

# Effects of Changing Modes on Item Nonresponse in Panel Surveys

*Oliver Lipps<sup>1</sup>, Marieke Voorpostel<sup>1</sup>, and Gian-Andrea Monsch<sup>2</sup>*

To investigate the effect of a change from the telephone to the web mode on item nonresponse in panel surveys, we use experimental data from a two-wave panel survey. The treatment group changed from the telephone to the web mode after the first wave, while the control group continued in the telephone mode. We find that when changing to the web, “don’t know” answers increase moderately from a low level, while item refusal increases substantially from a very low level. This is the case for all person groups, although socio-demographic characteristics have some additional effects on giving a don’t know or a refusal when changing mode.

*Key words:* telephone; web; don’t know; item refusal.

## 1. Introduction

Several household panel surveys have explored the feasibility of changing from an (expensive) interviewer-based survey mode to the (cheaper) web mode (Voorpostel et al. 2021). One concern with such a change is that the web mode produces more item nonresponse (INR) than interviewer-based modes such as telephone surveys (Bowling 2005; De Leeuw 2005; Guzy and Leitgöb 2015). Reasons for this are in particular the absence of an interviewer to motivate the respondent and to provide help to find a substantive answer. Differences between interviewer-based modes and the web on INR has been mostly researched for cross-sectional surveys (e.g., Lipps and Monsch 2022). The problem in a cross-sectional design is that it is difficult to separate effects from selection into a survey mode and measurement issues (Vannieuwenhuyze and Loosveldt 2013).

In this article, we overcome this shortcoming by using experimental data from a two-wave panel, in which the treatment group changes from the telephone to the web survey mode while the control group keeps the telephone mode. This allows us to investigate individual changes for each variable using a difference-in-difference design. Our research question is whether changing the mode leads to more or less “don’t know” (DK) or item refusal (REF) in the second wave and whether our findings differ across different socio-demographic groups.

<sup>1</sup>Foundation for Research in Social Sciences (FORS), University of Lausanne, Bâtiment Géopolis, CH – 1015 Lausanne, Switzerland. Emails: [oliver.lipps@fors.unil.ch](mailto:oliver.lipps@fors.unil.ch) and [marieke.voorpostel@fors.unil.ch](mailto:marieke.voorpostel@fors.unil.ch)

<sup>2</sup>Swiss Federal Statistical Office Espace de l’Europe 10, CH - 2010 Neuchâtel, Switzerland Email: [gianandrea.monsch@bfs.admin.ch](mailto:gianandrea.monsch@bfs.admin.ch)

**Acknowledgments:** This study has used data collected by the Swiss Household Panel, which is based at the Swiss Centre of Expertise in the Social Sciences (FORS) and financed by the Swiss National Science Foundation.

### 1.1. What Affects Item (Non-)responses?

The literature mentions four main factors that affect INR: the respondent, the question (Lipps and Monsch 2022), the interviewer (Silber et al. 2021), and the survey mode. In this research, we focus on effects from the respondent and in particular the survey mode.

As for “respondent characteristics”, older, very young, lower educated, and those with a higher response burden people have more problems with question comprehension and have a higher likelihood of giving an INR (Fricker et al. 2005). Holbrook et al. (2006) suggest that the way survey questions are constructed relies on cultural conceptions of the dominant cultural group such that respondents from language and cultural minorities have more difficulties to understand survey questions and to give adequate answers. Finally, question comprehension and motivation of respondents is related to INR (Kleiner et al. 2015) because motivation decreases satisficing and increases the cognitive effort of the interviewee when responding to survey questions (Krosnick 1991; Shoemaker et al. 2002).

While respondent characteristics are well established factors in the literature on INR, research on *mode* effects, most importantly, the presence or absence of an interviewer, is relatively new. The presence of an interviewer generally lowers INR because the interviewer motivates respondents to complete the task and can provide more explanation when a question is not understood (De Leeuw and Hox 2018; Groves et al. 2011; Silber et al. 2021). See Gooch and Vavreck (2019) for an exception.

However, an interviewer may also exert social pressure to give a substantively valid answer even if the respondent does not want or cannot answer a question (Chang and Krosnick 2010). INR often remains “hidden” in interviewer-based surveys as respondents, instead of not answering a question, may provide an answer that is socially accepted, for example a mid-scale response (Lipps and Monsch 2022; Sturgis et al. 2014). The mode of interview also affects responses through variation in other characteristics, such as the pace of the interview, presentation (visual or auditory), and the timing of the interview (Christian et al. 2007). Although more time and flexibility for web respondents and the possibility to reread the question may decrease INR, the evidence points to the opposite (Fricker et al. 2005). While an interviewer may classify a response as DK or REF without explicitly proposing this option to the respondent, in web surveys, the DK response category is often offered to the respondent to account for the fact that DK may be a valid response. Skipped questions are then interpreted as refusals (REF). Presenting DK and/or REF answers as visually explicit options is an important reason for obtaining a high rate of INR in the web mode.

### 1.2. Distinguishing DK and REF in Telephone and Web Surveys and by Respondent Characteristics

DK and REF are different types of INR, which result from different “disruptions of the cognitive response process” (Silber et al. 2021, 1; Beatty and Herrmann 2002). In the cognitive response process (Tourangeau and Rasinski 1988), respondents need to “understand” the question, retrieve adequate *information* to answer the question, evaluate whether they “are willing” to give this answer, and finally find the right category on the response scale. Beatty and Herrmann (2002) distinguish different states of the retrieval step, ranging from answers being easily available and retrievable to not being known. DK

answers will be given when a question is not well understood or a valid answer cannot be provided (Chang and Krosnick 2010; Young 2012), while REF will occur if the respondent has understood the question but does not want to give a valid answer. Often, REF results from social desirability if the respondent considers the valid answer to be inadequate (Krumpal 2013). Studies using unfolding brackets when answers about income are not given provide interesting information about the difference between DK and REF (Juster and Smith 1997). After an INR, bracket questions start by asking if the income is greater than  $x$  (e.g., EUR 5,000), then – once a substantive answer is given (e.g., no) – by asking if the income is greater than  $y$  (e.g., EUR 3,000), and so on. After all brackets (usually three or four) are answered, the ultimate income range is relatively small. Juster and Smith (1997) find that while 80% of initial DK respondents completed all brackets, this was the case for only 40% of initial REF respondents. Such findings show that DK answers may be truly valid but more substantive values can be retrieved if help or motivation is provided.

Self-administered modes generally provide more INR in particular to questions that are difficult or do not relate to respondents' life circumstances (Greene et al. 2008). More specifically, Lipps and Monsch (2022) found more DK answers in the web mode if calculation is needed, suggesting that interviewers can motivate and help respondents to answer such questions.

There is not much evidence about how socio-demographic characteristics interact with survey modes. Lipps and Monsch (2022) find that older, better educated, and male respondents give fewer DK responses in web surveys than in telephone surveys, while the contrary holds for foreigners from countries that speak a different language than the survey language. REFs in telephone surveys are more common among foreigners, and the more educated. Younger people, the more educated and men refuse answering a question less often in web surveys than in telephone surveys.

Given these still inconclusive findings, apart from an increase in REF answers in web surveys, we do not have clear-cut hypotheses about possible mode effects for different socio-demographic person groups. However, we expect more DK after a switch from telephone to web, especially from less educated respondents or those less familiar with the question topic.

## 2. Data and Method

### 2.1. Data

We use data from a two-wave pilot of the Swiss Household Panel (SHP) comparing telephone to web. The SHP is a longitudinal study that interviews randomly selected households and their members on an annual basis since 1999, predominantly by telephone (Tillmann et al. 2021). In preparation of the third refreshment sample, which was launched in 2020, a mode experiment conducted in 2017–2018 compared the standard telephone-based recruitment and fieldwork strategy with two web alternatives. This article is based on only a part of the data collected in the experiment. For the complete mode experiment, see Voorpostel et al. (2020).

The SHP conducts interviews with the household reference person (HRP) to collect information on the household and its members (household questionnaire) as well as with

all household members to complete an individual questionnaire. We include all HRP who participated in a telephone interview in wave 1 of the pilot. Part of this sample completed the household questionnaire in the second wave again by telephone (tel-to-tel group) and part of the sample completed a web survey (tel-to-web group).

Figure 1 gives an overview of the modes of the tel-to-tel group and the tel-to-web group in both waves.

While Lipps and Monsch (2022) used data from the individual questionnaire of wave 1 of this experiment, we use data from the household questionnaire answered by the HRP in wave 1 and 2 in this article, because only for the household questionnaire did a mode-switch take place. Unlike Lipps and Monsch (2022), we are not able to distinguish question characteristics such as length or complexity of the questions, as the questions included in the household questionnaire were mostly factual questions. The sample for the study was a simple random sample of individuals stratified by region, drawn from a sampling frame based on population registers maintained by the Swiss Federal Statistical Office. The households of the sampled individuals were randomly assigned to one of the treatment groups. The sampled individual was approached first as an HRP, although in both waves and treatment groups, households were free to select an alternative HRP than the one initially approached. Although face-to-face and web were offered as alternatives if no telephone number was available and to initial refusals, our study only includes HRPs who participated by telephone in the first wave.

Response rates, RRI (AAPOR 2016) in the first wave on the household level were 53% for the tel-to-tel group ( $n = 1'214$  responding households) and (a not significantly different on the 5% level) 52% for the tel-to-web group ( $n = 342$  responding households). The reason that the sample of the tel-to-tel group is much larger is that telephone on the household level in both waves was used in the control group, but also in one of the experimental groups. These two groups were combined in this analysis. All households that completed at least the grid in the first wave and that had not left the study were re-approached for the second wave. Response rates on the household level in wave 2 were 77% for the tel-to-tel group (935 households), and (a significantly different) 74% for the tel-to-web group (253 households).

We analyse INR by HRPs who participated in both waves and who answered in the assigned mode ( $N = 967$  households). In addition, we only keep households with the same HRP in both waves ( $N = 889$  households). These comprise 699 households in the tel-to-tel group and 190 households in the tel-to-web group. All analyses are done using Stata 16 SE.

## 2.2. Measures

We analyze all questions asked in the household questionnaire. The questions are typical for large household panel surveys and comprise questions in the domains of the division of





	Group 1 (telephone-to-telephone)	Group 2 (telephone-to-web)
Wave 1		
Wave 2		

Fig. 1. Research design for the two-wave pilot study of the SHP\_IV (adapted from Voorpostel et al. 2020).

unpaid work, outsourcing of household tasks, household income, material deprivation, and housing. For each variable we code whether the HRP gave a (1) substantive answer, a (2) don't know (DK), or a (3) item refusal (REF). Then, we calculate for each HRP and for each question the following transitions between the two waves:

- Substantive answer to substantive answer (subs-to-subs)
- Substantive answer to DK (subs-to-DK)
- Substantive answer to REF (subs-to-ref)
- INR (DK or REF) in the first wave (INR-to-any)

INR in the latter group is not further split out according to answers in the second wave, because there are only few INR in the first wave. We provide univariate descriptive statistics distinguished by mode for the dependent variables in [Table 1](#).

All transitions are significantly different across the two groups ( $\Pr(|T| > |t|) < 0.05$ ). In particular, changing to the web mode produces significantly more substantive answers to change to both DK and REF than when keeping the telephone mode. Considering the tel-to-tel group alone, there are fewer transitions from a substantive answer to INR (transitions 2 and 3) than starting from an INR in the first wave (transition 4). This suggests positive panel conditioning effects (e.g., [Sun et al. 2019](#)).

### 3. Multivariate Models

The main independent research variable is the transition to web versus repeating a telephone interview. The regression models described below include socio-demographic variables associated with survey participation and panel attrition ([Roberts and Vandenplas 2017](#)). The goal is to (partially) control for differences in the sample composition between the two groups. Except for education, all socio-demographic variables are included in the sampling frame and do not have missing values. First, we included whether the household has a registered landline (1 = yes, 0 = no (reference category)), gender (1 = male, 0 = female (reference category)), old (1 = over 54 years, 0 = 54 years or younger (reference category)), foreigner (1 = foreign nationality, 0 = Swiss nationality (reference category)), education (1 = tertiary level, 0 = lower than tertiary level (reference category)), and finally the number of question asked. Our substantive interest is the different effect of these variables on INR in the tel-to-tel and the tel-to-web group. In [Table 2](#), we show descriptive statistics of the independent variables by survey mode in the second wave.

Although there are only small insignificant differences (on the 5% level) in sample composition by mode, with one exception. The exception concerns the dummy variable

*Table 1. Distribution of dependent variables (Transitions) by mode in the second wave.*

Transition	Tel.-to-tel.		Tel.-to-web	
	mean	s.e.	mean	s.e.
1) Substantive answer to substantive answer	.986	.0010	.969	.0042
2) Substantive answer to DK	.004	.0004	.011	.0016
3) Substantive answer to REF	.001	.0005	.014	.0032
4) Item non-response in the first wave	.009	.0008	.006	.0010

Note: N(tel- to-tel.) = 699 households, N(tel.-to-web) = 190 households.

Table 2. Distribution of independent variables by mode in second wave.

Variable	Tel-to-tel.		Tel.-to-web	
	mean	s.e.	mean	s.e.
Registered landline at home	.73	.02	.71	.03
Male	.42	.02	.48	.04
Foreign nationality	.13	.01	.13	.02
Tertiary education level	.30	.02	.31	.03
Old (> 54 years)*	.52	.02	.39	.04
Number of questions	45.97	.20	45.87	.40

Note: N(tel.-to-tel.) = 699 households, N(tel.-to-web) = 190 households.

\*: mean(tel.-to-tel.) statistically (5%) different from mean (tel.-to-web).

indicating whether the HRP is older than 54: the tel-to-tel group is more likely to include these older respondents in wave 2 compared with the tel-to-web group.

Like [Lipps and Monsch \(2022\)](#), to analyze the effects of respondent characteristics in the tel-to-tel and the tel-to-web conditions, we transform the data to a “cross-classified” structure (see [Fielding and Goldstein 2006](#)). Thus, we account for the crossed structure of the two levels respondents and questions: questions are clustered in respondents and respondents are clustered in questions. This structure allows modeling respondent and question variance in the same two-level model and to disentangle respondent and question effects on INR. Cross-classified models can be estimated using Bayesian Markov Chain Monte Carlo (MCMC) methods (see [Browne 2019](#)). To compare nested models, we use the Bayesian Deviance Information Criterion (DIC), which is an MCMC penalized goodness of fit measure, equivalent to the Akaike Information Criterion (AIC) used in maximum likelihood estimation. A smaller DIC means a better model fit. We use the `runmlwin` ([Leckie and Charlton 2013](#)) command in Stata version 16, which estimates the models by using the MLwiN software.

After dropping variables that were answered by fewer than 30 respondents, the data set consists of 40,766 respondent-question combinations (32,066 telephone and 8,700 web), including 92 variables and between 32 and 889 respondents per variable. Since each of the measured variables has potentially four nominal categories (subs-to-subs, subs-to-DK, subs-to-REF, INR-to-any), we would have to use a multinomial logit model (using, e.g., subs-to-subs as reference category). Because the cross-classified model with a multinomial distribution does not converge, we resort to using three binary models, modeling subs-to-DK, subs-to-REF, and INR-to-any (versus subs-to-subs) separately. We start with the base models (only estimating variance components), then add the socio-demographic variables, and finally add the interaction of the socio-demographic variables with tel-to-web (as opposed to tel-to-tel) to test whether the mode effects vary by the socio-demographic variables. [Table 3](#) presents the results.

We interpret for each of the dependent variables (subs-to-DK, subs-to-REF, INR-to-any) only the best-fitting model, that is, the model with the lowest DIC statistics. This is for subs-to-DK and for subs-to-REF the model which includes the interaction variables (c), and for INR-to-any the model which includes the socio-demographic variables only (b).

Web in the second wave increases the likelihood to provide a REF or a DK rather than again giving a substantive answer compared to telephone. For both telephone and web, we find that

Table 3. Beta-coefficients from cross-classified logit models.

Variables	Sub-to-subs (ref.)	Subs-to-DK	Subs-to-REF	INR-to-any
(a) base models (variance components)				
Constant		-6.326**	-5.875**	-6.166**
Variance (Respondent level)		1 (fixed)	1 (fixed)	1 (fixed)
Variance (Question level)		2.641**	1.065**	3.730**
DIC		2250	1898	2920
(b) main effects				
Web		1.165**	2.636**	-0.253
Registered landline		0.0728	0.0781	0.225
Male		-0.849**	0.436*	-0.524**
Foreign Nationality (Ref.: Swiss)		0.126	0.603**	0.601**
Tertiary education level		-0.615**	-0.637**	0.0573
Old (Ref.: < 55 years old)		0.0914	0.368	0.694**
Number of questions answered		-0.0329**	-0.0544**	-0.00750
Constant		-4.867**	-5.093**	-6.213**
Variance (Respondent level)		1 (fixed)	1 (fixed)	1 (fixed)
Variance (Question level)		2.828**	1.280**	3.950**
DIC		2149	1642	2860
(c) interaction effects				
Web		2.911*	2.421**	-0.903
Registered landline		-0.184	1.190	0.229
Male		-1.011**	0.489	-0.515**
Foreign Nationality (Ref.: Swiss)		0.0427	0.0570	0.625**
Tertiary education level		-0.619**	0.899**	-0.0162
Old (Ref.: < 55 years old)		0.761**	1.972**	0.755**
Number of questions answered		-0.0157	-0.112**	-0.00849
Web X Registered landline		0.619	-1.265	0.0565
Web X Male		0.353	-0.0449	-0.115
Web X Foreigner Nationality		0.227	0.523	-0.0537
Web X Tertiary education level		0.144	-2.228**	0.519
Web X Old		-1.660**	-1.936**	-0.297
Web X Number of questions		-0.0362	0.0727**	0.0125
Constant		-5.827**	-5.219**	-6.267**
Variance (Respondent level)		1 (fixed)	1 (fixed)	1 (fixed)
Variance (Question level)		2.907**	1.274**	3.955**
DIC		2130	1601	2866
N (respondent-question combinations)		40'280	40'215	40'393

Note: \*\*  $p < 0.01$ , \*  $p < 0.05$ . Data: Swiss Household Panel. Shaded: best-fitting models (lowest DIC).

men give fewer transitions to DK and fewer INR from the onset than women, and foreigners more INR. Better educated HRP provide less DK but more REF than lower educated and older HRP more DK, REF, and initial INR than younger HRP. The number of questions is negatively associated with REF. Note that these effects may change if only main effects models (b) are considered (e.g., the effect of high education on DK changes sign).

As for the effect when transferred from telephone to the web as compared to keeping the telephone mode (interactions), no clear pattern emerges. Higher educated and older HRP who change to the web increase their REF less than lower educated and younger HRP. Older HRP are also less likely to switch to DK. This suggests that older HRP are more likely to provide a substantive answer when switched to the web. Finally, more questions lead to a higher likelihood to refuse in the tel-to-web group than in the tel-to-tel group. The comparatively high variance on the question level shows the importance to model respondent and question variance separately, using crossed models.

#### 4. Conclusion

Based on data from the two-wave SHP IV pilot study, which incorporated a mode experiment including a control (tel-to-tel) and a treatment group (tel-to-web), we explored the impact of the mode change on INR in the household questionnaire. We find from simple descriptive statistics that compared to the telephone mode, the web survey produces more transitions from subs-to-DK (tel-to-tel: 0.4%, tel-to-web: 1.1%) and considerably more transitions from subs-to-REF (tel-to-tel: 0.1%, tel-to-web: 1.4%). This is in line with most findings from cross-sectional studies (e.g., [De Leeuw and Hox 2018](#)). While cross-sectional studies may suffer from selection effects, our study using longitudinal study clearly shows that this increase is a mode effect.

Furthermore, in addition to recent research that uses cross-sectional data ([Lipps and Monsch 2022](#)) showing that survey modes affect both DK and REF to different extents, we now show that this holds in a longitudinal context: people with different socio-demographic characteristics have a different tendency to switch to DK or REF when transferred to the web in the second wave. However, there is no clear pattern emerging.

While offering DK and REF may drive up INR, it produces DK answers that are likely to reflect not knowing the answer to the question more accurately, rather than unwillingness to answer. As an additional exploratory examination of what drives DK answers, we analyzed the median time needed for the HRP to reply to one question (minimum monthly income needed to make ends meet). Respondents who gave a substantive answer in both waves took slightly less time to answer this question in the second wave compared to the first (33 to 29 seconds in the tel-to-tel group, 34 to 30 seconds in the tel-to-web group), suggesting a learning effect. To the contrary, a DK was given much faster in the second wave compared to the time it took them to give a substantive answer in the first wave (50 to 36 seconds in the tel-to-tel group, 36 to 14 seconds in the tel-to-web group). This suggests that a DK does not result from a long search process, especially in the web mode. This points to the need to probe the respondent after giving a DK or using unfolding brackets although this may lengthen the interview duration considerably ([De Leeuw et al. 2016](#)).

Being able to isolate DK from REF may be of substantive interest for researchers analyzing the data. Reasons for more REF in web surveys may be due to less motivation and guidance but probably also the tendency to give a more polite answer to an interviewer. It may be that both DK and REF decrease once the web is an even more established and accepted survey mode, people are more familiar with it, or more motivation, interaction, or entertainment is provided (e.g., [Keusch 2020](#)). We find support for this idea by the much lower likelihood for higher educated HRPs to give a REF. Higher educated respondents may be less burdened by a web survey than less well-educated people. Contrary to our expectations, we do not find less educated respondents to switch more from a substantive answer to a DK in the web. This shows that our web version produces more DK because this category is explicitly offered and that respondents with different education levels do not selectively give or hide a DK according to the mode.

This study has some limitations. First, the sample size of the telephone-to-web group is small. Second, the rate of INR is rather small, as questions from the household questionnaire are most often factual and not very prone to INR. Third, there are probably still small



selective attrition effects such that HRP who are less internet affine in the telephone-to-web group are more likely to attrite after the first (telephone) wave. We identified an association between (older) age and attrition (Voorpostel et al. 2020) but there are most likely other characteristics that remained unobserved. A higher drop-out of older respondents in the telephone-to-web group with probably more INR rates in the second wave was likely to decrease INR differences between telephone and web in the second wave due to attrition. Finally, since we cannot observe the counterfactual, we do not know whether our findings are an effect of changing modes or an effect of the use of web. These limitations withstanding, we believe that our design is well-suited to analyze effects on INR when changing mode in a panel survey. We show that switching to the web is possible without risking a severe increase in DK or REF. Future research could use larger samples in the web mode and different (more complex, more sensitive, more non-factual) questions with higher likelihoods of DK or REF already in an interviewer-administered mode. This will provide more insights into the trade-off between more REF in the web mode and more substantive but untrue answers in interviewer-administered modes.

## 5. References

- AAPOR (The American Association for Public Opinion Research). 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, (9th edition). Illinois: AAPOR. Available at: [http://www.aapor.org/AAPOR\\_Main/media/publications/Standard-Definitions20169theditionfinal.pdf](http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf) (accessed November 2022).
- Beatty, P., and D. Herrmann. 2002. "To Answer or not to Answer: Decision Processes Related to Survey Item Non-response." In *Survey Nonresponse*, edited by R. Groves, D. Dillman, J. Eltinge, and J. Little. (pp. 71–85). New York: Wiley.
- Bowling, A. 2005. "Mode of Questionnaire Administration can have Serious Effects on Data Quality." *Journal of Public Health* 27: 281–291. DOI: <https://doi.org/10.1093/pubmed/fdi031>.
- Browne, W.J. 2019. "MCMC Estimation in MLwiN, v3.03." Centre for Multilevel Modelling, University of Bristol. Available at: <http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-03/mcmc-web.pdf> (accessed November 2, 2022)
- Chang, L., and J.A. Krosnick. 2010. "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires. An Experiment." *Public Opinion Quarterly* 74(1): 154–167. DOI: <https://doi.org/10.1093/poq/nfp090>.
- Christian, L.M., D.A. Dillman, and J.D. Smyth. 2007. "The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys." In *Advances in Telephone Survey Methodology*, edited by J.M. Lepkowski, C. Tucker, M. Brick, E.D. de Leeuw, L. Japac, P.J. Lavrakas, M.W. Link, and R.L. Sangster: 250–275. Hoboken, Wiley.
- De Leeuw, E.D. 2005. "To Mix or not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics* 21(2): 233–255. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/to-mix-or-not-to-mix-data-collection-modes-in-surveys.pdf>.
- De Leeuw, E.D., and J. Hox. 2018. "Internet Surveys as Part of a Mixed-Mode Design." In *Social and behavioral research and the Internet: Advances in applied methods and*

- research strategies*, edited by M. Das, P. Ester, and L. Kaczmirek. 45–76. New York: Routledge.
- De Leeuw, E.D., J. Hox, and A. Boeve. 2016. “Handling do-not-know Answers: Exploring new Approaches in Online and Mixed-mode Surveys.” *Social Science Computer Review* 34(1): 116–132. DOI: [10.1177/0894439315573744](https://doi.org/10.1177/0894439315573744).
- Fielding, A., and H. Goldstein. 2006. “Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review,” University of Birmingham. Available at: <http://www.bristol.ac.uk/cmm/team/hg/full-publications/2006/cross-classified-models-review-for-dfes.pdf> (accessed November 2022)
- Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. 2005. “An Experimental Comparison of Web and Telephone Surveys.” *Public Opinion Quarterly* 69(3): 370–392. DOI: <https://doi.org/10.1093/poq/nfi027>.
- Gooch, A., and L. Vavreck. 2019. “How Face-to-Face Interviews and Cognitive Skill Affect Item Non-response: A Randomized Experiment Assigning Mode of Interview.” *Political Science Research and Methods* 7(1): 143–162. DOI: <https://doi.org/10.1017/psrm.2016.20>.
- Greene, J., H. Speizer, and W. Wiitala. 2008. “Telephone and Web: Mixed-Mode Challenge.” *Health services research* 43(1): 230–248. DOI: [10.1111/j.1475-6773.2007.00747.x](https://doi.org/10.1111/j.1475-6773.2007.00747.x).
- Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2011. *Survey Methodology*. Hoboken: Wiley.
- Guzy, N., and H. Leitgöb. 2015. “Assessing Mode Effects in Online and Telephone Victimization Surveys.” *International Review of Victimology* 21(1): 101–131. DOI: <https://doi.org/10.1177/0269758014547995>.
- Juster, F.T., and J.P. Smith. 1997. “Improving the Quality of Economic Data: Lessons from the HRS and AHEAD.” *Journal of the American Statistical Association* 92(440): 1268–1278. DOI: [10.1080/01621459.1997.10473648](https://doi.org/10.1080/01621459.1997.10473648).
- Holbrook, A., Y.I. Cho, and T. Johnson. 2006. “The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties.” *Public Opinion Quarterly* 70(4): 565–595. DOI: <https://doi.org/10.1093/poq/nfl027>.
- Keusch, F. 2020. “Gamification in Web Surveys.” In *SAGE Research Methods Foundations*, edited by P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, and R.A. Williams. 1–8. London: Sage.
- Kleiner, B., O. Lipps, and E. Ferrez. 2015. “Language Ability and Motivation Among Foreigners in Survey Responding.” *Journal of Survey Statistics and Methodology* 3(3): 339–360. DOI: <https://doi.org/10.1093/jssam/smv015>.
- Krosnick, J.A. 1991. “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys.” *Applied Cognitive Psychology* 5: 231–236. DOI: <https://doi.org/10.1002/acp.2350050305>.
- Krumpal, I. 2013. “Determinants of Social Desirability Bias in Sensitive Surveys: a Literature Review.” *Quality & Quantity* 47(4): 2025–2047. DOI: <https://doi.org/10.1007/s11135-011-9640-9>.
- Leckie, G., and C. Charlton. 2013. “runmlwin – A Program to Run the MLwiN Multilevel Modelling Software from Within Stata.” *Journal of Statistical Software* 52: 1–40. DOI: <https://doi.org/10.18637/jss.v052.i11>.

- Lipps, O., and G.-A. Monsch. 2022. "Effects of Question Characteristics on Item Non-Response in Telephone and Web Survey Modes." *Field Methods*. DOI: <https://doi.org/10.1177/1525822X221115838>.
- Roberts, C., and C. Vandenplas. 2017. "Estimating Components of Mean Squared Error to Evaluate the Benefits of Mixing Data Collection Modes." *Journal of Official Statistics* 33(2): 303–334. DOI: <http://dx.doi.org/10.1515/JOS-2017-0016>.
- Shoemaker, P.J., M. Eichholz, and E.A. Skewes. 2002. "Item Nonresponse: Distinguishing Between Don't know and Refuse." *International Journal of Public Opinion Research* 14(2): 193–201. DOI: <https://doi.org/10.1093/ijpor/14.2.193>.
- Silber, H., J. Roßmann, T. Gummer, S. Zins, and K.W. Weyandt. 2021. "The Effects of Question, Respondent and Interviewer Characteristics on Two Types of Item Nonresponse." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. DOI: <https://doi.org/10.1111/rssa.12703>.
- Sun, H., R. Tourangeau, and S. Presser. 2019. "Panel Effects: Do the Reports of Panel Respondents Get Better or Worse over Time?" *Journal of Survey Statistics and Methodology* 7(4): 572–588. DOI: <https://doi.org/10.1093/jssam/smy021>.
- Sturgis, P., C. Roberts, and P. Smith. 2014. "Middle Alternatives Revisited: How the Neither/nor Response Acts as a Way of Saying "I don't know." *Sociological Methods & Research* 43(1): 15–38. DOI: <https://doi.org/10.1177/0049124112452527>.
- Tillmann, R., M. Voorpostel, E. Antal, N. Dasoki, H. Klaas, U. Kuhn, F. Lebert, G.-A. Monsch, and V.-A. Ryser. 2021. "The Swiss Household Panel (SHP)." *Jahrbücher für Nationalökonomie und Statistik*. DOI: <https://doi.org/10.1515/jbnst-2021-0039>.
- Tourangeau, R., and K.A. Rasinski. 1988. "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Psychological bulletin* 103(3): 299–314. DOI: <https://doi.org/10.1037/0033-2909.103.3.299>.
- Vannieuwenhuyze, J.T., and G. Loosveldt. 2013. "Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects." *Sociological Methods & Research* 42(1): 82–104. DOI: <https://doi.org/10.1177/0049124112464868>.
- Voorpostel, M., U. Kuhn, R. Tillmann, G.-A. Monsch, E. Antal, V.-A. Ryser, F. Lebert, H.S. Klaas, and N. Dasoki. 2020. *Introducing Web in a Refreshment Sample of the Swiss Household Panel: Main Findings From a Pilot Study*. FORS Working Paper Series, paper 2020-2. Lausanne: FORS. Available at: [https://serval.unil.ch/resource/serval:BIB\\_7C300C68F2D1.P001/REF](https://serval.unil.ch/resource/serval:BIB_7C300C68F2D1.P001/REF) (accessed November 2022)
- Voorpostel, M., O. Lipps, and C. Roberts. 2021. "Mixing Modes in Household Panel Surveys: Recent Developments and new Findings." In *Advances in longitudinal survey methodology*, edited by P. Lynn. 204–226. Hoboken: Wiley. DOI: <https://doi.org/10.1002/9781119376965.ch9>.
- Young, R. 2012. *Don't Know Responses in Survey Research*. Unpublished dissertation. Department of Sociology: Pennsylvania State University. Available at: [https://etda.libraries.psu.edu/files/final\\_submissions/7219](https://etda.libraries.psu.edu/files/final_submissions/7219) (accessed November 2022)

Received July 2021

Revised February 2022

Accepted November 2022