlds by maintaining its status as a benchmark for representativeness (i.e., by recruiting a sizable number of respondents from a fresh probability sample), while also opening new possibilities to follow individual voters over time (i.e., by including respondents from an ongoing internet panel).

References

Dahlgard, J., Hansen, J., Hansen, K., & Bhatti, Y. (2019). Bias in self-reported voting and how it distorts turnout models: Disentangling nonresponse bias and overreporting among Danish voters. *Political Analysis*, 1–9.

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