

Mode Effects on Socially Desirable Responding in Web Surveys Compared to Face-to-Face and Telephone Surveys

Nejc Berzelak¹

Vasja Vehovar²

Abstract

This paper elaborates upon differences in socially desirable responding as being the result of mode effects between web, telephone, and face-to-face survey modes. Social desirability is one of the main threats to comparability of data between different modes. The paper conceptualises socially desirable responding as a specific type of mode effect, which is not only a result of inherent characteristics of a survey mode, but is also mediated and moderated by complex interdependencies of specific survey implementations, contextual factors, and characteristics and behaviours of respondents. While web surveys are generally less prone to socially desirable responding, it is essential to be wary of circumstances that may reduce the perceived privacy of the survey situation and lead to biased reporting. The presented empirical study analyses the answers to a large number of items used in a pilot implementation of the Generations and Gender Survey across the three modes to gain insights into the incidence of socially desirable responding and its role in the observed differences in estimates. The comparison of means, distributions, and proportions of extreme responses to scale questions is performed across 89 survey items. The results are in line with the previous findings on lower susceptibility of web surveys to social desirability bias. More importantly, the findings suggest that the problem of socially desirable responding is likely to be a major contributor to the differences in mean estimates, response distributions, and the level of extreme responding between the studied modes.

1 Introduction

Web surveys have become one of the main data collection tools in many research areas. However, the transition from traditional survey modes to web data collection often proves challenging and requires careful elaboration to assure effective implementation of web surveys and their compliance with the data quality requirements.

As previously recognized by Deming (1944), potential differences in answers caused by a mode change are one of the central challenges of introducing any new survey mode. The underlying causes of such differences have been embraced by the concept of “mode

¹ Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia; nejc.berzelak@fdv.uni-lj.si

² Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia; vasja.vehovar@fdv.uni-lj.si

effects” (e.g., Aquilino & Lo Sciuto, 1990), a type of measurement error that arises because a specific survey mode is used for data collection. Mode effects are of particular concern in regard to mixed-mode designs as well as changing survey modes in longitudinal studies. However, even with one-time and single-mode surveys, potential influences of the survey mode on respondents’ answers may introduce significant measurement biases into the data.

Between-mode differences in answers to sensitive questions are among the most commonly observed consequences of mode effects. It was discovered early on that, in the presence of an interviewer, respondents are less willing to report behaviours and attitudes that are deemed socially undesirable (Bradburn, Sudman, Blair and Stocking, 1978; Hochstim, 1967). These findings were replicated by many studies, including those comparing web surveys to traditional survey modes (Tourangeau, Conrad and Couper, 2013). However, the social desirability bias as a mode effect has been rarely studied by observing its effects on the differences in estimates between different survey modes across a large number of questionnaire items.

This paper contributes new insights to the various levels of socially desirable responding by evaluating the estimates and response patterns in web surveys compared to telephone and face-to-face surveys. We begin by establishing a conceptual framework that links mode characteristics and mode effects to socially desirable responding. This introduces some important considerations about factors that may increase or decrease the incidence of social desirability bias in surveys. In the empirical section, we rely on the data from an experimental survey, designed to study differences in the incidence of socially desirable responding between web and two interviewer-administered modes. The analyses are performed across a number of items, taking into account the susceptibility of items to social desirability. The obtained findings offer further indications about the role of social desirability bias in the observed differences in estimates caused by mode effects.

2 Background

Nederhof (1985) defines social desirability as a reflection of the tendency, on behalf of the subjects, to deny socially undesirable traits and to claim socially desirable ones, as well as the tendency to say things which place the speaker in a favourable light. In surveys, this is manifested either by underrepresentation of undesirable behaviours or overrepresentation of desirable ones.

How prone a survey question is to social desirability mainly depends on whether some answers to the question are more acceptable than others according to the relevant social norms (Tourangeau, Rips and Rasinski, 2000; Tourangeau & Yan, 2007). Respondents are likely to distort their responses in the socially desirable direction when they feel their attitudes, traits, or behaviours are not favoured by social norms (Bradburn et al., 1978). What is deemed a desirable response to a question can therefore vary between respondents from different social backgrounds and environments (Näher & Krumpal, 2012).

The level of distorted reporting also importantly depends on a respondent’s characteristics, with some respondents being more inclined to provide socially desirable responses owing to their personality characteristics, such as conformity or the need for social approval (Tourangeau et al., 2000). Paulhus (2002) further explains that respondents may

distort their answers either due to purposive impression management or unrealistic self-deception. While the differences in socially desirable responding between survey modes are likely caused by the varying incidence of impression management rather than self-deception, in the present study, we do not distinguish between the two aspects, but refer to social desirability as a general term.

Before turning to the discussion of how specific characteristics of survey modes contribute to differing levels of socially desirable responding, it is important to briefly outline the mechanism of social desirability in the context of the survey response process. Response errors due to social desirability stem from the respondents' altered performance of the response process, which, according to Tourangeau et al. (2000), consists of question comprehension, information retrieval, judgment of the retrieved information, and reporting of an answer in line with the requirements of a survey question. When a respondent resorts to social desirability, they may perform the response process thoroughly and derive an accurate answer to a survey question, but may ultimately decide to edit the answer before reporting it (Cannell, Miller and Oksenberg, 1981; Tourangeau et al., 2000). The overediting of an answer in the reporting stage has been empirically demonstrated by Holtgraves (2004), who confirmed longer response times when social desirability was presumably affecting the response process.

2.1 Mode Characteristics and Mode Effects

Each survey mode can be described by a set of specific characteristics that distinguish it from other modes. Although neither the term "survey mode" nor identifying properties of individual modes have been ultimately defined in survey methodology, Berzelak (2014) used previous conceptualisations by various authors to identify the following six characteristics that distinguish between common survey modes: information transmission medium, main question presentation channel, response channel, interviewer involvement during the data collection, and closeness of interaction with the respondent. Table 1 specifies the inherent characteristics of the three modes within the scope of this paper.

The inherent characteristics of a survey mode determine basic principles of communication and information transmission between a respondent and the survey questionnaire. They present the foundation and constrain the range of possible options for building and implementing the survey design. However, the chosen survey mode still allows for many variations in the implementation of the actual survey. Depending on a variety of design decisions, individual surveys implemented using the same mode may vary substantially in characteristics, such as the level of control over the survey situation left to the respondent, flexibility of the question presentation order, availability of verbal, nonverbal, and paralinguistic communication channels, sense of impersonality, pace of interview, and others (Berzelak, 2014; Couper, 2011; de Leeuw, 1992).

Another set of survey characteristics is grounded in specific social and individual contexts in which surveying takes place. Factors like familiarity and the use of the survey medium (e.g., the telephone or the World Wide Web), sincerity of the purpose conveyed by the medium, social and individual perceptions of an appropriate pace of conversation (de Leeuw, 1992, 2005), and the degree of privacy available to the respondent (Couper, 2011) are only some examples of contextual characteristics. Data collection procedures

Table 1: Selected and observed questionnaire length and complexity determinants

	Web	CAPI	CATI
Information transmission medium	Internet	In person	Telephone
Main question presentation channel	Visual	Auditory (visual supplement)	Auditory
Response channel	Electronic input	Oral	Oral
Interviewer involvement	No interviewer (self-administered)	Interviewer administers	Interviewer administered
Closeness of interaction between interviewer and respondent	Not applicable (no interviewer)	Face-to-face	Remote
Use of computer technology in data collection	Used by respondents	Used by interviewer	Used by interviewer

may be further affected by factors such as social norms and values, respondents' characteristics and abilities, and many other factors specific to the survey environment.

The distinction between inherent, implementation-specific, and contextual characteristics of a survey mode is beneficial for the conceptualisation of mode effects. Surveying is a form of specific and standardised conversation (Tourangeau & Rasinski, 1988) that is importantly determined by the characteristics of a survey mode. How information is transmitted to and from the respondent, to what degree interviewers are involved in the communication, and what medium and tools are used for this purpose determine not only the nature of the communication itself, but also the respondent's cognitive tasks and behaviour in the survey situation.

Mode effects are the result of influences of a survey mode's inherent characteristics on the response process and, consequently, on the obtained survey estimates. However, such influences are, to a large extent, moderated and mediated by further complex interdependences of implementation-specific and contextual factors, content of the survey questionnaire, and characteristics and behaviours of individual respondents. For example, a study by Bennink et al. (2013) demonstrated that mode differences between web and face-to-face surveys emerge only with some combinations of visual presentation, question topics, availability of a non-substantive response category, order of the answer categories, mandatory answers, and so on. The incidence of mode effects, therefore, depends on several factors that may be only indirectly related to the characteristics of a survey mode. In the next section, we discuss key factors of mode effects that result in socially desirable responding in different survey modes, with an emphasis on the three modes within the scope of the presented study.

2.2 Social Desirability as the Result of Mode Effects

The respondents' perceptions of privacy in a survey situation were recognised early on as a major factor of differences in social desirability between modes, particularly when respondents may be concerned about the interviewer's approval or disapproval regarding the reported answer. A study by Hochstim (1967) found that mail surveys produce higher reporting on sensitive questions than face-to-face and telephone surveys. This general pattern was later confirmed by many further experimental verifications, summarised in a meta-analysis by Tourangeau and Yan (2007). Consistently, web surveys elicited better performance on sensitive topics than telephone (Chang & Krosnick, 2009; Jäckle, Roberts and Lynn, 2006; Lee, Kim, Couper and Woo, 2018; Lozar Manfreda & Vehovar, 2002; Milton, Ellis, Davenport, Burns and Hickie, 2017) and face-to-face interviewing (Jäckle et al., 2006; Zhang, Kuchinke, Woud, Velten and Margraf, 2017). A meta-analysis of ten studies comparing the web mode to other survey modes by Tourangeau et al. (2013) further confirmed these observations.

The differences in privacy perceptions offered by the various survey modes predominantly stem from two inherent characteristics of a mode: the involvement of an interviewer, and the closeness of interaction between the interviewer and the respondent. The impersonal nature of interaction in self-administered modes reduces the respondent's sense of disclosing their answers to a third party (Tourangeau et al., 2000; Tourangeau & Yan, 2007), subsequently lowering their tendency toward socially desirable reporting compared to interviewer-administered surveys. Theoretical elaborations and empirical evidence, therefore, strongly support better performance of web surveys in terms of socially desirable reporting compared to telephone and face-to-face surveys. Nevertheless, it is important to elaborate upon some implementation-specific and contextual factors that may lead to reduced privacy perceptions with the web mode and may increase the incidence of social desirability.

According to de Leeuw (2008), some respondents may experience a lower degree of privacy with the use of computers, while others may perceive computerised data to be more secure against third-party access. While empirical evidence on such effects is largely inconsistent (Dodou & de Winter, 2014; Fang, Wen and Prybutok, 2014), there is some indication that respondents might be less willing to disclose sensitive information in computerised surveys than in paper-based surveys when they are aware that the survey participation is not anonymous (Smither, Walker and Yap, 2004).

The respondents' sense of privacy in web surveys may also be reduced by specific survey implementations that use highly interactive questionnaire features, in which case Tourangeau and Yan (2007) caution the effect of media presence. Interactive capabilities of computerised questionnaires, such as video clips of individuals asking questions, have been known to potentially create the illusion of the interviewer's presence and trigger effects commonly found with interviewer-administered modes (Krysan & Couper, 2003). Overexploiting computerisation to humanise web questionnaires can thus reduce the benefits of higher reporting on sensitive topics by introducing mode effects similar to those caused by the presence of an interviewer.

Potentially negative influences on privacy perceptions in web questionnaires may also arise from the survey environment. Self-administration allows respondents to choose the place and setting of survey completion, which gives researchers little control over the en-

vironment in which surveying occurs. Research has shown that the presence of others decreases the perception of privacy and increases false reporting even when the survey is self-administered (Aquilino, Wright and Supple, 2000; Beebe, Harrison, Mcrae, Anderson and Fulkerson, 1998; Castelli & Tomelleri, 2008). With a high flexibility of settings in which the survey can be completed using the web as a medium of information transmission, web surveys can be expected to produce a higher variability of effects related to the surveying environment.

Finally, the perceptions of the web as a medium itself may influence the willingness of respondents to disclose sensitive information. Characteristics of the medium, as well as social and personal attitudes toward it, importantly influence the mode's ability to convey legitimacy of the survey, which has been linked to the sincerity of reports on sensitive behaviours (Tourangeau et al., 2000). The absence of interviewers in web surveys limits their capacity to establish legitimacy. Furthermore, attitudes regarding the web can further contribute to this issue. A large amount of spam e-mail messages, fraudulent websites, media-fostered privacy concerns, and reports on security breaches are some examples of perils that undermine the general trustworthiness of the web. This is most directly reflected in lower response rates in web surveys compared to other modes (Lozar Manfreda, Berzelak, Vehovar, Bosnjak and Haas, 2008).

In summary, self-administration is expected and has been shown by previous studies to importantly reduce the tendencies toward socially desirable responding in web surveys compared to telephone and face-to-face surveys. Yet, a variety of implementation-related and contextual factors specific to web surveys can still negatively affect an individual's privacy perceptions and potentially increase the social desirability bias. Among the sources of such issues are negative influences of survey legitimacy, specific environments in which surveying occurs, negative attitudes toward computer technology, and effects of media presence caused by highly interactive questionnaire features.

3 Empirical Study Description

The empirical study analyses differences in answers obtained by an experimental survey conducted using web, telephone, and face-to-face modes. Its primary focus is to identify the presence of mode effects on socially desirable responding and to compare the magnitude and direction of between-mode differences on items that are more likely or less likely prone to the social desirability bias.

In contrast to most of the existing studies in the field, the presented analyses are performed across a large number of questions and items rather than focusing on a few target variables. While this occurs at the expense of less thorough item-level analysis, it offers the benefit of a more general evaluation of both the identified effects and their consistency (Jäckle et al., 2006).

3.1 Hypotheses

Following the conceptual elaboration and the aim of the study, the analyses strive to verify the following hypotheses:

- Hypothesis 1: *Between-mode differences in mean estimates and distributions are more commonly observed among the studied items that are prone to social desirability.* Differences in social desirability among survey modes is one of the most consistently observed results of mode effects. More significant and stronger effects can be therefore expected on items prone to socially desirable responding.
- Hypothesis 2: *Web respondents express lower tendencies of socially desirable responding.* The likelihood of socially desirable responding is importantly determined by the respondent's perception of privacy. Despite possible implementation-specific and contextual variations that can reduce privacy perceptions in web surveys, the self-administered nature of this mode is likely to result in web respondents being less reluctant to provide answers that may be regarded as socially less desirable.
- Hypothesis 3: *Respondents of the two interviewer-administered modes more likely select extreme scale answers than web respondents, with more pronounced effects observed on questions prone to social desirability.* Some previous studies (de Leeuw, Hox and Scherpenzeel, 2010; Dillman et al., 2009) have found lower levels of extreme responding in web surveys compared to interviewer-administered surveys. While various mode characteristics may contribute to this, including primacy or recency effects due to different question presentation channels, Ye, Fulton, and Tourangeau (2011) suggested the important role of social desirability in heightening the incidence of extreme responding in interviewer-administered modes.

3.2 Questionnaire and Analysed Items

The data for this study were collected as part of an experimental pilot implementation of the Generations and Gender Survey (GGS) in Slovenia. The objective of the pilot was to evaluate different aspects of data accuracy and comparability among web survey, telephone survey (CATI), and face-to-face survey (CAPI).

The GGS questionnaire consisted of eleven thematic modules and covered a variety of demographics and related topics. It contained approximately 340 questions and items with an average face-to-face duration of approximately 45 minutes.

Single-item and multiple-item scale questions with the numbers of answer categories ranging from two ("yes"/"no") to eleven were selected for the analysis. Questions applicable only to some respondents and those included in various within-questionnaire experimental manipulations were excluded to enable a comparison of the results across several questions for the same set of respondents.

The final selection included 89 items covering a broad range of topics related to health, personality, income, religiosity, life satisfaction, attitudes toward family and gender, and the survey experience. Additional details of the selected questions, including the full wording of each, are provided in the online material supplementary to the paper.

3.3 Sampling and Response Rates

The sample of respondents was obtained from an online access panel maintained by the leading Slovenian market research company Valicon. Although the sampling method

used was non-probability, it offers the important advantage of minimising confounding non-coverage effects between modes and reaching a demographically diverse population.

The company obtained the necessary contact information (e-mail, postal address, and telephone number) from 847 panel members who were randomly assigned to one of the three modes. The overall final response rates were 87 % for the web mode, 61 % for CATI, and 74 % for CAPI. In total, data from 623 respondents were used for the analyses. There were very minor variations in a generally low breakoff rate and item nonresponse rates across the analysed items. The number of cases for each item, therefore, vary between 611 and 618. A comparison of socio-demographic composition of the three mode groups revealed relatively small and nonsignificant differences.

3.4 Identification of Items Susceptible to Social Desirability

To identify questionnaire items prone to social desirability bias, each item was rated by three survey methodology experts using a developed coding scheme. The susceptibility to social desirability was measured in line with the definition of purposive impression management as a potential for overclaiming as a means of presenting oneself in a more favourable light (Nederhof, 1985; Paulhus, 2002). The potential for overclaiming was rated on a three-point scale, with the value “1” indicating no potential and the value “3” indicating high potential. An item was considered potentially susceptible to social desirability if the mean rating by three reviewers was “2” or higher. The majority (76 %) of 89 items included in the analysis was found to be potentially susceptible. An overall agreement between experts, measured using Krippendorff’s alpha, was moderate (0.51), but was considered acceptable.

Because many of the analysed items were opinion-based, it was often difficult to identify scale pole with a higher likelihood of selection under the influence of social desirability. Depending upon which social values and norms are invoked by a respondent during the response process, answers to such questions may be often shifted to either side of the scale (Näher & Krumpal, 2012; Schwarz & Oyserman, 2001). Without taking the risk of over-guessing, the direction was assigned to 49 items.

3.5 Approach to Analysis

Like a majority of other empirical studies, the analysis of mode effects was grounded on the differences between the web mode and both interviewer-administered modes (CATI and CAPI). Because the true value of any estimated parameter is unknown, the decision about which mode’s effects are causing the between-mode difference in the estimates needs to be predominantly theoretically-driven. In studying social desirability bias, the standard assumption is that the mode with a higher tendency toward socially desirable answers is more severely affected (Bradburn et al., 1978).

The analysis of between-mode differences was performed using ordinary least squares (OLS) regressions, logistic regressions, and partial proportional odds models (PPO). Although no significant differences were found in the socio-demographic composition of the experimental groups, the effects were controlled for basic socio-demographics (gender, age, and higher education) to reduce the potential confounding influence of differential unit nonresponse.

The use of both OLS and PPO is beneficial because the former measures the effects on means and the latter helps reveal more subtle patterns of mode effects by considering the ordinal-level structure of variables. This is important since mode effects may differently influence some rather than all response categories (Jäckle et al., 2006).

We fit one or more of these models to each of the 89 studied items. There is no universal agreement on if and how the problem of testing many null hypotheses should be adjusted for, particularly when the study objectives are exploratory. Rothman (1990) cautions that one should not be too conservative if such adjustments could lead to missing potentially important findings. We set a general threshold for interpretation of the results as significant at 0.01; however, wherever applicable, we also report a significance at the levels adjusted using the Benjamini-Yekutieli method (Benjamini & Yekutieli, 2001) and the more conservative Bonferroni method (see supplementary material).

Some comparisons are additionally based on the calculated effect sizes for explained variance (partial η^2) and mean difference (Glass's delta, Δ_G , calculated as the difference in means between the web mode and each of the self-administered modes, divided by the standard deviation for this item in the web mode). For some descriptive comparisons of effects across several items, the effect sizes were summarised using simple averages (Turner & Bernard, 2006).

4 Results

The presentation of the results follows the sequence of the three hypotheses within the scope of this study. However, before verifying the hypotheses, we expose some general patterns of differences in means and distributions between modes, which are important for further interpretation of the results.

Because the analyses are based on fitting the models to a large number of items and details of individual models bring little added value apart from transparency of the analyses, only some details at the level of individual items are presented in the text and the rest is available in online supplementary material. Question names are included for easier reference.

4.1 Overview of Between-Mode Differences in Estimates

We begin by analysing differences in means of target items between modes using OLS models, where mode and socio-demographic control variables were fitted to each of 73 items with four or more ordered scale points (Table 2). The means are significantly different at $p < 0.01$ between web and CATI on sixteen items (22%), and between web and CAPI on twenty items (27%). For six items, the mean value obtained by the web mode significantly differs from both interviewer-administered modes.

Table 2: The differences in adjusted means between modes and their effect sizes for items with four or more scale points (items with a significant effect of mode at $p < 0.01$ on at least one comparison)

Item	Web	CATI compared to Web		CAPI compared to Web	
	\bar{X}_m	Δ_{C-W}	Δ_G	Δ_{C-W}	Δ_G
Personality (7.05)					
<i>1 – does not apply / ... / 7 – applies perfectly</i>					
Does through job (b)	5.490	0.200	0.165	0.470 ^{##}	0.389
Talkative (c)	5.232	0.269	0.183	0.451 [#]	0.307
Outgoing, sociable (h)	4.999	+0.498 ^{##}	0.364	+0.668 ^{##}	0.488
Values artistic, aesthetic experience (j)	4.571	-0.034	-0.023	+0.579 ^{##}	0.392
Relaxed (n)	4.719	+0.467 [#]	0.339	+0.320	0.232
Sense of control (7.06)					
<i>1 – strongly agree / ... / 5 – strongly disagree</i>					
Cannot solve own problems (a)	3.568	+0.354 [#]	0.318	+0.407 ^{##}	0.366
Feel pushed around (b)	3.635	+0.287	0.261	+0.435 [#]	0.395
Depression (7.06)					
<i>1 – seldom or never / ... / 4 – most or all of the time</i>					
Could not shake off blues (a)	1.407	-0.284 ^{##}	-0.445	-0.015	-0.023
Felt depressed (b)	1.332	-0.220 ^{##}	-0.426	-0.082	-0.158
Thought life is a failure (c)	1.351	-0.241 ^{##}	-0.568	-0.167 [#]	-0.394
Felt fearful (d)	1.526	-0.364 ^{##}	-0.674	-0.181 [#]	-0.335
Felt lonely (e)	1.486	-0.313 ^{##}	-0.502	-0.165 [#]	-0.265
Had crying spells (f)	1.293	-0.240 ^{##}	-0.501	-0.099	-0.206
Felt sad (g)	1.642	-0.351 ^{##}	-0.571	-0.126	-0.205
Income adequacy (10.02)					
<i>1 – with great difficulty / ... / 4 – very easily</i>					
Making ends meet	3.402	+0.227	0.184	+0.435 ^{##}	0.353
Imp. of religious ceremonies (10.04)					
<i>1 – strongly agree / ... / 4 – strongly disagree</i>					
Religious wedding (b)	3.782	+0.403 [#]	0.340	+0.160	0.135
Planning for future (10.04)					
<i>1 – I plan f. fut. as much as possible / ... / 10 – I just take each day as it comes</i>					
Planning for future	4.700	-0.296	-0.116	-0.903 ^{##}	-0.354

... continued

Item	Web	CATI compared to Web		CAPI compared to Web	
	\bar{X}_m	Δ_{C-W}	Δ_G	Δ_{C-W}	Δ_G
Marriage and children (11.08)					
<i>1 – strongly agree / ... / 5 – strongly disagree</i>					
Living unmarried together all right (b)	2.005	-0.040	-0.040	-0.284 ^{##}	-0.288
Divorce having children all right (d)	1.838	+0.273 [#]	0.254	+0.036	0.034
Woman w/o stable relationship with man having a child (h)	2.489	-0.217	-0.210	-0.385 ^{##}	-0.373
Elderly-care responsibilities (11.11)					
<i>1 – strongly agree / ... / 5 – strongly disagree</i>					
Children should adjust work to parents' needs (b)	3.367	+0.108	0.108	+0.277 [#]	0.277
Children should financially help parents (c)	2.402	+0.307 [#]	0.333	+0.143	0.154
Gender roles (11.11)					
<i>1 – strongly agree / ... / 5 – strongly disagree</i>					
Women really want home and children (a)	3.363	-0.181	-0.153	-0.311 [#]	-0.262
Man's task earning, woman's family (c)	4.054	+0.118	0.136	+0.261 [#]	0.302
Not good if woman works, man cares for children (d)	3.593	-0.003	-0.003	-0.320 ^{**}	-0.239
Working woman same relation with child (e)	2.212	-0.235	-0.233	-0.415 ^{##}	-0.412
Family life suffers because men too concentrated on work (h)	2.798	+0.375 [#]	0.295	+0.199	0.157
Survey feedback (12.02)					
<i>1 – definitely not / ... / 5 – definitely yes</i>					
Questions difficult (a)	1.731	-0.099	-0.125	-0.276 [#]	-0.347
Questions made think (c)	3.697	-0.513 ^{##}	-0.375	-0.626 ^{##}	-0.458
Questionnaire too long (e)	1.991	+0.847 ^{##}	0.714	-0.073	-0.062
Mean $ \Delta_G $	-	-	0.171	-	0.188

Note: \bar{X}_m – the marginal mean based on the OLS regression; Δ_{C-W} – the difference between the marginal means of Web and each of the compared mode; control variables: gender, age, and higher education; ** $p < 0.01$, # $p < \alpha_{yek} = 0.0052$, ## $p < \alpha_{bnf} = 0.0006$

The differences in means between the web and interviewer-administered modes are comparatively small; the mean absolute Glass's delta (Δ_G) is 0.171 for the difference between web and CATI and 0.188 for the difference between web and CAPI. The largest effect ($\Delta_G = 0.71$) was observed between web and CATI respondents for the question regarding the questionnaire length (Q12.02E). Unsurprisingly, CATI respondents experienced the lengthy questionnaire as "too long" to a larger degree than did web respondents. Other medium-to-large effect sizes ($|\Delta_G| > 0.5$) were identified in web-CATI comparison for a majority of depression scale items (Q7.09), with telephone respondents claiming, on average, lower frequency of all depression symptoms. Effects of a similar size for web-CAPI comparison occurred less frequently. The most notably highlighted differences include higher self-portrayal of CAPI respondents as being outgoing and sociable (Q7.05H, $\Delta_G = 0.49$), and higher reporting of web respondents about "being made to think by questions" (Q12.02C, 0.46).

The analysis with partial proportional odds modelling (PPO) offers additional insight into potential mode effects by exploring differences at the level of individual scale values. We attempted to fit the models for all 89 selected items, but the estimation process failed to converge for nine of them, presumably due to very low cell frequencies of some answer categories. The effect of mode was found to be significant at $p < 0.01$ for 33 of 80 (41 %) items in the web-CATI comparisons and 39 (49 %) items in the web-CAPI comparisons.

The PPO models are beneficial because they expose answer categories with the strongest reflection of differences between the web and the compared mode, which is important when studying social desirability bias. Four specific patterns of significant differences in answers between modes were identified across the items. The first two patterns include items for which web respondents tended to select higher or lower responses across the whole range of response values. The second two patterns relate to items for which web mode significantly differed only in a less frequent selection of upper or lower extreme answers. The four patterns can be summarised as follows:

1. Generally higher responses in the web mode

- (a) Web respondents are generally more likely than CATI respondents to report:
 - Not having enough close people (Q7.08F).
 - Being less able to shake off morning blues (Q7.09A), feeling depressed (Q7.09B), fearful (Q7.09D), lonely (Q7.09E), and sad (Q7.09G) more.
 - Inability to afford monthly dining out (Q10.03D).
 - Being made to think by the questions more (Q12.02C).
- (b) Web respondents are generally more likely than CAPI respondents to report:
 - Having worse health (Q7.02).
 - Being more easily nervous (Q7.05I).
 - Claim not to have enough close people (Q7.08F).
 - Feeling fearful (Q7.09D) and lonely (Q7.09E) more.
 - Inability to afford one-week holidays (Q10.03B), furniture replacement (Q10.03C), and monthly dining out (Q10.03D).
 - Taking each day more as it comes instead of Planning for future (Q11.07).

- Less agreement and more disagreement toward living together unmarried (Q11.08B), woman having a child without stable relationship with a man (Q11.08H), and equality of relationship between working woman and her child (Q11.12E).
- Finding questions more difficult (Q12.02A), and being made to think by the questions more (Q12.02C).

2. Generally lower responses in the web mode

- (a) Web respondents are generally more likely than CATI respondents to report:
- Being less outgoing and sociable (Q7.05H) and being less relaxed (Q7.05N).
 - More agreement and less disagreement with the importance of religious wedding (Q11.04B).
 - Not finding questionnaire too long (Q12.02E).
- (b) Web respondents are generally more likely than CAPI respondents to report:
- Being less outgoing and sociable (Q7.05H), less value artistic and aesthetic experience (Q7.05J) and being less considerate and kind (Q7.05K).
 - Being more agreeable and less disagreeable that they cannot solve own problems (Q7.06A) and feel being pushed around (Q7.06B).
 - Experiencing sense of emptiness (Q7.08B).
 - Having more difficulties making ends meet with their income (Q10.02).
 - Being unable to pay for utilities (Q10.04C) and loans (Q10.04D) in last 12 months.
 - Being more agreeable and less disagreeable that children should adjust work to the needs of their parents (Q11.11B) and that man's task is earning while woman's task is family (Q11.12C).
 - Finding the overall survey experience more unpleasant and less enjoyable (Q12.01).

3. Less high extreme responses in the web mode

- (a) Web respondents are generally more likely than CATI respondents to report:
- Extremely disagree that they cannot solve own problems (Q7.06A) and feel pushed around (Q7.06B).
 - Extremely disagree that it is important for an infant to be registered in a religious ceremony (Q11.04A), that marriage is outdated (Q11.08A) and should not end (Q11.08C), grandparents should help childcare (Q11.10A), parents should adapt life to help adult children (Q11.10C), man's task is earning while woman's task is family (Q11.12C), and that family life suffers if mother works (Q11.12G).
- (b) Web respondents are generally more likely than CAPI respondents to report:
- Extremely disagree that grandparents should help childcare (Q11.10A), children should live with parents for care (Q11.11D), and that family life suffers because men are too concentrated on work (Q11.12H).

4. Less low extreme responses in the web mode

(a) Web respondents are generally more likely than CATI respondents to report:

- Strongly agree that they have little control over things (Q7.06C).
- Claim with certainty to have plenty of people to lean on (Q7.08A), and many people to count on (Q7.08D).
- Strongly agree that living unmarried together is all right (Q11.08B), mother and father are needed for a happy child (Q11.08G), children should care for parents (Q11.11A), what women really want is home and children (Q11.12A), and that a working woman have the same relation with her child (Q11.12E).

(b) Web respondents are generally more likely than CAPI respondents to report:

- Strongly agree that they have little control over things (Q7.06C).
- Claim with certainty to have plenty of people to lean on (Q7.08A).
- Strongly agree that parents should adapt life to help adult children (Q11.10C), children should care for parents (Q11.11A), and that surveys enable own opinion articulation (Q12.05C).

5. Other patterns

(a) Web respondents are generally more likely than CATI respondents to report:

- Web respondents less likely disagree and strongly disagree that religious funeral is important (Q11.04C)
- Web respondents are more likely agreeable or neutral that children should financially help parents (Q11.11C).
- Web respondents are more likely agreeable or neutral that family life suffers because men are too concentrated on work (Q11.12H)
- Web respondents less likely strongly agree, but are also less likely disagreeable that surveys enable own opinion articulation (Q12.05C)

(b) Web respondents are generally more likely than CAPI respondents to report:

- Web respondents less likely answer 6 or 7 (applies perfectly) for having a forgiving nature (Q7.05F).
- Web respondents are more likely to disagree or strongly disagree that marriage should not end (Q11.08C).
- Web respondents are more likely to be agreeable that children should financially help parents (Q11.11C).
- Web respondents are less likely to strongly agree, but are also less likely to be disagreeable that what women really want is home and children (Q11.12A).
- Web respondents less likely agreeable that it is not good if woman works and man cares for children (Q11.12D).

Before turning to a detailed analysis of the direction of observed effects, it is worth drawing attention to the difference in the number of identified significant effects using the two modelling approaches. Because the PPO models did not converge for all items and OLS regressions were used only for items with four or more answer categories, both models were successfully estimated for 64 items. The ordinal-level approach revealed thirteen more items with a significant difference when comparing web and CATI, and seventeen more items when comparing between web and CAPI. All effects significant in the OLS regressions were also found significant by the PPO models. This demonstrates that mode can affect a specific part of the variable's distribution without significantly altering the mean; such effects go unnoticed in mean comparisons.

4.2 Mode Effects on Socially Desirable Responding

Of 68 items rated as potentially prone to social desirability bias (see the supplementary material for the list of these items), differences between modes were found significant either by OLS or PPO models at $p < 0.01$ for 37 % of these items in web-CATI comparisons and on 51 % of them in web-CAPI comparisons. As evident from Table 3, significant effects are generally more frequently observed on items prone to social desirability bias, with the exception of web-CATI comparison using the PPO models. Despite this inconsistency, the role of social desirability tendencies in identified mode effects is further strengthened by somewhat larger mean effect sizes for mean differences (Δ_G) on susceptible items. Although the effect sizes are generally small, this observation is largely in line with the hypothesis of more commonly found between-mode differences among the studied items that are prone to social desirability; this is particularly true for the differences between web and CAPI modes.

Table 3: Number of significant effects and effect size for mean difference found using the two modelling approaches by susceptibility of items to social desirability bias

	Web-CATI		Web-CAPI	
	Prone	Not prone	Prone	Not prone
OLS				
Tested items	52	21	52	21
Sig. effects ($p < 0.01$)	13	3	19	1
Mean $ \Delta_G $ for mean diff.	0.19	0.14	0.22	0.11
PPO				
Tested items	61	19	61	19
Sig. effects ($p < 0.01$)	22	11	32	7

Differences in socially desirable responding between modes can be most thoroughly examined for 49 items for which we were able to identify the likely direction of socially desirable responding (which is the most socially desirable option).

A quick inspection of specific response patterns listed in the previous section already gives support to the hypothesis of lower social desirability tendencies among web respondents. In a vast majority of cases, web respondents show significantly lower odds of

Table 4: The number of items with identified likely direction of socially desirable responding and a significant effect of mode by the pattern of difference identified using PPO models

Pattern of difference	Lower response desirable		Higher response desirable	
	Web vs. CATI	Web vs. CAPI	Web vs. CATI	Web vs. CAPI
Generally higher responses	7 ^a	9 ^a	1	1
Generally lower responses	0	0	2 ^a	10 ^a
Less high extreme	0	0	2 ^a	0 ^a
Less low extreme	2 ^a	1 ^a	1	1
Other patterns	0	0	0	1
No significant effect	11	10	18	11
Total items	20	20	24	24

^a Cell represents response pattern that is consistent with the hypothesised lower social desirability tendency among Web respondents

selecting more socially desirable answers, either across all categories (pattern types 1 and 2) or primarily at the extreme values (types 3 and 4). This is further demonstrated by the counts of items in the labelled cells of Table 4 that are consistent with lower social desirability tendencies of web respondents. Only two items for which PPO models indicated significant mode difference deviate from this: compared to both interviewer-administered modes, web respondents have lower odds of strongly agreeing to having little control over things (Q7.06C), and generally claim to having felt obligated to think more deeply about the questions (Q12.02C).

Lower impression management tendencies of web respondents are also reflected in mean differences, calculated using OLS regressions for items with four or more scale values. The most obvious advantage over telephone interviewing is observed for depression scale items (question 7.09), where absolute effect sizes ($|\Delta_G|$) of mean difference range from 0.426 to 0.674. Interestingly, face-to-face respondents reported the presence of depression symptoms much more frequently, although still significantly less so than web respondents ($|\Delta_{G_{\max}}| = 0.394$). An opposite example includes finance-related questions about affordable goods (Q10.03) and payment inability (Q10.04), where the difference against web is higher among CAPI than CATI respondents.

The presented findings strongly confirm the hypothesis on generally lower social desirability tendencies among web respondents. An additional review of the response patterns listed in the previous section also indicates a relatively consistent performance of the web mode on opinion items, for which we did not identify the likely direction of socially desirable answers. The most notably highlighted area is the more traditional (or less liberal) position of web respondents on most opinion items regarding marriage (Q11.08) and gender roles (Q11.12).

4.3 Extreme Responding and Social Desirability

Some interesting initial information about the extreme responding in the web mode compared to the two interviewer-administered modes can be obtained from the listed patterns

Table 5: Median and mean odds ratios for extreme responses by susceptibility of items to social desirability bias

	Median (mean) OR for lower extreme answers compared to Web		Median (mean) OR for upper extreme answers compared to Web		No. of items
	CATI	CAPI	CATI	CAPI	
Lower desirable	2.71 ^a (3.02)	1.72 ^a (1.70)	0.88 (0.89)	1.00 (1.02)	17
Upper desirable	1.11 (1.68)	1.00 (1.51)	1.33 ^a (1.44)	1.62 ^a (1.65)	22
Not prone to soc. des.	1.57 (1.58)	1.36 (1.43)	2.06 (2.93)	1.54 (1.82)	21
All items	1.68 (2.31)	1.41 (1.98)	1.32 (1.71)	1.33 (1.50)	79

^a Cell represents response pattern that is consistent with the hypothesised lower social desirability tendency among Web respondents

of differences in answers, which reveal the lack of pattern types in the direction of more extreme answers in the web mode.

To further examine the differences in extreme responding and its incidence on items susceptible to social desirability bias, we limit the analysis to 79 items with at least three answer categories. We fitted logistic regressions with two binary dependent variables indicating whether a respondent selected the lower or upper extreme response for each of these items. As with other models, gender, age, and higher education were included as the control variables.

The findings are largely consistent with the observed patterns of differences in answers. Significant odds ratios ($\alpha = 0.01$) of selecting lower extreme answers were found on nineteen of 79 items (24 %) in a comparison between web and CATI, and on eighteen items (23 %) between web and CAPI. Interviewer-administered modes exhibit a higher likelihood of extreme answers in all but one of these cases. The exception is the item about questionnaire length, where web respondents were more likely than CATI respondents to give an extreme answer that the questionnaire was “definitely not” too long (12.02E).

CATI and CAPI respondents were also found to have significantly higher odds of selecting upper extreme answers than web respondents on fourteen items (18 %) in the web-CATI comparison and on thirteen items (17 %) in the web-CAPI comparison. For none of these items are the odds of selecting an upper extreme answer significantly higher on the web.

Table 5 provides insight into the relationship between socially desirable and extreme responding. It summarises average and median² odds ratios across the items, with higher

²Exceptionally high odds ratios in a few items relatively strongly affect the overall means of odds ratios. We therefore prefer to use median values for interpretation.

odds indicating higher odds of selecting an extreme response in one of the interviewer-administered modes. While there is a general pattern of lower extreme responding among web respondents across all items, the observed effects are the highest for extreme responses identified as more socially desirable. In these cases (marked with a note in Table 5), a lower extremity of answers by web respondents syndicate with their lower tendency toward socially desirable responding. On the other hand, a lower extremity of the web mode largely disappears for extreme answer categories opposite to the desirable end of the scale.

These findings are in line with the hypothesis that respondents more likely select extreme answers to questions prone to social desirability, where the effect appears to be in synergy with a generally lower tendency of extreme responding in web surveys.

5 Conclusions

The presented study elaborated the differences in estimates between three survey modes caused by mode effects with a focus on socially desirable responding. The comparisons of web mode against telephone and face-to-face interviewing remains a topical research area, particularly due to an ever-increasing interest in the transition from traditional to online surveys.

The results of the empirical study agree with a majority of the existing research observing lower social desirability tendencies in web surveys compared to interviewer-administered modes. Significant differences were observed for 37 % of items rated as potentially susceptible to social desirability through comparisons between web and CATI and for 51 % of such items through comparisons between web and CAPI. An analysis of 49 items for which we were able to identify the likely desirable response categories revealed lower social desirability tendencies of web respondents where significant differences were observed between modes. Furthermore, while web respondents were generally less likely to choose extreme scale values compared to CATI and CAPI respondents, the differences were particularly pronounced on items susceptible to social desirability. This strongly suggests that social desirability bias might play a central role in differences between modes. Although further research is needed to confirm such indications, social desirability has been the most consistently observed consequence of mode effects by previously-conducted mode comparisons studies.

The performed analysis across a large number of variables enabled insights into the prevalence of differences between modes caused by socially desirable responding. The use of different indicators of response patterns (differences in means, distributions, and extreme answers) offered additional added value by demonstrating how the (non)detection of mode differences can strongly depend on the type of the estimated parameter. On the other hand, such analysis comes at the cost of lower attention paid to specific factors of individual items or scales. Therefore, it would be worth performing further item-level analyses in the future, particularly to compare the reliability and validity of the measurement between modes.

It is important to highlight some limitations of the presented study. While using an online access panel for sampling assisted with assuring the comparability between modes by reducing the non-coverage bias and reaching a broader demographic structure of re-

spondents, the yielded sample may be specific compared to the general population in other ways. Most importantly, the panellists are used to regularly participate in web surveys, which may have affected the obtained results to some degree. For example, the self-administration of the questionnaire and the use of computer technology may present a smaller burden for panellists. The higher response rate achieved by the web compared to the interviewer-administered modes also indicates that the study participants may favour web surveying over the interviewer-administered modes.

As with other studies on social desirability, the decision on whether or not an item is prone to social desirability is a largely subjective one, and the underlying assumption is that a higher reporting of less desirable answers is more accurate. The expert evaluations proved to be helpful in alleviating the former issue, but it would be beneficial to use a more granular measurement of susceptibility to social desirability rather than a simple binary one. This is especially true since most of the items were rated as susceptible, but no significant difference between modes was found for many of them. A more sensitive question evaluation methodology would allow further investigation of these observations. In addition, we cannot rule out the possibility of potentially different results for questions for which we did not identify the likely direction of socially desirable responses (mostly questions regarding values), although there are little theoretical and empirical grounds for expecting inconsistent findings.

The results thus strengthen the advantageous position of the web mode over telephone and face-to-face surveys on sensitive and socially desirable topics. While this is clearly beneficial in terms of accuracy, it may prove problematic with mode changes in longitudinal studies, mixed-mode designs, and other surveys in which the comparability of data between modes is of paramount importance. There is currently little action survey practitioners may take regarding this issue apart from weighing the importance of various data quality dimensions and costs when considering the transition to web surveys. Pilot evaluations are an essential decision-making tool for large survey projects to assess the magnitude of potential differences caused by the differences in socially desirable responding and other mode effects.

Finally, the lower incidence of socially desirable responding should not lead one to regard this mode as being immune to the problem. As the theoretical elaboration in this paper pointed out, the survey mode itself merely presents a foundation for the survey implementation. Mode effects emerge under different circumstances, where complex relations and interactions between mode characteristics are accompanied by other specific survey-related and respondent-related factors. Highly versatile modes, such as web surveys, offer numerous possibilities, but also involve the danger of damaging data quality by careless use of available features, such as overexcitement with media-related and interactive capabilities. Survey practitioners should therefore carefully weigh the benefits of exploiting specific characteristics of the web mode, especially when using features with currently unclear methodological implications.

Acknowledgements

The authors acknowledge that the project “Integration of mobile devices into survey research in social sciences: Development of a comprehensive methodological approach”

(J5-8233) was financially supported by the Slovenian Research Agency.

References

- [1] Aquilino, W. S. and Lo Sciuto, L. A. (1990): Effects of interview mode on self-reported drug use. *Public Opinion Quarterly*, **54**, 362–393. <https://doi.org/10.1086/269212>
- [2] Aquilino, W. S., Wright, D. L. and Supple, A. J. (2000): Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse*, **35**, 845–867. <https://doi.org/PMID:10847214>
- [3] Beebe, T. J., Harrison, P. A., Mcrae, J. A., Jr., Anderson, R. E. and Fulkerson, J. A. (1998): An evaluation of computer-assisted self-interviews in a school setting. *Public Opinion Quarterly*, **62**, 623–632. <https://doi.org/10.1086/297863>
- [4] Benjamini, Y. and Yekutieli, D. (2001): The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, **29**, 1165–1188.
- [5] Bennink, M., Moors, G. and Gelissen, J. (2013): Exploring response differences between face-to-face and web surveys: A qualitative comparative analysis of the dutch European values survey 2008. *Field Methods*, **25**, 319–338. <https://doi.org/10.1177/1525822X12472875>
- [6] Berzelak, N. (2014): Mode effects in web survey (Doctoral dissertation). University of Ljubljana, Ljubljana. Retrieved from dk.fdv.uni-lj.si/doktorska_dela/pdfs/dr_berzelak-jernej.pdf
- [7] Bradburn, N. M., Sudman, S., Blair, E. and Stocking, C. (1978): Question threat and response bias. *Public Opinion Quarterly*, **42**, 221–234. <https://doi.org/10.1086/268444>
- [8] Cannell, C. F., Miller, P. V. and Oksenberg, L. (1981): Research on interviewing techniques. In S. Leinhardt (Ed.): *Sociological Methodology*, 389–437. San Francisco, CA: Jossey-Bass.
- [9] Castelli, L. and Tomelleri, S. (2008): Contextual effects on prejudiced attitudes: When the presence of others leads to more egalitarian responses. *Journal of Experimental Social Psychology*, **44**, 679–686. <https://doi.org/10.1016/j.jesp.2007.04.006>
- [10] Chang, L. and Krosnick, J. A. (2009): National surveys via RDD telephone interviewing versus the internet. *Public Opinion Quarterly*, **73**, 641–678. <https://doi.org/10.1093/poq/nfp075>
- [11] Couper, M. P. (2011): The future of modes of data collection. *Public Opinion Quarterly*, **75**, 889–908. <https://doi.org/10.1093/poq/nfr046>

- [12] de Leeuw, E. D. (1992): *Data quality in mail, telephone and face-to-face surveys*. Amsterdam, NL: TT-Publikaties. Retrieved from <https://eric.ed.gov/?id=ED374136>
- [13] de Leeuw, E. D. (2005): To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, **21**, 233–255.
- [14] de Leeuw, E. D. (2008): The effects of computer-assisted interviewing on data quality (Unpublished paper). Retrieved from <http://igitur-archive.library.uu.nl/fss/2010-0601-200229/EdL-effect 2002.pdf>
- [15] de Leeuw, E. D., Hox, J. and Scherpenzeel, A. (2010): Mode effect or question wording? In Proceedings of the Survey Research Methods Section (pp. 5959–5967). Alexandria, VA, US: American Statistical Association. Retrieved from <http://www.amstat.org/sections/srms/proceedings/y2010/Files/400117.pdf>
- [16] Deming, W. E. (1944): On errors in surveys. *American Sociological Review*, **9**, 359–369. <https://doi.org/10.2307/2085979>
- [17] Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J. and Messer, B. L. (2009): Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, **38**, 1–18. <https://doi.org/10.1016/j.ssresearch.2008.03.007>
- [18] Dodou, D. and de Winter, J. (2014): Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, **36**, 487–495. <https://doi.org/10.1016/j.chb.2014.04.005>
- [19] Fang, J., Wen, C. and Prybutok, V. (2014): An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. *Computers in Human Behavior*, **30**, 335–343. <https://doi.org/10.1016/j.chb.2013.09.019>
- [20] Hochstim, J. R. (1967): A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association*, **62**, 976–989. <https://doi.org/10.2307/2283686>
- [21] Holtgraves, T. (2004): Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, **30**, 161–172. <https://doi.org/10.1177/0146167203259930>
- [22] Jäckle, A., Roberts, C. and Lynn, P. (2006): Telephone versus face-to-face interviewing: Mode effects on data quality and likely causes (ISER Working Paper No. 2006–41). Essex, GB: Institute for Social & Economic Research.
- [23] Krysan, M. and Couper, M. P. (2003): Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly*, **66**, 364–383. <https://doi.org/10.2307/1519835>

- [24] Lee, H., Kim, S., Couper, M. P. and Woo, Y. (2018): Experimental comparison of PC web, smartphone web, and telephone surveys in the new technology era. *Social Science Computer Review*, advance online publication. <https://doi.org/10.1177/0894439318756867>
- [25] Lozar Manfreda, K., Berzelak, N., Vehovar, V., Bosnjak, M. and Haas, I. (2008): Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, **50**, 79–104. <https://doi.org/10.1177/147078530805000107>
- [26] Lozar Manfreda, K. and Vehovar, V. (2002): Mode effects in web surveys. In *Proceedings of the Survey Research Methods Section*. Alexandria, VA, US: American Statistical Association. Retrieved from <http://www.amstat.org/sections/srms/Proceedings/y2002/Files/JSM2002-000972.pdf>
- [27] Milton, A. C., Ellis, L. A., Davenport, T. A., Burns, J. M. and Hickie, I. B. (2017): Comparison of self-reported telephone interviewing and web-based survey responses: Findings from the second Australian Young and Well national survey. *JMIR Mental Health*, **4**, e37. <https://doi.org/10.2196/mental.8222>
- [28] Näher, A.-F. and Krumpal, I. (2012): Asking sensitive questions: the impact of forgiving wording and question context on social desirability bias. *Quality & Quantity*, **46**, 1601–1616. <https://doi.org/10.1007/s11135-011-9469-2>
- [29] Nederhof, A. J. (1985): Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, **15**, 263–280. <https://doi.org/10.1002/ejsp.2420150303>
- [30] Paulhus, D. L. (2002): The evolution of a construct. In H. I. Braun, D. N. Jackson and D. E. Wiley (Eds.): *The role of constructs in psychological and educational measurement*, 49–69. Mahwah, NJ: Lawrence Erlbaum Associates.
- [31] Rothman, K. J. (1990): No adjustments are needed for multiple comparisons: *Epidemiology*, **1**, 43–46. <https://doi.org/10.1097/00001648-199001000-00010>
- [32] Schwarz, N. and Oyserman, D. (2001): Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, **22**, 127–160. <https://doi.org/10.1177/109821400102200202>
- [33] Smither, J. W., Walker, A. G. and Yap, M. K. T. (2004): An examination of the equivalence of web-based versus paper-and-pencil upward feedback ratings: Rater- and ratee-level analyses. *Educational and Psychological Measurement*, **64**, 40–61. <https://doi.org/10.1177/0013164403258429>
- [34] Tourangeau, R., Conrad, F. G. and Couper, M. (2013): *The science of web surveys*. Oxford, GB: Oxford University Press.

- [35] Tourangeau, R. and Rasinski, K. A. (1988): Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, **103**, 299–314.
- [36] Tourangeau, R., Rips, L. J. and Rasinski, K. A. (2000): *The psychology of survey response*. Cambridge, GB: Cambridge University Press.
- [37] Tourangeau, R. and Yan, T. (2007): Sensitive questions in surveys. *Psychological Bulletin*, **133**, 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>
- [38] Turner, H. M., III and Bernard, R. M. (2006): Calculating and synthesizing effect sizes. *Contemporary Issues in Communication Science and Disorders*, **33**, 42–55.
- [39] Ye, C., Fulton, J. and Tourangeau, R. (2011): More positive or more extreme? A meta-analysis of mode differences in response choice. *Public Opinion Quarterly*, **75**, 349–365. <https://doi.org/10.1093/poq/nfr009>
- [40] Zhang, X., Kuchinke, L., Woud, M. L., Velten, J. and Margraf, J. (2017): Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behavior*, **71**, 172–180. <https://doi.org/10.1016/j.chb.2017.02.006>