Mode matters

Haan, Marieke

Chapter 4

A Humanized Web Survey and Mode Effects*

* This Chapter was adapted from: Haan, M., Ongena, Y.P., Vannieuwenhuyze, J., & Glopper, C.M., de. (submitted). A Humanized Web Survey and Mode Effects.

4.1 Introduction

Web surveys are a popular method for data collection in survey research. Survey organizations are increasingly using web surveys due to time and cost efficiencies (Callegaro et al., 2014), the ever-improving technological possibilities, and the increase in Internet use in Western countries. In addition, in self-administered questionnaires (SAQ) such as web, respondents are inclined to answer especially sensitive questions more truthfully than in interviewer-administered modes (e.g., Tourangeau & Smith, 1996). Yet, the risks of survey satisficing are higher in SAQ than in interviewer-administered modes (Heerwegh & Loosveldt, 2008). In web surveys, there is no interviewer present who has control over the interview situation and respondents can easily be distracted. Thus, respondents can essentially act at their own discretion without having to justify their answers to an interviewer.

To benefit from the advantages of web surveys and to reduce their weaknesses, researchers have integrated human-like features in SAQ in order to simulate interviewer presence (visibility and audibility). One way of humanizing web surveys is the use of a voice in audio computer self-interviewing (ACASI). Since the 1990s researchers have been using ACASI modes in which respondents listen to a pre-recorded voice reading the questions but self-administer their answer using a digital device (O'Reilly, Hubbard, Lessler, Biemer, & Turner, 1994). Another way to humanize web surveys is the inclusion of pictures of researchers and interviewers in web questionnaires (Krysan & Couper, 2006; Tourangeau, Couper, & Steiger, 2003). A third way of humanizing web surveys is to incorporate both auditory and visual human aspects in a web survey, by including videos of interviewers. However, studies on the effects of videos in web surveys on response behavior are limited and often executed in a laboratory setting.

Two aspects are remarkable in the existing literature on integration of human-like features in SAQ. The first is the single focus on socially desirable responding. ACASI modes have been found to provide data quality comparable to web surveys with regard to reporting of sensitive behavior (Couper, Singer, & Tourangeau, 2003). Put differently, respondents report more socially undesirable answers in ACASI compared to face-to-face surveys (Tourangeau & Smith, 1996; Tourangeau & Yan, 2007). Pictures do not seem to evoke socially desirable responding either (Krysan & Couper, 2006; Tourangeau et al., 2003). Secondly, human-like features in SAQ are primarily used to reduce satisficing behavior compared to traditional Text-web surveys but, remarkably, none of the above mentioned studies have focused on satisficing behavior. Due to humanized features, web surveys with videos of interviewers might lead to a decrease in satisficing behavior compared to Text-web surveys.

Another striking aspect in the existing literature is that the prior studies are either done on a limited set of items (Fuchs & Funke, 2007), or they are based on analysis

at the respondent level (Krysan & Couper 2003; Fuchs, 2009; Gerich, 2008). Thus, in prior studies, the correlation between responses within respondents and within items is not taken into account. In order to adequately study the effects throughout an entire questionnaire, analyses should be done at the response level, using mixed-effect models (Yan & Tourangeau, 2008).

Triggered by the possibilities of improving data quality of web surveys by incorporating human-like features, we developed a Video-web mode for our study. In this survey mode respondents can see pre-recorded clips of an interviewer reading the questions, and then self-administer their answer. Firstly, we investigate the possibility that this Video-web mode can decrease the amount of satisficing behavior that is found in traditional Text-web. Since we expect Video-web to reduce satisficing behavior, we expect satisficing behavior in Video-web to be at a lower level than traditional Text-web, and at a level more similar to computer assisted telephone interviewing (CATI), and computer assisted personal interviewing (CAPI). However, there are also risks involved when incorporating videos in web as these human aspects may also increase socially desirable answering behavior (Tourangeau et al., 2003). For example, Sproull, Subramani, Kiesler, Walker, and Waters (1996) found that individuals present themselves more positively when responding to human-like displays than text displays on computer screens. Therefore, web surveys should implement motivational human-like aspects that are subtle in such a way that they are easy to standardize and do not reach the degree of social presence of interviewer-administered modes. Since Video-web can increase the level of socially desirable responses we, secondly, investigate the effect of Video-web on socially desirable answering behavior. By only implementing subtle human cues we expect the degree of social presence not to evoke as much socially desirable responding as in CAPI and especially CATI.

The Chapter proceeds as follows. In section 4.2, we start by providing what is known about videos in web surveys and which variables are important to study response behavior. In section 4.3 we present our experimental design and describe our statistical methods. Then, our results will follow in section 4.4. Finally, section 4.5 contains a discussion and possibilities for future work.

4.2 Background

Research on the use of videos in SAQ is limited thus far and there is a lot of variation within the few existing studies (see Table 4.1). Most existing studies implement video-recorded interviewers, but some use virtual interviewers (Lind, Schober, Conrad, & Reichert, 2013; von der Pütten, Hoffmann, Klatt, & Kramer, 2011). Krysan and Couper (2003) were the first to implement video-recordings of interviewers in a SAQ in a labo-

ratory setting. Their findings show that interviewer's race effects operate similarly for video-recorded and live interviewers; hence interviewer effects are shown for Video-CASI. Fuchs and Funke (2007) were the first to compare Video-web and Text-web, but found no differences with respect to social desirability bias. Results opposite to the expected direction were found with respect to social presence differences: respondents in the Text-web condition had a higher score than respondents in the Video-web condition. However, social presence was measured by means of only three items that were also rather suggestive. The only study documenting an experiment that compares Video-CASI with a real recorded interviewer with face-to-face and ACASI reports mixed results (Gerich, 2008). The comparison modes of all studies reported in Table 4.1 vary greatly, different contact modes are used across the studies, and the samples involve students or local populations. Except for Fuchs and Funke's (2007) and Fuchs (2009) studies, who used student-samples in a field setting, all prior studies were conducted in laboratory settings. Due to these differences and the limited availability of Video-web studies in a field setting it is hard to draw firm conclusions about the advantages and weaknesses of this mode.

Although there are no studies published in which satisficing behavior is investigated directly in Video-web surveys, the literature provides some indirect support of less satisficing behavior in Video-web compared to Text-web. Due to the added human cues, Video-web surveys are partially comparable to CAPI modes which are known for their low risk of satisficing (Heerwegh & Loosveldt, 2008). Also, as more senses are used in Video-web (e.g., seeing a video-interviewer, hearing the question, and seeing the response options) the risk of being distracted by environmental stimuli while participating in the survey might be lower than in Text-web (Ansolabehere & Schaffner, in Messer 2013). Because respondents are more involved and focused in Video-web, this can lead to less satisficing in Video-web than in Text-web. In addition, the responses of Video-web might be better comparable to interviewer-administered modes, which could reduce effects between survey modes.

Regarding social desirability, it is known that socially desirable answering behavior positively correlates with the degree of social presence in surveys (especially for sensitive questions). The intensity of social presence is the degree to which the interviewer and the respondent are in the same physical and social space using a communication medium (Malakhoff & Jans, 2011; Short, Williams, & Christie, 1976). The highest degree of social presence can be found in face-to-face modes (e.g., CAPI) where the interviewer and the respondent are co-present; they share the same physical environment where they can see and hear each other (Clark, 1996). Nonetheless, it is likely that the relationship between interviewer presence and socially desirable responding is not a linear one, and other factors play a role. For example, Holbrook, Green, and Krosnick (2003) compared CAPI to CATI and found more socially desirable responding in CATI, although

Table 4.1 Videos in web surveys used in methodological survey studies (presented in chronological order)

Study	Sample members	Sample size & Response rate*	Contact mode	Response mode	Human-like feature	Analysis & Main results
Krysan & Couper, 2003	Lab experiment: White and African American subjects from Ann Arbor/ Ypsilanti US	N=160 Rate: -	f2f, newspaper advertisement, posters, church and community group visits	CAPI, Video-CASI** (random allocation)	Videos of real interviewers	Anova at respondent level. Interviewer's race operates similarly for virtual (video-recorded) and live interviewers; hence interviewer effects are shown for Video-CASI.
Fuchs & Funke, 2007	Field experiment: University of Kassel students Germany	N=398 Rate Video-web: 37% Rate Text-web: 50%	email invitation	Text-web, Video-web (random allocation / Text-web for second part survey)	Videos of real interviewers	Chi-Square tests at item level. Unit nonresponse rate and break-off rate higher in Video-Web than Text-web. No difference in social desirability bias. Scores social presence higher for Text-web than Video-web, but only three items were used to measure this.
Gerich, 2008	Lab experiment: University of Linz students Austria	N=200 Rate: -	f2f in public area of the main building of the university	PAPI, PASI, ACASI, Video-CASI (random allocation / PAPI for second part survey)	Videos of real interviewer	Anova and regression at respondent level. Higher reported norm deviation in Video-CASI, effect was mainly due to a low score in ACASI.
Fuchs, 2009	Field experiment: Online access panel University of Kassel students Germany	N=880 Rate: 37.5%	e-mail invitation	Text-web, Video-web (random allocation)	Videos of real interviewers	Chi-Square tests at item level. Mixed effects of interviewer gender on socially desirable behavior.

Table 4.1 Videos in web surveys used in methodological survey studies (presented in chronological order) (continued)

Study	Sample members	Sample size & Response rate*	Contact mode	Response mode	Human-like feature	Analysis & Main results
von der Pütten et al., 2011	Lab experiment: University of Duisburg-Essen students Germany	N=81 Rate: -	general advertising on campus	Video-CASI (PAPI for demographic items)	Virtual interviewer 'Thomas'	Anova at respondent level. Effect of specific virtual interviewer manipulations of verbal patterns on disclosing personal information.
Lind et al., 2013	Lab experiment: Subjects from New York City area US	N=235 Rate: -	online advertising on Graigslist, print, and online ads in Village Voice	CAPI, ACASI, Video-CASI (random allocation / PAPI for debriefing questions)	Virtual interviewers	Anova and logistic regressions at respondent level for individual items. Responses to virtual interviewers fall between ACASI and CAPI.

* N and Rate are based on the number of subjects who completed the questionnaire.

** In laboratory setting: Video-CASI, in field setting: Video-web.

social presence is higher in CAPI. The authors suggest that there might be trust issues when a respondent cannot see the interviewer and also that the lack of interviewer's rapport in CATI can cause this behavior. In addition, the respondent might not feel as comfortable during CATI because of the fast pace of the interview.

Web is known for its lowest social presence; therefore, in this mode, respondents are inclined to answer (especially) sensitive questions more truthfully than in interviewer-administered modes (Heerwegh, 2009). Based on the assumption that the degree of social presence correlates with socially desirable answering behavior it is likely that respondents provide more socially desirable responses in Video-web than in Text-web surveys. However, due to the fact that social presence in Video-web is not at the same level as social presence in interviewer-administered modes, it can be expected that there is less socially desirable behavior in a Video-web mode than in interviewer-administered modes. Furthermore, in a mixed-mode design with interviewer- and self-administered modes, the mode effects could be reduced when using Video-web instead of Text-web. Since mode inherent effects are smaller between Video-web and interviewer-administered modes (e.g., the degree of social presence), the provided answers are more comparable than when comparing Text-web to interviewer-administered modes.

Besides the response mode, it is important to take other factors into account that could affect answering behavior. One variable in mixed-mode studies that could possibly affect response behavior is whether respondents can directly choose their response mode or not. Allowing respondents to choose their response mode can create goodwill (de Leeuw, 2005) and can reduce respondents' satisficing behavior. Conrad et al. (2013) found less satisficing when sample members could choose a response mode compared to when respondents were allocated to a mode. They found fewer non-differentiation behavior and less rounding of numerical answers for the group of respondents that was offered a mode choice. However, it should be noted that the rounding of numerical answers, also known as response heaping, is not always a clear indicator for satisficing behavior (Holbrook et al., 2014). Furthermore, Smyth, Olson and Kasabian (2014) found fewer satisficing for multiple answer items and open-ended items when respondents answered these questions in a preferred mode than in a non-preferred mode. The possibility to choose a mode could also decrease social desirability effects. Respondents who can select their preferred mode might be more willing to give honest answers to sensitive questions, since they were able to choose the mode that feels most comfortable for responding. To our knowledge, there are no studies investigating this matter.

Furthermore, respondents' social-demographic characteristics can influence response behavior as well (e.g., gender, age, and education). In some (single-mode) studies, satisficing behavior is stronger with lower educated respondents; they are more likely to select don't know options (Holbrook et al., 2003; Krosnick et al., 2002). In addition, for some groups, such as the elderly, lower educated respondents, or females,

it is assumed that they are more eager to present themselves in a favorable way as a sort of defense mechanism for their vulnerable position in society (Green, Krosnick, & Holbrook, 2001). This can lead to potential socially desirable answering behavior. In addition, certain response behaviors can be activated by the particular topic of the question (Vannieuwenhuyze & Loosveldt, 2013). For example, it is likely that respondents provide more don't know responses for opinion questions than for social-demographic questions.

4.3 Methods

4.3.1 Data Collection

In order to compare response behavior in Video-web with Text-web, CAPI, and CATI we implemented an experiment in an additional round of the Dutch European Social Survey (ESS). Data were collected from March to June 2012 by GfK Panel Services Benelux. For the selection of our sample, data of the ESS 2010 round was used to study in which municipalities sample members lived who fulfilled criteria for being a hard-to-survey sample member (see Chapter 3). When a respondent of the ESS round 2010 fulfilled the criteria for being a hard-to-survey sample member the municipality in which the individual resided was selected. These selected communities made up two-third of the initial sample of communities. The other one third of the initial sample was a mere random selection of municipalities, not over-sampling communities with hard-to-survey groups. This initial selection among the 441 Dutch municipalities left us with 283 municipalities. Subsequently, 169 out of these 283 municipalities were selected based on the location of GfK's employed interviewers.

Based on the remaining municipalities a multistage sample design was used. In the first stage, a stratified sample of 40 municipalities out of the 169 was drawn taking into account the location (equal selection within the 12 Dutch provinces) and the urbanization levels of the municipalities. In the second stage, address-based sampling was applied by means of databases with information on population characteristics for the ZIP code areas within the 40 municipalities (see also Chapter 3 of this thesis for the procedure). The total sample at the end of stage two consisted of 3,496 households. In the third stage, an adapted form of the last-birthday method was used to select one individual within each selected household (when more than one individual lived in the household). This standard sampling method of the ESS requires that the interviewer asks which person in the household had his or her birthday closest to a randomly chosen date. The identified individual should then be selected for the survey (no substitute can be taken). An experimental concurrent mixed-mode design combined with an experimental non-concurrent mixed-mode design was used to collect the data (see

Table 4.2). First, all the sample members received a letter at their home in which the goal of the survey was introduced and their selection for this study was explained. The sampled households were randomly allocated to five experimental groups. One group was contacted face-to-face and could choose between CAPI and web. A second group was contacted by telephone and could choose between telephone and web. Sample units in the other three groups were randomly allocated to CAPI (group 3), CATI (group 4), or web (group 5) after being contacted by telephone. In groups 1, 2, and 5, half of the web respondents started with Text-web and then switched to Video-web, the other half started with Video-web and then switched to Text-web halfway through the questionnaire.

Nonetheless, these web respondents were not aware of the existence of two web modes prior to the switch, and which mode they received first was determined by a random procedure. Interviewers continued visiting and calling sample members until a sufficient number was reached for all response modes. Of the 3,496 households selected for this study, 824 participated in the survey. Another 327 households were willing to participate but were not contacted again for the survey because of budget reasons or expiration of the data-collection period. Combining all groups, the response rate was 37.5%, and the cooperation rate 42.1% (AAPOR 2011, RR1 and COOP1). The highest

Table 4.2 Outcome rates experimental groups

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
Contact Mode	Face-to-face	Telephone	Telephone	Telephone	Telephone
Choice	CAPI or WEB	CATI or WEB	Random allocation to CAPI	Random allocation to CATI	Random allocation to WEB
CAPI	117 (40.6%)		100 (100%)		
CATI		100 (44.5%)		106 (100%)	
Text-web then Video-web	96 (33.3%)	59 (26.2%)			47 (44.8%)
Video-web then Text-web	75 (26.1%)	66 (29.3%)			58 (55.2%)
Subtotal	N=288	N=225	N=100	N=106	N=105
Total N=824					
Response Rates - RR1	54.9%	34.8%	20.8%	31.3%	36.7%
Cooperation Rates - COOP1	60.6%	40.6%	23.6%	32.8%	41.4%

response and cooperation rates were found for group 1 (54.9 % and 60.6%), followed by group 5 (36.7% and 41.4%), and then group 2 (34.8% and 40.6%). The lowest rates were found for group 3 (20.8% and 23.6%) and group 4 (31.3% and 32.8%)

4.3.2 Design Video-Web Mode

The Video-web mode consisted of pre-recorded clips of an interviewer reading the questions to the respondent (see Figure 4.1). In our study, one white female interviewer was taped for the entire interview, as most of the interviewers in the CAPI and CATI mode were white females as well. The questions were not repeated in writing on the screen, but the respondent could watch every clip as often as desired. When the questions were read by the interviewer, the response options were already presented on the screen. To make the Video-web mode more comparable with the interviewer-administered modes (CAPI and CATI), for some questions the video-interviewer read the response options out loud as well. When more explanation for a question was necessary (e.g., 'Only one answer possible, select the most important.'), this was displayed on the screen. After

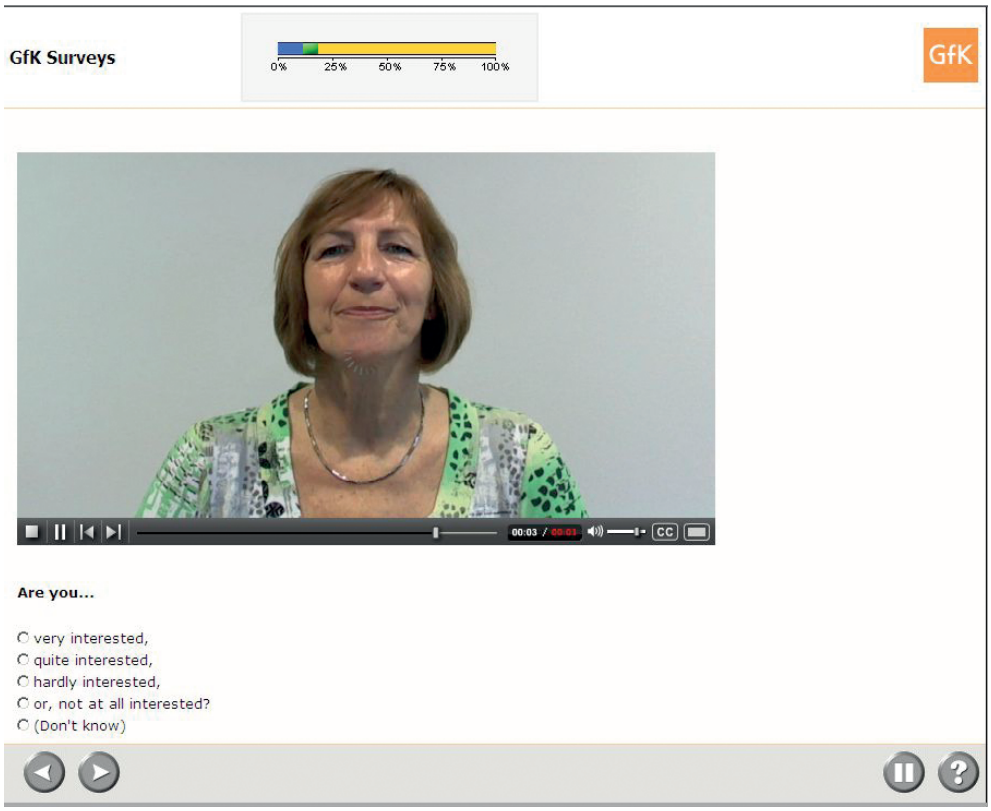

Figure 4.1 Example Video-web display

hearing the question and reading its response options, the respondent could answer the question by self-administration. Although the respondent was able to see the interviewer in the video, no interaction was possible between the interviewer and the respondent and no interactive feedback was provided by the video-interviewer.

4.3.3 Dependent Variables

We compared response behavior in Video-web with Text-web, CAPI and CATI in terms of satisficing and socially desirable responding. In our study, satisficing includes the number of don't know responses and non-differentiation answering behavior. We used a slightly adjusted version of the ESS round 5 core questionnaire sections on politics (41 items), subjective wellbeing (36 items), and social demographics (53 items) (European Social Survey, 2015). The items on politics appeared before the web mode switch and the items on subjective wellbeing and social demographics appeared after the web mode switch.

Satisficing Behavior: Non-Differentiation

Our questionnaire contained five batteries of items that were used to measure non-differentiation (see Appendix A.2). Four batteries were presented with questions on politics (18 items in total), and one battery was presented related to subjective wellbeing topics (4 items in total). To measure non-differentiation, we calculated the mean of the root of the absolute differences in the answers between all pairs of items for each response given within the battery. The answering scale was the same for all items within one battery, but differed between batteries, and therefore the non-differentiation score was computed separately for each battery. We used a formula of Mulligan et al. (2001; see also Chang & Krosnick, 2009), the next formula is an example for a battery with four items:

$$X = \frac{(\sqrt{|Item1-Item2|} + \sqrt{|Item1-Item3|} + \sqrt{|Item1-Item4|} + \sqrt{|Item2-Item3|} + \sqrt{|Item2-Item4|} + \sqrt{|Item3-Item4|}}{6}$$

To yield an index where higher scores indicated more non-differentiation, the highest level of observed differentiation was subtracted from each score and then divided by the negative value of the highest observed differentiation level (Chang & Krosnick, 2009):

$$X = \frac{(\text{score} - \text{highest level of observed differentiation})}{-\text{highest level of observed differentiation}}$$

So, non-differentiation can range from 0 to 1: 0 meaning the least non-differentiation possible (thus maximum variation), and 1 meaning the most non-differentiation pos-

sible (thus no variation). Finally, the non-differentiation score of each battery within a respondent was used as a continuous dependent variable in the analysis models.

Satisficing Behavior: Don't Know Responding
Within our version of the ESS questionnaire, 117 items offered a 'don't know' option (see Appendix A.1). The questions on politics and subjective wellbeing all contained a don't know option and 40 out of the 53 questions on social demographics had a don't know option. For each of the 824 respondents we removed the items that were not applicable. To measure don't know answering behavior, we created a binary indicator that took value 1 if a don't know response was given and 0 if any other answer was provided.

Socially Desirable Responding
To find out whether socially desirable responding varies across modes, 21 items were selected that have been shown to be sensitive to socially desirable answering behavior or have widely shared social desirability connotations (see Appendix A.3), such as income, religious service attendance, voting, and interest in politics (Jäckle, Roberts, & Lynn, 2006; 2010). Other topics were included about which we thought respondents might be hesitant to disclose their true opinions (such as immigration, and equality between the sexes). For most items, the responses on one side of the scale (both the extreme and moderate options) were considered to be the potentially socially desirable answers similar to the study of Jäckle et al. (2006). For some of the items the middle response options were more likely to be the socially desirable answer (e.g., religiosity, church attendance, income). At the item level we created a binary indicator that took value 1 if the response was considered as a potential socially desirable answer and 0 if any other answer was provided.

4.3.4 Statistical Approach

In this study the analysis units are the responses on all different items for all respondents. For that reason, we fitted mixed-effects models in order to take into account the correlation between responses within respondents and within the items (Yan & Tourangeau, 2008). More specifically, we fitted two-level models with crossed random intercepts for the respondents and items. However, for the non-differentiation model, item was not included as a random factor because here the analysis units are item batteries instead of the separate items. Since we only had five batteries at our disposal we decided not to include battery as a random factor.

Our primary interest lies in the effect of mode on the response outcomes. For that reason, mode is included as the first and principal fixed-effect. Additionally, we also included interactions between mode on the one hand and question topic (politics, subjective wellbeing, or social demographics) and item respectively on the other hand.

The interaction between question topic is included because we expect mode effects to differ between the topics. For example, we expect larger mode effects on the occurrence of socially desirable answers for questions about political opinions or subjective wellbeing than for questions about social demographics. Additionally, we also expect that mode effects may differ from item to item within each topic as well. For that reason, the interaction between item and mode was also included in the model through random regression coefficients.

Substantial effects of mode on the outcome can, however, not straightforwardly be interpreted as causal effects because they may also be caused by differences between the respondents of the different modes. A common technique to avoid this problem is the inclusion of additional covariates to the regression equation which explain the selection of different respondents into the different modes. In our analysis, we included contact mode, gender, age, and education as covariates in order to neutralize imbalanced selection of respondents between modes. As shown by previous studies, these covariates are relevant regarding response mode preferences (Loges & Jung, 2001; Schneider et al., 2005). It should be remarked that this set of covariates is still somewhat limited. As a result, causal conclusions from our analysis about mode effects should be handled with care.

As we are dealing with dichotomous dependent variables for socially desirable responses and don't know responses, generalized linear mixed models were fit using the glmer function in package lme4 (Bates, 2005), which is an extension package to R. For the continuous dependent variable non-differentiation we used the lmer function of the same package.

4.4 Results

4.4.1 Results Satisficing Behavior

Non-differentiation

Regarding non-differentiation, our results yielded very small effects for mode and none of these effects were significant (see Table 4.3). These results mean that Video-Web cannot improve data quality compared to Text-web, CATI and CAPI on the level of non-differentiation, simply because there are no clear differences with Text-Web, CATI, and CAPI from the outset. On the positive side, Video-Web does not seem to deteriorate data quality either compared to the other three modes.

Furthermore, no significant differences were found for the five batteries that were used to measure non-differentiation. Also, no significant differences were found for the control variables. The group to which the respondents were randomly allocated did not

affect non-differentiation behavior, and the effects for gender, age, and education are very small.

Table 4.3. Non-differentiation in Video-web

Fixed effects	**Estimate**	**SE**
Intercept	0.5167	0.0135
Mode(capi)	0.0015	0.0099
Mode(cati)	0.0000	0.0100
Mode (t-web)	0.0026	0.0093
Group (contact f2f choice)	0.0060	0.0082
Group (contact tel choice)	0.0172	0.0086
Gender (fem)	0.0044	0.0065
Age (middle)	0.0008	0.0093
Age (old)	0.0034	0.0093
Educ (high)	-0.0002	0.0073
Educ (unknown)	-0.0354	0.0197
Batteryset 2	-0.0481	0.0100
Batteryset 3	-0.1164	0.0100
Batteryset 4	0.2756	0.0100
Batteryset 5	-0.1405	0.0100
Random effects	**Variance/SD**	
Respondent		
Intercept	0.0004 / 0.0217	
AIC	-1351.4	

The significance of the fixed-effect factors was evaluated by means of the Wald test for the coefficients in the models, †p<0.1, *p<0.05, **p<0.01, ***p<0.001

Don't know

Regarding don't know responses, our results yield less don't know responding in CATI and CAPI than in Video-web for questions about political opinions and subjective well-being (see Table 4.4). The largest difference can be found between CATI and Video-web for questions about subjective wellbeing where the odds of a don't know answer in CATI are almost only one tenth (1/0.103) of the odds in Video-web. Contrary to our expectation, the odds for don't know answers in Text-web are also lower for both political opinion and subjective wellbeing questions, although the differences with Video-web are much smaller compared to CATI and CAPI and not significant. These results thus do not confirm our expectations that Video-web may help reducing satisficing compared to Text-web and provide answers more similar to CAPI and CATI.

Table 4.4 Don't know in Video-web

Fixed effects	**Estimate**	**Exp(B)**	**SE**	
topic politics				
Intercept	-4.7078***	0.0090	0.2884	
Mode(capi)	-1.5852***	0.2049	0.2864	
Mode(cati)	-1.1155***	0.3277	0.2673	
Mode (t-web)	-0.2821	0.7541	0.2136	
topic subjective wellbeing				
Intercept	-5.2035***	0.0054	0.3038	
Mode(capi)	-1.9985***	0.1355	0.5528	
Mode(cati)	-2.2688***	0.1034	0.5646	
Mode (t-web)	-0.4031	0.6681	0.3845	
topic social demographics				
Intercept	-4.6467***	0.0095	0.2883	
Mode(capi)	-0.8017*	0.4485	0.3863	
Mode(cati)	-0.1333	0.8751	0.3412	
Mode (t-web)	-0.3649	0.6942	0.3663	
Group (contact f2f choice)	0.0857	1.0895	0.1452	
Group (contact tel choice)	0.0108	1.0108	0.1522	
Gender (fem)	0.6714***	1.9570	0.1191	
Age (middle)	-0.0855	0.9180	0.1631	
Age (old)	-0.0631	0.9388	0.1645	
Educ (high)	-0.7437***	0.4753	0.1355	
Educ (unknown)	0.8501**	0.4273	0.3055	
Random effects	**Variance/SD**	**Correlations**		
Item				
Intercept	1.2479 / 1.1171			
Mode(capi)	0.8388 / 0.9158	0.34		
Mode(cati)	0.6865 / 0.8286	-0.02	0.90	
Mode (t-web)	0.3399 / 0.5830	0.57	0.78	0.45
Respondent				
Intercept	1.5517 / 1.2457			
AIC	11778.6			

The significance of the fixed-effect factors was evaluated by means of the Wald test for the coefficients in the models, †p<0.1, *p<0.05, **p<0.01, ***p<0.001

For the questions on social demographics, all mode effects were much smaller than for the other two topics, probably because the combination of self-administered modes and the political and subjective wellbeing items are more prone to don't know reporting.

Furthermore, mode choice does not seem to influence the reporting of don't knows as we did not find differences for the group of respondents that had no mode choice compared to the choice groups.

Additionally, the results also suggest a fairly large variation between the items regarding the odds for don't know answers and the mode effects on these odds. Even after controlling for question topic, the variances of the random effects for mode over the different items is considerably large. For example, the odds ratios for don't knows answers between CAPI and Video-web still differ by a factor 2.498 (exp(0.9158)) on average. This means that for some items, the effect of CAPI versus Video-web is much more extreme than the conclusions reported above, while for other items the effect of CAPI versus Video-web might be absent or even opposite. The control variable gender shows that females were almost twice as likely to give don't know answers than males. Older and middle-aged respondents reported less don't know answers than younger respondents, but this difference is not significant. As expected we find that higher educated respondents gave significantly less don't know answers than the lower educated (odds ratio 0.48). This outcome is similar to earlier findings on education and don't know responding (Holbrook et al., 2003; Krosnick et al., 2002). Probably, lower educated respondents invested less cognitive effort in answering the questions than the higher educated. For respondents with an unknown educational level we find that they were twice as likely to give don't know answers than the lower educated. These results may confirm that our control variables do correct for imbalanced selection between the modes.

4.4.2 Results Socially desirable responding

For socially desirable answering behavior, our results only show a significant difference between CATI and Video-web for the subjective wellbeing items (see Table 4.5); more socially desirable responses were found in CATI than in Video-web. CATI respondents were 1.5 times as likely to give a socially desirable response on subjective wellbeing items than Video-web respondents. We did not find this effect between CAPI and Video-web. This may be the result of interviewers' better opportunities to build rapport in CAPI, due to which socially desirable responding can be at a low level in CAPI (Holbrook et al., 2003). To verify whether CAPI indeed had a lower level of social desirability than CATI we ran the same model with CATI as a reference category. This model revealed only marginal significant effects. For the subjective well-being items CAPI had a lower score than CATI, for the other topics no significant effects were found. Although most differences between modes were small and not significant, the analysis does show that more socially desirable responses were given in CAPI and in CATI compared to Video-web for the political and subjective wellbeing items. This outcome means that by implementing subtle human cues in Video-web the degree of social presence is not as

Table 4.5 Socially desirable responding in Video-web

Fixed effects	**Estimate**	**Exp(B)**	**SE**	
topic politics				
Intercept	0.2712	1.3115	0.4108	
Mode(capi)	0.1251	1.1333	0.1257	
Mode(cati)	0.0814	1.0848	0.1335	
Mode (t-web)	-0.1252	0.8823	0.1016	
topic subjective wellbeing				
Intercept	-0.4513	0.6367	0.4770	
Mode(capi)	0.0974	1.1023	0.1696	
Mode(cati)	0.3863*	1.4715	0.1821	
Mode (t-web)	0.0011	1.0011	0.1767	
topic social demographics				
Intercept	0.9826	2.6715	0.9422	
Mode(capi)	-0.5259†	0.5910	0.2796	
Mode(cati)	-0.4683	0.6260	0.3016	
Mode (t-web)	-0.0484	0.9527	0.2579	
Group (contact f2f choice)	-0.1803*	0.8349	0.0738	
Group (contact tel choice)	-0.0720	0.9304	0.0779	
Gender (fem)	0.1085†	1.1146	0.0593	
Age (middle)	0.2087*	1.2321	0.0845	
Age (old)	0.3813***	1.4641	0.0839	
Educ (high)	0.4974***	1.6445	0.0666	
Educ (unknown)	0.0134	1.0135	0.1809	
Random effects	**Variance/SD**	**Correlations**		
Item				
Intercept	1.7106 / 1.3079			
Mode(capi)	0.0547 / 0.2340	-0.81		
Mode(cati)	0.0716 / 0.2678	-0.54	0.56	
Mode (t-web)	0.0032 / 0.0573	-0.49	0.91	0.40
Respondent				
Intercept	0.4189 / 0.6473			
AIC	17991.8			

The significance of the fixed-effect factors was evaluated by means of the Wald test for the coefficients in the models, †p<0.1, *p<0.05, **p<0.01, ***p<0.001

high as in interviewer-administered modes, possibly resulting in less socially desirable responding in Video-web.

Contrary to our expectation, the differences between Text-web and Video-web are also very small for all the items. So Video-web does not seem to provide significantly lower data quality than Text-web relating to socially desirable reporting. However, as noted before, the question topic and web-switch were confounded in our design. Respondents answering the subjective wellbeing questions in the Video-web mode, had answered the preceding politics questions in Text-web, and respondents answering the subjective wellbeing questions in Text-web had answered the political questions in Video-web. Thus, lack of effects for Text-web versus Video-web may be due to the switch of modes; after the switch respondents had experienced both modes. Furthermore, it is also possible that the human cues in Video-web were too subtle resulting in a mode that was very similar to a traditional Text-web survey. Assessment of socially desirable answering behavior was relevant for only two out of 53 of the social demographical questions. For these two questions the model shows that there is less socially desirable responding in CAPI, CATI and Text-web compared to Video-web, but the differences are not significant.

Additionally, significantly less socially desirable responses were found in the face-to-face contacted group with choice compared to the telephone contacted group without choice (odds ratio 0.84). The difference in responses between the telephone contacted group with choice and without choice was very small. Contacting sample members face-to-face and offering them a choice in mode seems to have a positive effect on socially desirable responding.

Furthermore, we did not find a very large variation between the items regarding the odds for socially desirable responses and the mode effects on these odds. The odds ratio for socially desirable responses differ by a factor 1.263 for CAPI and Video-web, by a factor 1.307 for CATI and Video-web, and by a factor 1.058 for Text-web and Video-web. This result indicates that the answers given in the modes were not very far from the mean and each other.

For the control variables age, and education our results show significant differences in response behavior. Compared to young respondents, middle-aged (odds ratio 1.23) and older respondents (odds ratio 1.46) gave more socially desirable responses. Possibly, older respondents are eager to present themselves in a favorable way because of their vulnerable position in society (Green et al., 2001). Noteworthy, the higher educated were 1.64 times as likely to give more socially desirable answers than lower educated respondents. We expected the opposite, but this effect can be explained on the basis of the topic of the questions and the fact that determining which answers are socially desirable is based on assumptions. For example, higher educated respondents might be more likely to vote and are very interested in politics. So possibly, the socially desirable answers were the honest answer to the items.

4.5 Discussion

As web surveys are used more and more, questions are raised about how data quality of this mode can be improved. The aim of this Chapter was to compare data quality of a Video-web survey to CAPI, CATI and a traditional Text-web survey. Because of the human cues in the Video-web mode, we expected satisficing behavior in Video-web to be at a lower level than Text-web, and at a more similar level in CAPI and CATI. Furthermore, we investigated the degree of socially desirable responding in Video-web in relation to the other three survey modes. Compared to Video-web, more socially desirable behavior was expected in CATI, slightly more in CAPI and less in Text-web.

To study satisficing we analyzed non-differentiation behavior and don't know reporting. For non-differentiation, no clear differences were found comparing CAPI, CATI, and Text-web to Video-web. This may explain why studies comparing Video-web to other modes have paid no attention to satisficing; it is possible that non-differentiation was analyzed but not reported. On the positive side, Video-web does not seem to deteriorate data quality either compared to the other three modes.

A lower level of don't know responding was found in CAPI and CATI compared to Video-web. This may demonstrate that the Video-web mode was not able to motivate respondents at the same level as the interviewer-administered modes. Interviewers may be persistent (either verbally or non-verbally) in obtaining a substantial response from respondents, in that way lowering the level of don't know responses. Due to the fact that Video-web provided no interaction, this type of persistence was not present in this mode. Furthermore, lower scores for don't know responding were also found for Text-web than Video-web but differences were not significant. However, we should be careful interpreting these outcomes despite of the fact that don't know reporting can be an appropriate indicator for survey satisficing (Krosnick, 1991). Taking don't know answers as an indicator for satisficing behavior is complicated. In this study, don't know options in the Text-web and Video-web mode were only offered by presenting the response option visually, while in the interviewer-administered modes don't know options were not always explicitly mentioned. To make the don't know option less prominent in the web surveys, this option was placed between brackets and presented as the last item of the list, which due to primacy effects would also lower the chance of respondents selecting this option.

Using a subtle humanized Video-web survey only resulted in significant differences between CATI and Video-web for the subjective wellbeing items. More socially desirable responses were provided in CATI than in the Video-web mode. Small non-significant differences were found between Video-web and the interviewer-administered modes for political and subjective wellbeing items; less socially desirable responses were provided in the Video-web mode compared to CATI and CAPI. Contrary to our expectation,

we also found small differences between Text-web and Video-web. Possible reasons for this result are the confounding of question topic and web-switch, the human cues in Video-web being too subtle, or that the selected questions might not have elicited much socially desirable response behavior. It is important to mention that socially desirable answering is based on assumptions; it is possible that differences in for instance the number of respondents who report to have voted reflects real differences between respondents. Although one could argue that differences between modes may therefore reflect selection effects, we have taken this into account by adding respondent characteristics that are likely corresponding to selection effects (e.g., education, age and gender). Nevertheless, not all possible underlying variables have been taken into account in the models.

One limitation of the current study is that mode switch and topic were confounded in the experimental design. Therefore, the differences we found for the subjective wellbeing and social demographical items could also be due to the mode switch that took place after the political items. Future studies should investigate which topics are more prone to sensitive reports without having respondents switching modes. Another point worth discussing is the possibility that lack of differences between Video-web and Text-web is due to respondents not listening to the video-recorded questions and instantly choosing response options. However, we do not think this is an explanation for our results for two reasons. First, it was not possible to give meaningful answers when not listening to the video since the question was not shown on the screen. Second, in almost all cases the question could not be derived from the response options (which were shown on the screen to allow respondents to select an alternative).

Further research in this area could explore how data quality in web surveys could be improved by using human features. Cassel and Miller (2008) suggested the idea of making use of Embodied Conversational Agents (ECA's). Through a webcam, ECA's can see and hear respondents, and can respond to nonverbal behavior (e.g., body posture) and verbal utterances. Some work on ECA's is executed by Conrad et al. (2008), and Malakhoff and Jans (2011), but using ECA's in surveys is still at an early stage.