# Questionnaire design decisions when transitioning from an interviewer-administered to a self-administered online mode (Version 1.0)

Nießen, Désirée; Hadler, Patricia; Neuert, Cornelia

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

**Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:**
GESIS - Leibniz-Institut für Sozialwissenschaften

**Empfohlene Zitierung / Suggested Citation:**
Nießen, D., Hadler, P., & Neuert, C. (2023). *Questionnaire design decisions when transitioning from an interviewer-administered to a self-administered online mode (Version 1.0).* (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/gesis-sg_en_046

gesis Leibniz-Institut für Sozialwissenschaften

# Questionnaire Design Decisions When Transitioning from an Interviewer-Administered to a Self-Administered Online Mode

*Désirée Nießen, Patricia Hadler & Cornelia Neuert*

July 2023, Version 1.0

## Abstract

Large-scale surveys are increasingly moving from (face-to-face or telephone-based) interviewer-administered to self-administered online modes. To ensure high measurement quality and maximum comparability and equivalence between the source questionnaire and its adaptation——across modes as well as across survey waves——various aspects of question design must be considered and several decisions need to be made concerning question presentation and wording. This survey guideline summarizes good practices on how to transition questionnaires from interviewer- to self-administered web surveys and gives recommendations and examples for major adaptation issues as well as general questionnaire design elements relevant to web surveys. In this context, we focus on the switch from an interviewer-based to an online mode (although mixed-mode designs are also conceivable).

## Citation

Nießen, D., Hadler, P., & Neuert, C. (2023). Questionnaire design decisions when transitioning from an interviewer-administered to a self-administered online mode. Mannheim, GESIS – Leibniz Institute for the Social Sciences (GESIS – Survey Guidelines).

DOI: 10.15465/gesis-sg_en_046

# 1 Introduction

Large-scale surveys are increasingly switching to online or mixed-mode data collection due to time pressure, rising costs of solely interviewer-administered surveys, and declining response rates as well as societal changes (e.g., the ever-increasing relevance of digitization and declining willingness to participate in face-to-face surveys). Although web surveys require programming, they do not necessarily need professional programmers (Hansen et al., 2016) and are ultimately much less expensive (Breton et al., 2017). Moreover, during the Covid-19 pandemic, several major large-scale surveys decided to switch modes to ensure data collection. However, a switch in survey modes or the combination of modes can lead to so-called measurement mode effects (i.e., responses, participation rates, response rates, break-off rates, etc. differing between modes)——for example, triggered by the fact that participants interpret the question differently because it is presented in a different way (e.g., Dillman, 2017; Hox et al., 2015). Presenting response scales in different layouts or adapting them in a way that is answerable on mobile devices might lead to a shift in responses. Therefore, adaptations of a questionnaire from an interviewer-administered (e.g., face-to-face or telephone-based) to a self-administered online mode have to minimize measurement error. At the same time, they should maintain maximal equivalence with the interviewer-administered source questionnaire, to enable comparison across modes and waves.

The demand to keep the measurement error as small as possible and at the same time to guarantee the best possible comparability between modes is already the first hurdle and question to be asked when moving to web surveys. More specifically, researchers must decide whether to focus on the comparability with previous waves or whether to reduce measurement error in the adapted online mode at the expense of comparability over time. The same applies to mixed-mode designs, if researchers want to implement several modes in parallel. Here, too, they must ask themselves whether they want to realize the most error-free measurement in each mode individually (i.e., mode-specific design) or whether to optimize the comparability between modes (i.e., unified mode design) and also make adaptations in the other direction——that is, in the interviewer-administered source questionnaire (e.g., to achieve consistency of scales in both modes).[1]

This survey guideline summarizes knowledge from the literature and from a pretest project[2] on how to transition questionnaires from interviewer-administered surveys to an online mode by providing recommendations on questionnaire design and implementation (for survey guidelines on web probing and cognitive pretesting, see Behr et al., 2017 and Lenzner et al., 2015). In the first part, we address aspects of questionnaire design that need to be adapted due to the absence of an interviewer who administers the questions and provides clarifications when needed. In doing so, the best possible comparability between modes is emphasized and only changes are

---

[1] We refer interested readers to previous survey guidelines addressing web surveys (Bandilla, 2015), methodological advantages and disadvantages of mixed-device and mobile web surveys (Beuthner et al., 2019), as well as benefits and drawbacks of using multiple modes for data collection (i.e., effects on response rates, sample balance, survey costs, and measurement mode effects; Stadtmüller et al., 2021).

[2] Hadler, P., Lenzner, T., Schick, L., Neuert, C. (2022). *European Working Conditions Survey 2024: Preparation and cognitive testing of the online questionnaire* (Working Paper WPEF22035). Eurofound. https://www.eurofound.europa.eu/sites/default/files/wpef22035.pdf

Hadler, P., Lenzner, T., Schick, L., Neuert, C., Steins, P., & Behnert, J. (2022). *European Working Conditions Survey 2024. Cognitive pretest* (GESIS Project Report 2022/08). GESIS-Pretestlabor. https://doi.org/10.17173/pretest116

Nießen, D., Hadler, P., Lenzner, T., & Neuert, C. (2022). *Working conditions and sustainable work. Good practices in transitioning to an online mode in 3MC questionnaire design* (Working Paper WPEF22034). Eurofound. https://www.eurofound.europa.eu/sites/default/files/wpef22034.pdf

recommended that are mandatory and compatible with the interviewer-administered source questionnaire. In the second part, we discuss general design aspects that should be considered in web surveys.

# 2 Aspects of questionnaire design in the transition to an online mode

The main difference between interviewer- and self-administered survey modes is the presence or absence of an interviewer, which has several implications for questionnaire design. This first main section addresses issues of questionnaire design to consider when adapting an interviewer-based source questionnaire into a self-administered web survey (for a comprehensive overview, see Olson et al., 2019):

- Addressing the respondents and question wording
- Definitions and clarifications
- Substantive response options
- Non-substantive response options
- Survey length

## 2.1 Addressing the respondents and question wording

The adaptation to a web survey requires many adjustments in question wording. This is most prominently the case when respondents were previously directly addressed, or the interviewers referred to themselves directly (e.g., when the next survey topic is introduced).

### 2.1.1 Introductory pages and bridges

In computer-assisted personal interviewing (CAPI) or computer-assisted telephone interviewing (CATI) questionnaires, the interviewer generally introduces the next topic with sentences such as "I'm now going to ask you some questions about…." In online questionnaires, four adaptations must be made: First, the adaptation must replace the first person singular ("I") with a neutral wording, such as "The next questions are about…." Second, the introductory text to a set of questions should be placed either at the top of a survey page with the next question or on a separate survey page prior to the respective question. Third, it is recommended that important words, which an interviewer would emphasize, are visually highlighted, for instance, using bold font. For example, in a survey on working conditions, for respondents with multiple jobs, the words "main" or "all" should be highlighted, depending on whether respondents should base their answer only on their main job or whether they should consider all jobs (e.g., "The next questions are about […] your **main** job. By 'main' job, we mean…"; Hadler et al., 2022, p. 18). Fourth, the introductory sections of interviewer-administered question texts are often sentence fragments that read incomplete, for instance, "In your spare time, do you…," followed by the items read out by the interviewer "… visit cultural events (e.g., exhibitions, museums, theatre, the opera)?" In the adapted online questionnaire, the question text and items should be turned into full sentences: "Do you carry out the following activities in your spare time?" and "I visit cultural events (e.g., exhibitions, museums, theatre, the opera)."

### 2.1.2 Show cards and instructions

Another common example of necessary adjustments to question wording and design is when show cards are used in CAPI mode. Normally, these show cards can simply be replaced with a

corresponding single-choice question or item-by-item battery without any adjustments to the question text. In cases in which the question text explicitly refers to the show card, this part of the question must be adjusted accordingly.

Furthermore, to ensure that respondents understand the survey questions as they are intended, instructions on how to respond to a survey question must be inserted (e.g., “Please select one answer.” or “Please select all answers that apply to you.”). In all cases, the question or item text should be adapted as minimally as possible to ensure comparability between survey modes.

## 2.2 Definitions and clarifications

In surveys conducted by an interviewer, there are instructions (i.e., definitions and explanations) that the interviewer reads aloud as needed. In online questionnaires, researchers must decide whether to present definitions and clarifications to all respondents or only “on request” using techniques such as mouse rollover or clickable references.

### 2.2.1 Explanations presented with the question text

In general, clarifications should be presented to all respondents if they are (potentially) relevant to all respondents, for instance, if central concepts of the questionnaire are introduced or if (ambiguous) terms could be misinterpreted, although respondents are familiar with the term (and thus would probably not actively seek additional information). In addition, it is recommended to explain to all respondents what they should include or exclude in an answer (e.g., when asking about the number of people living in the household, whether respondents should include themselves or not) and how detailed the information should be in an open-ended answer field. By contrast, clarifications only relevant to certain groups of respondents should be presented only to these groups (see also Olson et al., 2019). Whether a question is relevant for a respondent can be determined based on preceding filter questions.

In web surveys, it is generally recommended to offer definitions and clarifications on the survey page to which they pertain (and not centrally at the beginning of a survey), because this procedure increases the likelihood of respondents reading them (Christian & Dillman, 2004) and not already having forgotten the relevant information. If additional information is presented to all respondents, the placement of this information should (in most cases) occur directly after the question stem or below the general instruction in a series of questions (e.g., Metzler et al., 2015; Redline, 2013). However, for some questions, such as long ones, it can be preferable to place additional information in front of the question because respondents are more likely to expect the research intent of these items at this point and to ignore information after the question (Redline, 2013).

Visual techniques, such as the use of italics, parentheses or bolding, should help to differentiate definitions or clarifications from the main question text (e.g., Christian & Dillman, 2004; Dillman & Christian, 2005; Olson et al., 2019; Redline et al., 2003; Tourangeau et al., 2004). However, visual features should be used consistently throughout the survey (Dillman & Smyth, 2007) because there is growing evidence that the visual design of web surveys influences data quality (Hansen et al., 2016).

### 2.2.2 Explanations presented “on request”

Interviewer-administered surveys often include supplementary information that the interviewer only presents when it is deemed necessary to improve response quality. This includes (detailed) definitions of terms or clarifications of constructs used in the question text that can be read aloud

to respondents on demand. Typically, terms that may be clear to many respondents but unclear to others should be provided "on request" if it cannot be determined in advance who needs additional information, because presenting all potentially relevant clarifications to all respondents would immensely increase the length of some surveys and cause unnecessary burden to respondents.

If additional information is made available "on request," for example, via computer mouse rollover (resulting in a pop-up definition) or clickable words/links (resulting in a new window or tab), it should be made as easily accessible as possible without requiring multiple clicks or much effort (e.g., Conrad et al., 2006; Galesic et al., 2008; Olson et al., 2019; Peytchev et al., 2006). Rollover definitions are requested more frequently than click definitions (Conrad et al., 2006)——even though definitions are accessed most often when presented directly with the question text (Peytchev et al., 2006). But this also depends on the device type because it has been shown that smartphone users consult definitions regardless of their placement more often than tablet or laptop users (Tourangeau et al., 2017). Therefore, easy access to definitions is an important issue to consider in web surveys.

We recommend to visually highlight clarifications that are provided via mouse-over or clicking on the respective word to ensure that respondents are aware that they can receive additional information. Since several studies and cognitive pretests have shown that respondents do not read instructions and explanations thoroughly if they are only displayed "on request," researchers should decide in advance what information is necessary to increase question comprehension and avoid too complex and burdensome questions (e.g., Conrad et al., 2006; Hadler et al., 2022; Lenzner et al., 2014, 2015; Schober et al., 2000). Some questions can contain both clarifications visible to all respondents and clarifications via mouse rollover.

## 2.3 Substantive response options

In the end, the answers given by respondents produce the data of a survey. In order to maximize the quality of the answers, special attention must be paid to content-related answer options when creating an online survey. These include ordinal response scales (e.g., Likert response scales with two poles and gradations in between), multiple-answer/-item (i.e., grid) questions, and open-ended numerical questions.

### 2.3.1 Ordinal response scales

One well-documented effect observable when switching survey modes refers to attitudinal questions using ordinal scales. If presented orally, such questions are more likely to produce extreme (positive) responses, whereas visually presented scales tend to elicit an increase in intermediate categories (e.g., Christian et al., 2008; Dillman & Edwards, 2016; Heerwegh & Loosveldt, 2008; but see Heerwegh, 2009). Research has attributed this to differential cognitive processing of information obtained orally and visually, as well as providing several alternative explanations for the tendency to extreme positive responses in orally administered surveys, such as primacy and recency effects, socially desirable and acquiescent responding (e.g., Bishop et al., 1988; Cernat et al., 2016; Chang & Krosnick, 2010; Olson et al., 2019). Being aware of these effects of ordinal response scales on differential response behavior depending on the mode, it is necessary to provide respondents with as much assistance and facilitations as possible to keep these mode differences as small as possible.

A general recommendation for any type of survey, whether conducted by an interviewer or self-administered, is that the direction of the response scales should be consistent throughout the entire survey. Most previous studies have found that scale direction impacts response

distribution (e.g., Terentev & Maloshonok, 2019), but that effects are small (e.g., Höhne & Lenzner, 2015) and factor structures for latent constructs are not impacted by scale direction (for an overview, see Robie et al., 2022). At the same time, Liu and Keusch (2017) have found that in web surveys, acquiescence was higher when attitudinal scales began with the "agreement" or "applies" part of the scale. Therefore, for item batteries employing the same response scale across multiple items, "a simple solution to the inflated data obtained from descending-ordered scales is to present response scales in ascending order" (Chyung et al., 2018, p. 9), that is, from "does not apply at all" to "applies completely" (for an overview of important response scale characteristics, see DeCastellarnau, 2018). Finally, scale polarity in adapted questionnaires should not differ from the source questionnaire (e.g., Dorer, 2012). In addition to the consistency of the direction of response scales, the same response scale with the same number of categories and consistent labels is recommended whenever possible (e.g., a 5-point response scale; Porst, 2011; for a survey guideline on the design of rating scales in questionnaires, see Menold & Bogner, 2015).

### 2.3.2 Multiple-answer questions and multiple-item (grid) questions

The handling of multiple-answer and multiple-item or grid questions should also be considered when moving to a web-based survey mode. In multiple answer questions, multiple items are presented in a grid or matrix in which respondents must indicate whether or not each single item applies to them by selecting "yes" or "no" as response choices. This question format (also referred to as "forced-choice" or "yes-vs.-no" format) is often used in interviewer-administered surveys (Olson et al., 2019; Smyth et al., 2006). In an attempt to lower cognitive burden in visually presented questionnaires, multiple-answer questions are often adapted to a "check-all-that-apply" format (Olson et al., 2019; Smyth et al., 2006). In a "check-all-that-apply" format, the items are listed, and respondents are asked to select all items that apply to them.

However, it has been shown that these two formats yield significantly different responses (more response categories selected and longer completion times in the "yes-vs.-no" compared to the "check-all-that-apply" format; e.g., Neuert, 2020; Nicolaas et al., 2011, 2015; Smyth et al., 2006; Thomas & Klein, 2006). Moreover, the "check-all-that-apply" format is more prone to satisficing and primacy effects (e.g., Smyth et al., 2006), and it is more difficult to interpret what a missing choice really means (i.e., whether the item was intentionally or accidentally omitted; e.g., Neuert, 2020). To ensure comparability of data, it is generally recommended to use the same format in all modes and to prefer the "yes-vs.-no" format in most cases (e.g., Nicolaas et al., 2011) because it induces deeper processing of the answer options and is more comparable across modes (e.g., Smyth et al., 2006, 2008), even though it is also more likely to elicit acquiescent responding (e.g., Callegaro et al., 2015).

Even though it can generally not be recommended, the "check-all-that-apply" format can present a possibility for the query of factual information (e.g., countries visited) or questions with lots of answer options (e.g., leisure activities, spoken languages)——especially, considering evidence from previous studies that there were no differences in cognitive effort (measured by eye tracking) between the two modes for factual questions (e.g., Neuert, 2017). Here, an additional response option (e.g., "none of the above") should be added to avoid item non-response that would otherwise be difficult to interpret (Olson et al., 2019).

### 2.3.3 Open-ended numerical questions

Special attention should also be paid to the handling of open-ended numerical answers. For a series of open-ended numerical questions, programmed feedback that calculates the sum of numerical responses is a useful aid in eliciting more valid responses——for example,

corresponding to 100%——than without feedback. Specifically, concurrent feedback (i.e., while the answers are being entered) has been found to result in shorter response times than delayed feedback (i.e., after all answers in the question block have been entered; Conrad et al., 2005).

Another type of open-ended numerical questions, which requires special attention, are those in which answers can be given in different units of measurement or formats (e.g., height, money amounts, time indications, dates) to prevent wrong answers (e.g., minutes instead of seconds, comma instead of a period for numbers, cents instead of dollars; see also Olson et al., 2019). Providing respondents with the opportunity to select the unit that fits their need is intended to facilitate the response process. This means that the question asks about the most common unit (e.g., earnings per month), but respondents are given the option to answer the question in a different response unit (e.g., earnings per week or per year).

To ensure that respondents enter the correct format, two issues must be considered: First, the instructions should be clear enough and include, for example, a highly visible exemplary template in a separate box so that respondents can easily understand and answer the questions (e.g., Christian et al., 2007; Dillman & Smyth, 2007; Olson et al., 2019). Instead of single answer fields, it is also advisable to use separate answer fields for date questions or non-integer answers (e.g., Couper et al., 2011). Second, programmed validation checks that only allow numbers in a certain interval or format (e.g., from 1 to 100, with or without decimal numbers) and error messages that alert respondents to the allowable range can help capture correct values (Olson et al., 2019). Because error messages might be perceived as frustrating, helping respondents understand what unit of measurement or format is needed in the first place will reduce the likelihood of receiving error messages at all (e.g., Christian et al., 2007; Dillman et al., 2014). In general, we recommend programming plausibility checks and providing additional explanations on the range of valid answers (e.g., “Please enter a whole number between 1 and 100.”).

## 2.4 Non-substantive response options

Next to the presentation of substantive response options, survey designers adapting a questionnaire to an online mode must consider whether to explicitly offer non-substantive response options, such as “I don’t know,” “I don’t want to answer,” “Not applicable,” or “Refusal.”

### 2.4.1 General remarks on non-substantive response options

In interviewer-administered surveys, non-substantive response options are generally not read aloud, but noted by the interviewer when prompting the respondent does not elicit a response. In web surveys, researchers must decide whether to explicitly offer one or more non-substantive response options. Doing so generally results in a higher selection of these categories (e.g., Heerwegh, 2009; Heerwegh & Loosveldt, 2008; Olson et al., 2019). That is why they are sometimes considered as an “easy way out” and a satisficing response strategy (Krosnick et al., 2002; Schmid et al., 2023). For example, in the German questionnaire of the European Values Survey (EVS) 2017/2018, which was administered either as a face-to-face survey, as a self-administered mixed-mode survey in full length or in a matrix design, the share of “I don’t know” answers was much lower in the face-to-face-mode (1.7%) than in the mixed-mode matrix (4.6%) and the mixed-mode full survey (5.8%; Wolf et al., 2021).

Because there is no simple answer to the question of whether non-substantive answers should be provided in web surveys, we recommend making this decision for each question individually. We generally recommend offering these answer options only very sparingly (also for reasons of comparability with the interviewer-administered source questionnaire) and only when a non-

substantive response option poses a valid response. For attitudinal questions where researchers truly believe that respondents do not have an opinion on the issue, an explicit “I don’t know” option can be offered (Dillman et al., 2014; Krosnick et al., 2002).

In addition, if a question potentially does not apply to all respondents, a “Not applicable” option should be presented. If the non-substantive response applies only to certain respondent groups, we recommend offering it only to these respondents (e.g., job evaluation question for self-employed participants). All non-substantive response options should be visually detached from the substantive response options and written, for instance, in italics.

Sensitive questions such as on respondents’ age and personal income can, in the first place, be asked using an open-ended numerical format, thus, asking for very detailed level of reporting along with an explicit “Refusal” option like “I prefer not to answer this question. Respondents who select the “Refusal” option could then be directed to a second question asking about the approximate age or income range (for an illustrative overview of how income is measured in social science surveys, see Schneider et al., 2022). Generally, sensitive questions benefit from self-administered survey modes (e.g., lower item non-response, less socially desirable responding and acquiescence; Cernat et al., 2016; Chang & Krosnick, 2010; Duffy et al., 2005, Hansen et al., 2016; Olson et al., 2019) due to higher perceived anonymity.

### 2.4.2 Motivational statements (soft prompts)

Alternatively, researchers can decide not to explicitly offer non-substantive response options in web surveys, and respondents can simply leave single questions blank (referred to as skipping-allowed design; Kmetty & Stefkovics, 2022). Respondents should be informed about this option on the welcome page. This, however, has the clear disadvantage that data analysts cannot distinguish between respondents who accidentally skipped an item or question, those who did not know how to answer it, and those who did not want to answer it (Olson et al., 2019). It might, however, lead to lower rates of missing information than providing an “I don’t know” category (Kmetty & Stefkovics, 2022).

To prevent respondents from inadvertently skipping questions, soft prompts (or motivational statements) could be used. Respondents are still allowed to skip questions but receive a prompt pointing to which question(s) were not filled out along with a motivational statement underlining the importance of answering when a skip occurred (e.g., “Note: This question is very important for the further course of the questionnaire. Please try to answer it.”). If respondents confirm that they wish to leave the question blank, they are then directed to the next survey page.

Motivational statements (or soft prompts) that immediately follow a page containing an unanswered item have been shown to reduce item non-response to the same level as interviewer-based survey modes (e.g., Al Baghal & Lynn, 2015; see also DeRouvray & Couper, 2002). However, they should also be used only to a limited extent (e.g., for key questions or crucial constructs) because the motivating effect of such reminders decreases in the course of the survey (Oudejans & Christian, 2010).

In case the survey contains knowledge questions, it should be completely refrained from motivational statements. Compared to interviewer-based modes, an online mode is susceptible to biased responses for knowledge questions because respondents might look up the correct answer online, leading to a higher rate of correct answers (e.g., Clifford & Jerit, 2016; Domnich et al., 2015; Gummer & Kunz, 2022; Liu & Wang, 2014; Olson et al., 2019). A self-commitment in which respondents confirm that they will not cheat can help reduce this bias (Clifford & Jerit, 2016).

## 2.5 Survey length

Another important aspect that might arise when switching from an interviewer- to a self-administered web survey concerns the appropriate length of the questionnaire. This is particularly the case because questionnaires for interviewer-administered surveys are often designed to be longer. In contrast, the ideal length of a web survey is between ten and 15 minutes, and the maximum length of a web survey is between 20 and 28 minutes (e.g., Revilla & Höhne, 2020; Revilla & Ochoa, 2017).

If web surveys are too long, this bears the risk that respondents get tired and lose motivation, potentially leading to undesired response behavior (e.g., straightlining, survey break-off; Cernat et al., 2022; Littvay, 2009). However, this also depends on the survey itself, the survey topic, and the interest of the respondents. Initial research indicates that switching from an interviewer-administered to an online questionnaire has minimal impact on the length of the survey. In the European Value Survey (EVS), the average interview duration was 59 minutes for a face-to-face survey and 55 minutes for a web survey of the same content (Wolf et al., 2021).

Strategies to shorten the questionnaire without losing information about the target population are split questionnaire or matrix designs, in which the questionnaire is divided into shorter modules. Subsets of the modules are then administered to different respondents (Raghunathan & Grizzle, 1995). We refer readers interested in split-questionnaire designs to Little and Rhemtulla (2013) for an overview or to Axenfeld, Blom, et al. (2022) and Axenfeld, Bruch, et al. (2022).

# 3 Aspects of questionnaire design in web surveys

In addition to adaptations that must generally be taken into account when changing modes, there are specific aspects and challenges that arise due to the web mode. This second main section addresses these general issues of questionnaire design relevant to web surveys:

- Presentation of the questionnaire
- Visual design elements

## 3.1 Presentation of the questionnaire

### 3.1.1 Device types

Most web surveys offer respondents the possibility of choosing between different device types to fill out the survey: desktop devices (computers, laptops/notebooks) and mobile devices (tablets, smartphones). The choice of device can affect data quality and cause measurement error because the way a survey is displayed (e.g., the orientation of the response scale, the amount of information visible at once) can affect respondent behavior (Olson et al., 2019). This is, therefore, an important aspect in questionnaire design in an online mode.

Previous studies have found negligible or no differences in data quality between device types in terms of reliability and validity of responses (e.g., Keusch & Yan, 2017; Heerwegh, 2009; Olson et al., 2019; Sommer et al., 2017; Tourangeau et al., 2017), and in terms of undesired response behavior such as social desirability, primacy effects, and non-substantive responses (e.g., Mavletova, 2013). With regard to straightlining or non-differentiation, findings are mixed: Some studies have found less straightlining behavior among respondents using mobile devices than among those using a desktop device (e.g., Keusch & Yan, 2017; Lugtig & Toepoel, 2016), while

others have found more straightlining on smartphones compared to desktop PCs (e.g., Struminskaya et al., 2015).

Findings on the rate of missing data and length of open-ended answers varied in previous research: Some have found more item non-responses (e.g., Keusch & Yan, 2017; Lambert & Miller, 2015; Struminskaya et al., 2015) and shorter responses on mobile devices (e.g., Lambert & Miller, 2015; Mavletova, 2013; Struminskaya et al., 2015), whereas other researchers have found no differences across device types (e.g., Buskirk & Andrus, 2014). However, researchers have consistently reported that respondents using mobile devices terminate the survey more often and need more time to complete it than those using desktop devices (e.g., Buskirk & Andrus, 2014; Keusch & Yan, 2017; Lambert & Miller, 2015; Mavletova, 2013; Olson et al., 2019; Sommer et al., 2017). In addition, data from web surveys showed that estimates differ between mobile devices and desktop devices, primarily attributed to non-coverage (i.e., differences in characteristics between smartphone owners and non-owners) and potential selection errors (Antoun et al., 2019). Moreover, specific personal characteristics have been associated with the tendency to answer surveys using mobile devices: younger and more educated respondents, females, and those more familiar with online surveys (e.g., Keusch & Yan, 2017; Sommer et al., 2017). Furthermore, it has been shown that the shorter the survey, the more often respondents use their smartphones to complete it (Toepoel & Lugtig, 2018). Thus, it is strongly recommended to collect and provide information about the specific device type used, the number of logins to the survey as well as screenshots of the questionnaire format of both versions to understand and identify potential differences in measurement quality between different devices and formats within web surveys (Olson et al., 2019; for survey guidelines on web paradata in survey research and documentation of online surveys, see Kunz & Hadler, 2020 and Schaurer et al., 2020). Researchers also have the opportunity to consciously influence the device type, for example by offering a QR code to access the survey page or by indicating in the study invitation that it should ideally be completed on a desktop device.

### 3.1.2 Responsive and mobile-first design

Because mobile screens are smaller than desktop screens and orientated vertically instead of horizontally, researchers should choose a questionnaire layout that facilitates survey completion on mobile devices (e.g., prevent excessive scrolling or zooming; Antoun et al., 2018; De Bruijne & Winant, 2014), and ensure that the survey is displayed in the intended manner on all potentially used devices (Hansen et al., 2016). That is, the online questionnaire should be presented and formatted in the same way regardless of which device respondents use to fill out the questionnaire. To accommodate this, we recommend choosing a mobile-first layout. Antoun et al. (2018) provide a brief guide to effective smartphone-compatible layouts, including, for example, font sizes and touch targets that are large enough and the use of simple design and question type elements. An alternative would be a responsive design, where the programmed survey automatically adapts in its layout to the device used.

Regarding the implementation of multiple items that share the same response scale, presenting them in a matrix or grid design is a commonly used question format in web surveys due to the assumed response efficiency. As for multiple-answer questions, items are usually presented in rows, and the response scale with the response entry fields is presented in columns (Liu & Cernat, 2018). The size of the grid depends on the number of items and scale points (Couper et al., 2013). The effort required to answer the items increases with the size of the grid, as navigation becomes more difficult (Couper et al., 2013; Grady et al., 2019; Neuert et al., 2023). When large grids are displayed at full size on smartphone screens, this requires horizontal scrolling and zooming. Therefore, in a mobile-first survey design, multiple-item batteries should be presented in an item-by-item design, in which the items of a grid are presented separately with response options

displayed vertically (i.e., as stand-alone questions) either on the same survey page (scrolling design) or on separate pages (paging design; see also Antoun et al. 2018; Hadler et al., 2022; Liu & Cernat, 2018; Mavletova et al. 2018). In contrast, when using a responsive design, item batteries might be automatically adapted to the screen size by being converted into multiple single items by the software solution. In this case, the presentation format for respondents will differ depending on whether they are using a personal desktop or laptop computer or a mobile device.

In general, regardless of device type, grid formats have been found to slightly increase straight-lining, measurement error, and technical difficulty, and to decrease concurrent validity and survey evaluation, compared to item-by-item formats (Liu & Cernat, 2018; Mavletova et al., 2018), and are consequently not recommended in this form.

Moreover, research points to a trade-off between a higher number of items per screen, which decreases response time, and a higher level of item non-response or non-substantive answers (e.g., Mavletova et al., 2018; Roßmann et al., 2018; Toepoel et al., 2009). We recommend presenting a maximum of five items on one survey page to prevent excessive scrolling. Questions with more than five items should be split onto multiple survey pages because previous research has shown that five rows (and five columns for the response scale; i.e., a 5-x-5-matrix design) at the most provide the best data quality (e.g., Grady et al., 2019).

## 3.2 Visual design elements

One advantage of visually administered surveys——compared to orally administered ones——is the possibility to make questions more understandable to respondents by using visual design elements, such as different fonts, sizes, and typefaces, or even graphical tools, such as smileys, ladders, and maps (Olson et al., 2019). The most basic visual design decision involves implementing the corporate design of the research organization or field institute (i.e., inserting the logo and using the organization's color scheme and font). The framing of a survey is known to impact response behavior (Galesic et al., 2007) and to increase respondents' willingness to disclose information.

In general, web survey design should be guided by general principles of readability (Geisen & Bergstrom, 2017; Toepoel, 2017). For instance, using sans-serif fonts, adequate font size, and line spacing ensures readability on different screen sizes. It is also easier to read for people with visual impairment or low reading skills. Keywords and concepts should be highlighted using bold font (avoiding underlining or italics); highlighting longer text passages should be avoided. Background colors should be light and text dark, avoiding color schemes that put colorblind readers at a disadvantage (for an overview of factors impacting readability on the web, see Miniukovich et al., 2019). In surveys, alternating response options can be shaded, or items that have already been answered can be greyed out, decreasing item non-response (Galesic et al., 2007). However, supplementing scales with colors (i.e., using different shades of red to indicate disagreement with a statement and/or different shades of blue to indicate levels of agreement) is not recommended, as this influences response behavior (Tourangeau et al., 2007). Smileys should only be used if the instrument has been tested beforehand and there are convincing reasons for doing so (Chambers et al., 1999; Gummer et al., 2020; e.g., if the sample consists of children). In all other cases, it is not advisable to use smileys, as they have no advantages over verbally labelled response scales but increase response time (Emde & Fuchs, 2012; Gummer et al., 2020).

In addition, web surveys can assist respondents by using drag-and-drop or drop-down questions or response scales with a slider (for an overview, see Olson et al., 2019). Among different techniques, drag-and-drop formats have been shown to be most appropriate for ranked data in web surveys (Blasius, 2012). By contrast, the use of slider scales is not advisable because they

lead to higher item non-response and break-off rates, especially in lower-educated participants and among mobile device users (Funke, 2016; Funke et al., 2011). In general, it is recommended not to overload web surveys with too many (and different) technical tools (see also Funke et al., 2011). Instead, we recommend using visual design elements, such as different font sizes, but not the employment of graphical tools. This recommendation is based on previous research indicating that some graphic tools increase survey break-off rates (e.g., Funke et al., 2011).

# 4 Outlook

Surveys are increasingly switching from an interviewer-administered to a self-administered web mode, which affects aspects like question wording as well as presentation and layout of the questionnaire. This survey guideline provides an overview of adaptations that need to be considered as well as challenges associated. To adequately address the change in survey mode, some large-scale survey programs have implemented parallel runs of their questionnaires in different modes (e.g., the ESS). Empirical research needs to further study measurement mode effects to determine the effect of survey modes on respondents' response behavior.

# 5 References


Al Baghal, T., & Lynn, P. (2015). Using motivational statements in web-instrument design to reduce item-missing rates in a mixed-mode context. *Public Opinion Quarterly, 79*(2), 568–579. https://doi.org/10.1093/poq/nfv023

Antoun, C., Conrad, F. G., Couper, M. P., & West, B. T. (2019). Simultaneous estimation of multiple sources of error in a smartphone-based survey. *Journal of Survey Statistics and Methodology, 7*(1), 93–117. https://doi.org/10.1093/jssam/smy002

Antoun, C., Katz, J., Argueta, J., & Wang, L. (2018). Design heuristics for effective smartphone questionnaires. *Social Science Computer Review, 36(*5), 557–574. https://doi.org/10.1177/089443931772707

Axenfeld, J. B., Blom, A. G., Bruch, C. & Wolf, C. (2022). Split questionnaire designs for online surveys: The impact of module construction on imputation quality. *Journal of Survey Statistics and Methodology, 10*(5), 1236–1262. https://doi.org/10.1093/jssam/smab055

Axenfeld, J. B., Bruch, C., & Wolf, C. (2022). General-purpose imputation of planned missing data in social surveys: Different strategies and their effect on correlations. *Statistics Surveys, 16*, 182–209. https://doi.org/10.1214/22-SS137

Bandilla, W. (2015). *Online – Befragungen* [Online – Surveys] (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_003

Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2017). *Web probing – Implementing probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions* (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_023

Beuthner, C., Daikeler, J., & Silber, H. (2019). *Mixed-device and mobile web surveys* (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_028

Bishop G. F., Hippler H. J., Schwarz N., Strack F. (1988). A comparison of response effects in self-administered and telephone surveys. In R. B. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L.

Nicholls II, & J. Waksberg (Eds.), Telephone survey methodology (pp. 321–358). John Wiley & Sons.

Blasius, J. (2012). Comparing ranking techniques in web surveys. *Field Methods, 24*(4), 382–398. https://doi.org/10.1177/1525822X1244309

Breton, C., Cutler, F., Lachance, S., & Mierke-Zatwarnicki, A. (2017). Telephone versus online survey modes for election studies: Comparing Canadian public opinion and vote choice in the 2015 federal election. *Canadian Journal of Political Science, 50*(4), 1005–1036. https://doi.org/10.1017/S0008423917000610

Buskirk, T. D., & Andrus, C. H. (2014). Making mobile browser surveys smarter: Results from a randomized experiment comparing online surveys completed via computer or smartphone. *Field Methods, 26*(4), 322–342. https://doi.org/10.1177/1525822X145261

Callegaro, M., Murakami, M. H., Tepman, Z., & Henderson, V. (2015). Yes-no answers versus check-all in self-administered modes. *International Journal of Market Research, 57*(2), 203–223. https://doi.org/10.2501/IJMR-2015-014a

Cernat, A., Couper, M. P., & Ofstedal, M. B. (2016). Estimation of mode effects in the health and retirement study using measurement models. *Journal of Survey Statistics and Methodology, 4*(4), 501–524. https://doi.org/10.1093/jssam/smw021

Cernat, A., Sakshaug, J., Christmann, P., & Gummer, T. (2022). The impact of survey mode design and questionnaire length on measurement quality. *Sociological Methods & Research*. Advance online publication. https://doi.org/10.1177/00491241221140139

Chambers, C. T., Giesbrecht, K., Craig, K. D., Bennett, S. M., & Huntsman, E. (1999). A comparison of faces scales for the measurement of pediatric pain: Children's and parents' ratings. *Pain, 83*(1), 25–35. https://doi.org/10.1016/S0304-3959(99)00086-X

Chang, L., & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered computerized questionnaires. An experiment. *Public Opinion Quarterly, 74*(1), 154–167. https://doi.org/10.1093/poq/nfp090

Christian, L. M., & Dillman, D. A. (2004). The influence of symbolic and graphical language manipulations on answers to paper self-administered questionnaires. *Public Opinion Quarterly, 68*(1), 57–80. https://doi.org/10.1093/poq/nfh004

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping respondents get it right the first time: The relative influence of words, symbols, and graphics in web and telephone surveys. *Public Opinion Quarterly, 71*(1), 113–125. https://doi.org/10.1093/poq/nfl039

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2008). The effects of mode and format on answers to scalar questions in telephone and web surveys. In J. Lepkowski, C. Tucker, M. Brick, E. DeLeeuw, L. Japec, P. Lavrakas, M. Link, & R. Sangster (Eds.), *Advances in telephone survey methodology* (pp. 250–275). Wiley. https://doi.org/10.1002/9780470173404.ch12

Chyung, S. Y., Kennedy, M., & Campbell, I. (2018). Evidence-based survey design: The use of ascending or descending order of Likert-type response options. *Performance Improvement, 57*(9), 9–16. https://doi.org/10.1002/pfi.21800

Clifford, S., & Jerit, J. (2016). Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions. *Public Opinion Quarterly, 80*(4), 858–887. https://doi.org/10.1093/poq/nfw030

Conrad F. G., Couper M. P., Tourangeau R., & Galesic M. (2005). *Interactive feedback can improve quality of responses in web surveys* [Paper presentation]. 60th Annual Conference of the American

Association for Public Opinion Research, Miami Beach, FL. http://www.asasrms.org/Proceedings/y2005/files/JSM2005-000938.pdf

Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2006). Use and non-use of clarification features in web surveys. *Journal of Official Statistics, 22*(2), 245–269. https://www.proquest.com/scholarly-journals/use-non-clarification-features-web-surveys/docview/1266792618/se-2

Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing input fields for non-narrative open-ended responses in web surveys. *Journal of Official Statistics, 27*(1), 65–85. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3570266/

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review, 31*(3), 322–345. https://doi.org/10.1177/0894439312469 8

De Bruijne, M., & Wijnant, A. (2014). Improving response rates and questionnaire design for mobile web surveys. *Public Opinion Quarterly, 78*(4), 951–962. https://doi.org/10.1093/poq/nfu046

DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity, 52*, 1523–1559. https://doi.org/10.1007/s11135-017-0533-4

DeRouvray, C., & Couper, M. P. (2002). Designing a strategy for reducing "no opinion" responses in web-based surveys. *Social Science Computer Review, 20*(1), 3–9. https://doi.org/10.1177/089443930202000101

Dillman, D. A. (2017). The promise and challenge of pushing respondents to the web in mixed-mode surveys. *Survey Methodology, 43*(1), 3–31. https://www.nonresponse.org/wp-content/uploads/2021/03/DnD152346272825Dillman-WaksbergpaperSurveymethododlogyJune272017-1.pdf

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods, 17*(1), 30–52. https://doi.org/10.1177/1525822X04269550

Dillman, D. A., & Edwards, M. L. (2016). Designing a mixed-mode survey. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Eds.), *The Sage handbook of survey methodology* (pp. 255–268). London

Dillman, D. A., & Smyth, J. D. (2007). Design effects in the transition to web-based surveys. *American Journal of Preventive Medicine, 32*(5), S90–S96. https://doi.org/10.1016/j.amepre.2007.03.008

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons.

Domnich, A., Panatto, D., Signori, A., Bragazzi, N. L., Cristina, M. L., Amicizia, D., & Gasparini, R. (2015). Uncontrolled web-based administration of surveys on factual health-related knowledge: A randomized study of untimed versus timed quizzing. *Journal of Medical Internet Research, 17*(4), e94. https://doi.org/10.2196/jmir.3734

Dorer, B. (2012). *ESS Round 6 translation guidelines*. GESIS – Leibniz Institute for the Social Sciences. http://www.europeansocialsurvey.org/docs/round6/methods/ESS6_translation_guidelines.pdf

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research, 47*(6), 615–639. https://doi.org/10.1177/147078530504700602

Emde, M., & Fuchs, M. (2012). Exploring animated faces scales in web surveys: Drawbacks and prospects. *Survey Practice, 5*(1), 1–6. https://doi.org/10.29115/SP-2012-0006

Funke, F. A. (2016). Web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review, 34*(2), 244–254. https://doi.org/10.1177/0894439315575477

Funke, F., Reips, U. D., & Thomas, R. K. (2011). Sliders for the smart: Type of rating scale on the Web interacts with educational level. *Social Science Computer Review, 29*(2), 221–231. https://doi.org/10.1177/0894439310376896

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). *Using change to improve navigation in grid questions* [Paper presentation]. General Online Research Conference (GOR 07), Leipzig, Germany.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*(5), 892–913. https://doi.org/10.1093/poq/nfn059

Geisen, E., & Bergstrom, J. R. (2017). *Usability testing for survey research*. Elsevier.

Grady, R. H., Greenspan, R. L., & Liu, M. (2019). What is the best size for matrix-style questions in online surveys? *Social Science Computer Review, 37*(3), 435–445. https://doi.org/10.1177/0894439318773733

Gummer, T., & Kunz, T. (2022). Relying on external information sources when answering knowledge questions in web surveys. *Sociological Methods & Research, 51*(2), 816–836. https://doi.org/10.1177/0049124119882470

Gummer, T., Vogel, V., Kunz, T., & Roßmann, J. (2020). Let's put a smile on that scale: Findings from three web survey experiments. *International Journal of Market Research, 62*(1), 18–26. https://doi.org/10.1177/1470785319858598

Hadler, P., Lenzner, T., Schick, L., Neuert, C. (2022). *European Working Conditions Survey 2024: Preparation and cognitive testing of the online questionnaire* (Working Paper WPEF22035). Eurofound. https://www.eurofound.europa.eu/sites/default/files/wpef22035.pdf

Hansen, S. E., Lee, H. J., Lin, Y., & McMillan, A. (2016). *Instrument technical design* (Cross-cultural survey guidelines). Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI. https://ccsg.isr.umich.edu/chapters/instrument-technical-design/

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research, 21*(1), 111–121. https://doi.org/10.1093/ijpor/edn054

Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly, 72*(5), 836–846. https://doi.org/10.1093/poq/nfn045

Höhne, J. K., & Lenzner, T. (2015). Investigating response order effects in web surveys using eye tracking. *Psihologija, 48*(4), 361–377. https://doi.org/10.2298/PSI1504361H

Hox, J. J., De Leeuw, E. D., & Zijlmans, E. A. O. (2015). Measurement equivalence in mixed mode surveys. *Frontiers in Psychology, 6*, Article 87. https://doi.org/10.3389/fpsyg.2015.00087

Keusch, F., & Yan, T. (2017). Web versus mobile web: An experimental study of device effects and self-selection effects. *Social Science Computer Review, 35*(6), 751–769. https://doi.org/10.1177/0894439316675566

Kmetty, Z., & Stefkovics, Á. (2022). Assessing the effect of questionnaire design on unit and item-nonresponse: Evidence from an online experiment. *International Journal of Social Research Methodology, 25*(5), 659–672. https://doi.org/10.1080/13645579.2021.1929714

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Michael Hanemann, W., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, K., Moody, W. R., Green, M. C., & Conaway, M. (2002). The impact of "No opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly, 66*(3), 371–403. https://doi.org/10.1086/341394

Kunz, T., & Hadler, P. (2020). *Web paradata in survey research* (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_037c

Lambert, A. D., & Miller, A.L. (2015). Living with smartphones: Does completion device affect survey responses? *Research in Higher Education, 56*(2), 166–177. https://doi.org/10.1007/s11162-014-9354-7

Lenzner, T., Neuert, C., & Otto, W. (2015). *Kognitives Pretesting* [Cognitive pretesting] (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_010

Lenzner, T., Neuert, C., Otto, W., & Best, F. (2014). *wb-Personalmonitor 2014. Kognitiver Pretest* [wb personnel monitor 2014. Cognitive pretest] (GESIS-Projektbericht 2014/02). GESIS-Pretestlabor. http://doi.org/10.17173/pretest4

Lenzner, T., Neuert, C., Otto, W., Landrock, U., Adams, F., Disch, K., Gebhardt, S. & Menold, N. (2015). *Computer-assisted measurement and coding of educational qualifications in surveys (CAMCES). Kognitiver Pretest* [Cognitive pretest] (GESIS-Projektbericht 2015/05). GESIS-Pretestlabor. http://doi.org/10.17173/pretest25

Littvay, L. (2009). Questionnaire design considerations with planned missing data. *Review of Psychology, 16*(2), 103–114.

Little, T. D. & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7*(4), 199–204. https://doi.org/10.1111/cdep.12043

Liu, M., & Cernat, A. (2018). Item-by-item versus matrix questions: A web survey experiment. *Social Science Computer Review, 36*(6), 690–706. https://doi.org/10.1177/0894439316674459

Liu, M., & Keusch, F. (2017). Effects of scale direction on response style of ordinal rating scales. *Journal of Official Statistics, 33*(1), 137–154. http://doi.org/10.1515/JOS-2017-0008

Liu, M., & Wang, Y. (2014). Data collection mode effects on political knowledge. *Survey Methods: Insights from the Field*. Retrieved from https://doi.org/10.13094/SMIF-2014-00009

Lugtig, P., & Toepoel, V. (2016). The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error. *Social Science Computer Review, 34*(1), 78–94. https://doi.org/10.1177/0894439315574248

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review, 31*(6), 725–743. https://doi.org/10.1177/0894439313485201

Mavletova, A., Couper, M. P., & Lebedev, D. (2018). Grid and item-by-item formats in PC and mobile web surveys. *Social Science Computer Review, 36*(6), 647–668. https://doi.org/10.1177/0894439317735307

Menold, N., & Bogner, K. (2015). *Gestaltung von Ratingskalen in Fragebögen* [Design of rating scales in questionnaires] (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_015

Metzler, A., Kunz, T., & Fuchs, M (2015). The use and positioning of clarification features in web surveys. *Psihologija, 48*(4), 379–408. https://doi.org/0.2298/PSI1504379M

Miniukovich, A., Scaltritti, M., Sulpizio, S., & De Angeli, A. (2019). *Guideline-based evaluation of web read-ability* [Paper presentation]. CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), Glasgow, Scotland, UK. https://doi.org/10.1145/3290605.3300738

Neuert, C. E. (2017). Processing forced-choice versus check-all-that-apply question formats: Evidence from eye tracking. *Field Methods, 29*(4), 383–394. https://doi.org/10.1177/1525822X17703943

Neuert, C. E. (2020). Do forced-choice (FC) questions trigger deeper cognition than check-all-that-apply (CATA) questions? *Journal of Survey Statistics and Methodology, 8(*4), 617–635. https://doi.org/10.1093/jssam/smz015

Neuert, C. E., Roßmann, J., & Silber, H. (2023). Using eye-tracking methodology to study grid question designs in web surveys. *Journal of Official Statistics, 39*(1), 1–24. http://doi.org/10.2478/JOS-2023-0004

Nicolaas, G., Campanelli, P., Hope, S., Jäckle, A., & Lynn, P. (2011). *Is it a good idea to optimise question format for mode of data collection? Results from a mixed modes experiment* (ISER Working Paper Series, No. 2011-31). University of Essex, Institute for Social and Economic Research (ISER), Colchester. http://hdl.handle.net/10419/65965

Nicolaas, G., Campanelli, P., Hope, S., Jäckle, A., & Lynn, P. (2015). Revisiting "yes/no" versus "check all that apply": Results from a mixed modes experiment. *Survey Research Methods, 9*(3), 189–204. https://doi.org/10.18148/srm/2015.v9i3.6151

Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N., McCarthy, J., O'Brien, E., Opsomer, J., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C., & Wagner, J. (2019). *Report of the AAPOR task force on transitions from telephone surveys to self-administered and mixed-mode surveys*. American Association for Public Opinion Research. https://www.aapor.org/getattachment/Education-Resources/Reports/AAPOR-Mixed-Mode-Task-Force-Executive-Summary-FINAL.pdf.aspx

Oudejans, M., & Christian, L. M. (2010). Using interactive features to motivate and probe responses to open-ended questions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), Social and behavioral research and the internet: Advances in applied methods and research strategies (pp. 215–244). Routledge.

Peytchev, A., Couper, M. P., McCabe, S. E., & Crawford, S. D. (2006). Web survey design: Paging versus scrolling. *Public Opinion Quarterly*, 70(4), 596–607. https://doi.org/10.1093/poq/nfl028

Porst, R. (2011). *Fragebogen. Ein Arbeitsbuch* [Questionnaire. A workbook] (3rd edition). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92884-5

Raghunathan, T. E., Grizzle J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association, 90*(429), 54–63. https://doi.org/10.2307/2291129

Redline, C. (2013). Clarifying categorical concepts in a web survey. *Public Opinion Quarterly, 77*(S1), 89–105. https://doi.org/10.1093/poq/nfs067

Redline, C., Dillman, D. A., Dajani, A. N., & Scaggs, M. A. (2003). Improving navigational performance in U.S. Census 2000 by altering the visually administered languages of branching instructions. *Journal of Official Statistics, 19*(4), 403–420. https://www.proquest.com/scholarly-journals/improving-navigational-performance-u-s-census/docview/1266794088/se-2

Revilla, M. & Höhne, J. K. (2020). How long do respondents think online surveys should be? New evidence from two online panels in Germany. *International Journal of Market Research, 62*(5), 538–545. https://doi.org/10.1177/1470785320943049

Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research, 59*(5), 557–565. https://doi.org/10.2501/IJMR-2017-039

Robie, C., Meade, A. W., Risavy, S. D., & Rasheed, S. (2022). Effects of response option order on Likert-type psychometric properties and reactions. *Educational and Psychological Measurement, 82*(6), 1107–1129. https://doi.org/10.1177/00131644211069406

Roßmann, J., Gummer, T., & Silber, H. (2018). Mitigating satisficing in cognitively demanding grid questions: Evidence from two web-based experiments. *Journal of Survey Statistics and Methodology, 6*(3), 376–400. https://doi.org/10.1093/jssam/smx020

Schaurer, I., Kunz, T., & Heycke, T. (2020). *Documentation of online surveys* (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_031

Schmid, L., Kunz, T., & Naumann, E. (2023). Questionnaire design in the FReDA panel recruitment: Challenges in transitioning from a face-to-face to a self-administered mixed-mode design. *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2023-00006

Schneider, S. L., Ortmanns, V., Diaco, A., & Müller, S. (2022). *Die Erhebung sozio-demographischer Variablen in großen deutschen Umfragen: Ein Überblick über Möglichkeiten und Herausforderungen der Harmonisierung* [The collection of socio-demographic variables in large German surveys: An overview of harmonization opportunities and challenges (KonsortSWD Working Paper 2/2022). GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.5281/zenodo.6810973

Schober, M. F, Conrad, F. G, & Bloom, J. E. (2000). Clarifying word meanings in computer-administered survey interviews. *Proceedings of the Annual Meeting of the Cognitive Science Society, 22*. https://escholarship.org/uc/item/0846695d

Smyth, J. D., Christian, L. M., & Dillman, D. A. (2008). Does "yes or no" on the telephone mean the same as "check-all-that-apply" on the web? *Public Opinion Quarterly, 72*(1), 103–113. https://doi.org/10.1093/poq/nfn005

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly, 70*(1), 66–77. https://doi.org/10.1093/poq/nfj007

Sommer, J., Diedenhofen, B., & Musch, J. (2017). Not to be considered harmful: Mobile-device users do not spoil data quality in web surveys. *Social Science Computer Review, 35*(3), 378–387. https://doi.org/10.1177/0894439316633452

Stadtmüller, S. Beuthner, C., & Silber, H. (2021). *Mixed-mode surveys* (GESIS Survey Guidelines). Mannheim, GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_038

Struminskaya, B., Weyandt, K., & Bosnjak, M. (2015). The effects of questionnaire completion using mobile devices on data quality. Evidence from a probability-based general population panel. *Methods, data, analyses, 9*(2), 261–292. https://doi.org/10.12758/mda.2015.014

Terentev, E., & Maloshonok, N. (2019). The impact of response options ordering on respondents' answers to rating questions: Results of two experiments. *International Journal of Social Research Methodology, 22*(2), 179–198. https://doi.org/10.1080/13645579.2018.1510660

Thomas, R. K., & Klein, J. D. (2006). Merely incidental? Effects of response format on self-reported behavior. *Journal of Official Statistics, 22*, 221–244. https://www.proquest.com/scholarly-journals/merely-incidental-effects-response-format-on-self/docview/1266792627/se-2

Toepoel, V. (2017). Online survey design. In N. G. Fielding, R. M. Lee, G. & Blank (Eds.), *The SAGE handbook of online research methods* (2nd edition, pp. 184–202). SAGE Publications Ltd. https://doi.org/10.4135/9781473957992

Toepoel, V., Das, M., & Van Soest, A. (2009). Design of web questionnaires: The effects of the number of items per screen. *Field Methods, 21*(2), 200–213. https://doi.org/10.1177/1525822X08330261

Toepoel, V., & Lugtig, P. (2018). Modularization in an era of mobile web: Investigating the effects of cutting a survey into smaller pieces on data quality. *Social Science Computer Review, 40*(1), 150–164. https://doi.org/10.1177/0894439318784882

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly, 68*(3), 368–393. https://doi.org/10.1093/poq/nfh035

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly, 71*(1), 91–112. https://doi.org/10.1093/poq/nfl046

Tourangeau, R., Maitland, A., Rivero, G., Sun, H., Williams, D., & Yan, T. (2017). Web surveys by smartphone and tablets: Effects on survey responses. *Public Opinion Quarterly, 81*(4), 896–929. https://doi.org/10.1093/poq/nfx035

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting general social surveys as self-administered mixed-mode surveys. *Public Opinion Quarterly, 85*(2), 623–648. https://doi.org/10.1093/poq/nfab039