WORKING PAPER

Eurofound

Working conditions and sustainable work

# Good practices in transitioning to an online mode in 3MC questionnaire design

**Authors:** Désirée Nießen, Patricia Hadler, Timo Lenzner and Cornelia Neuert (GESIS – Leibniz Institute for the Social Sciences)

**Research managers:** Agnès Parent-Thirion and Eleonora Peruffo

**Contributors:** Isabella Biletta (Eurofound) and Karolina Predotova (Econresearchkp)

**Acknowledgements:** We thank Dorothée Behr, Christian Bruch, Tobias Gummer, and Kai Weyandt for sharing their expertise and valuable comments on the report. In particular, Dorothée Behr for her expertise in questionnaire translation, Christian Bruch for his helpful comments on questionnaire modularization, Tobias Gummer for advice on survey retention strategies, and Kai Weyandt for sharing information on the measurement of ISCO in the GESIS Panel. In addition, we thank Cara Nordhoff, Friederike Quint, and Patricia Steins for their able research assistance.

**Eurofound reference number:** WPEF22034



*Research carried out prior to the UK's withdrawal from the European Union on 31 January 2020, and published subsequently, may include data relating to the 28 EU Member States. Following this date, research only takes into account the 27 EU Member States (EU28 minus the UK), unless specified otherwise.*

The European Foundation for the Improvement of Living and Working Conditions (Eurofound) is a tripartite European Union Agency established in 1975. Its role is to provide knowledge in the area of social, employment and work-related policies according to Regulation (EU) 2019/127.

**European Foundation for the Improvement of Living and Working Conditions**

**Telephone:** (+353 1) 204 31 00

**Email:** information@eurofound.europa.eu

**Web:** www.eurofound.europa.eu

# Contents

# Introduction

Large-scale surveys are increasingly switching to self-administered or mixed-mode data collection, both due to rising costs of solely interviewer-administered surveys and declining response rates. Although web-based 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 major large-scale surveys also decided to switch modes to ensure collection of data. However, a switch in survey modes or the combination of modes can lead to so-called measurement mode effects – 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). Therefore, adaptations of a questionnaire from an interview-administered face-to-face to an online self-administered mode have to ensure high-quality data, while maintaining maximal equivalence with the face-to-face source questionnaire. When adapting cross-national surveys, the adaptation of the source questionnaire should put the multinational, multiregional, and multicultural context (3MC) perspective at the heart of all considerations.

This report summarizes knowledge on how to transition interviewer-administered, cross-national surveys to an online mode. The report is based on a literature review and complemented by input from experts in the areas of transitioning surveys to online, cross-cultural survey methodology, and survey modularization. Details on the experts and their contributions can be found in the acknowledgements.

# Aspects of question design to be considered during transition from face-to-face to online mode

This section is divided into two main parts. First, it addresses general issues of question design to consider when adapting an interviewer-based source questionnaire into a self-administered web-based questionnaire. Second, specific survey questions and issues are addressed that are of great importance for the European Working Conditions Survey (EWCS).

## General issues of question design

The following general issues of question design have to be considered when transitioning survey mode (for a comprehensive overview, see Olson et al., 2019):

(a) substantive response options and formats;

(b) inclusion of non-substantive response options;

(c) presentation of definitions or clarifications;

(d) addressing of the respondents and question wordings;

(e) presentation of the questionnaire.

### Substantive response options and formats

Ordinal response scales

One well-documented effect observable when switching survey modes regards opinion questions using ordinal scales. Aurally administered 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, such as primary and recency effects, socially desirable and acquiescent responding (e.g., Bishop et al., 1998; Cernat et al., 2016; Chang & Krosnick, 2010; Olson et al., 2019). To keep these mode differences as small as possible, it is necessary to provide respondents with as much assistance and facilitations as possible in self-administered surveys.

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 found that scale direction impacts response distribution (i.e., Terentev & Maloshonok, 2019), but that effects are small (i.e., 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) 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), 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 translated questionnaires should not differ from the source questionnaire (e.g., Dorer, 2012).

Visual design elements

One advantage of visually administered surveys is – compared to aurally administered ones – the possibility to make questions more understandable to respondents by using visual design elements as, for example, different fonts, sizes, and typefaces, or even graphical tools as, for example, 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 colour scheme and font). The framing of a survey is known to impact response behaviour (Galesic et al., 2007). In the case of the EWCS, clearly framing the survey as a research project on working conditions by Eurofound may increase respondents' willingness to disclose information on illicit or precarious work.

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 and is 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 colours should be light and text dark, avoiding colour schemes that put colourblind 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 colours (i.e., using different shades of red to indicate disagreement with a statement and/or different shades of blue to indicate levels of agreement) cannot be recommended, as this influences response behaviour (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-based 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 online-based surveys (Blasius, 2012). By contrast, the use of slider scales is not advisable because they lead to higher response and break-off rates, especially in lower-educated participants and among mobile device users (Funke et al., 2011; Funke, 2016). In general, it is recommended not to overload web surveys with too many (and different) technical tools (see also Funke et al., 2011).

For the adapted online questionnaire in the EWCS, we recommend using visual design elements, such as font size, but not the employment of graphical tools. This recommendation is based on previous research indicating that some graphic tools increase survey break-off rates and to maintain comparability to the interviewer-administered format.

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. Thereby, 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 require 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 period for numbers, cents instead of dollars; see also Olson et al., 2019). First, the instructions should be clear enough and include, for example, a highly visible good 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 the EWCS, several questions provide the opportunity to offer the respondents a choice of response units. 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). We recommend programming plausibility checks to ensure that respondents enter the correct format and provide additional explanation on the range of valid answers (e.g., “Please enter a whole number between 15 and 100.”).

Multiple-answer questions

The handling of multiple-answer questions should also be taken into account when moving to an online-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” and “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 self-administered and visually presented questionnaires, often an alternative option is used in which multiple-answer questions are 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 the items that apply to them.

However, it has been shown that both formats yield different responses (more categories selected and longer completion times in the yes-no compared to the check-all-that-apply format; e.g., Smyth et al., 2006; Neuert, 2020). 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). Therefore, it is

generally recommended to use the same format in all modes to ensure comparability of data and to prefer the yes-no format in most cases (e.g., Nicolaas et al., 2011) because it induces a deeper processing of the answer options and is more comparable across modes (e.g., Smyth et al., 2006, 2008), even though it is more likely to elicit acquiescent responding (e.g., Callegaro et al., 2015).

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 in order to avoid item non-responses that would be difficult to understand (Olson et al., 2019).

However, for the adapted online questionnaire of the EWCS, we tested a check-all-that-apply format for factual questions, such as when asking about health problems ("Over the last 12 months, did you have any of the following health problems?") and income ("Thinking about your earnings from your main job, what do they include?"). Results clearly showed that even for these questions, a check-all-that-apply format leads to low-quality responses and cannot be recommended.

Regarding the implementation of item batteries in an item-by-item design, the amount of information and the effort required to answer the items increases with the number of items, making navigation more difficult (Couper et al., 2013; Grady et al., 2019). 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 (Toepoel et al., 2009). In general, up to five items per screen are recommended (Grady et al., 2019; for a grid design, four to ten items have been suggested, Toepoel et al., 2009).

## Inclusion of non-substantive response options

Next to the presentation of substantive response options, survey designers adapting a questionnaire to a self-administered 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." In interviewer-administered surveys, non-substantive response options are generally not read aloud, but only noted by the interviewer when prompting the respondent does not elicit a response. In self-administered modes, 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). For example, in the German part of the European Values Survey (EVS) 2017/2018, which was administered either as face-to-face survey or as a self-administered mixed-mode survey in full length or in a matrix design, the share of "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).

Alternatively, researchers can decide not to explicitly offer a non-substantive response option, and respondents can simply leave single questions blank. 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). In this regard, interviewer-based survey modes benefit from the social context in the sense that interviewers can observe and support respondents, for example, by motivating them to answer a question, to complete the survey, by following up, or by giving clarifications or other help (Olson et al., 2019). However, in the case of one particular type of question, it should be refrained from motivations statements, namely, for knowledge questions. Compared to interviewer-based

interviews, self-administered (web-based) modes are 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; 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).

Handling item non-response in general

Concerning item non-response, both self-administered and online survey modes have advantages and disadvantages. Generally, self-administered modes show a slightly higher rate of item non-response (e.g., Heerwegh, 2009; Heerwegh & Loosveldt, 2008; Olson et al., 2019), but switching to a CAWI (computer-assisted web interviewing) mode has shown lower item non-response than switching to paper-and-pencil interviewing (PAPI) via mail, which can be explained by the fact that online surveys can either force or at least request respondents to give an answer (Olson et al., 2019). Motivational statements that immediately follow a page containing an unanswered item have been shown to reduce item non-response to the same level of interviewer-based survey modes (e.g., Al Baghal & Lynn, 2015; see also DeRouvray & Couper, 2002).

Because there is no simple answer to the question of whether non-substantive answers should be provided, we recommend making this decision for each question individually. In general, we recommend not to offer respondents explicit non-substantive response options in the online mode. Instead, respondents should be informed that they may leave questions unanswered on the welcome page. To prevent respondents from inadvertently skipping questions, they should receive a prompt pointing to which question(s) were not filled out along with a motivational statement underlining the importance of answering (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 empty, they should be directed to the next survey page.

If respondents try to skip a question or item, they can be presented a motivational statement, and also an explicit non-response option, either alongside the motivational statement or if the respondent tries to leave the item unanswered after being presented the motivational statement (Al Baghal & Lynn, 2015).

However, for certain types of questions in the EWCS, we recommend offering a non-substantive response option to respondents explicitly in the adapted online questionnaire. When it is plausible that respondents do not know the answer, providing a “Don’t know” option might be preferable (e.g., a question on the number of people working at one’s workplace, which might not be answerable by people who have joined a company only recently). 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).

Handling item non-response in sensitive questions

Especially sensitive questions, which are not so easy to ask in a personal interview (e.g., questions on the respondents’ income, discrimination at work, criminal past, (drug) abuse, or sexual behaviour), profit from web-based surveys leading to lower item non-response and thus better data quality (e.g., Hansen et al., 2016). This is probably due to the fact that online surveys are more anonymous than face-to-face, telephone, and even mail surveys. This more anonymous administration also in general lowers the tendency towards social desirability and acquiescence (e.g., Cernat et al., 2016; Chang &

Krosnick, 2010; Duffy et al., 2005, Olson et al., 2019) because it is not necessary to avoid embarrassing situations as if an interviewer was present (Tourangeau & Yan, 2007). Furthermore, the potential impact of the interviewer on attitudinal questions (e.g., by gender, ethnicity, or appearance) is prevented in self-administered survey modes (Olson et al., 2019) and self-administered modes have shown to produce lower random errors (Klausch et al., 2013).

For sensitive questions in the adapted online questionnaire of the EWCS, we recommend presenting an explicit refusal option (i.e., "I prefer not to answer this question"). Questions on respondents' age and personal income can, for example, be asked using an open-ended numerical format, thus, asking for very detailed level of reporting. Respondents who select the "refusal option" could then be directed to a second question asking about the approximate age or income range. For the question on income, an explicit refusal option should be offered as well. All non-substantive response options should be visually detached from the substantive response options and written, for instance, in italics.

## Presentation of definitions or clarifications

In self-administered 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.

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 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 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 based on their previous responses (see also Olson et al., 2019). EWCS-related examples of this are clarifications specific to the work situation of employed or self-employed participants and clarifications for respondents with more than one job.

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). If additional information is presented to all respondents, the placement of this information should (in most cases) occur directly after/below the question stem/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 before 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, parentheticals 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 of the visual design of online surveys influencing data quality (Hansen et al., 2016).

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 the survey and unnecessary burden 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). From these techniques, it has been found that rollover definitions are requested more frequently than click definitions (Conrad et al., 2006) – even though definitions are accessed most often when presented directly (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, the easiest access to definitions is an important issue to consider when transitioning to web-based 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. Some questions can contain both clarifications visible to all respondents and clarifications via mouse rollover.

## Addressing of the respondents and question wordings

The adaptation to a self-administered format 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).

Introductory pages and bridges

In the EWCS CAPI (computer-assisted personal interviewing) questionnaire, the interviewer generally introduces the next topic with sentences such as "I'm now going to ask you some questions about…." In self-administered online questionnaires, four adaptations have to be made: First, the introductory text should be placed either at the top of a survey page with the next question or on a separate survey page. Second, the adaptation must replace the first person singular ("I") with a neutral wording, such as "The next questions are about…." Third, it is recommended that important words that an interviewer would emphasize are visually highlighted, for instance, using bold font. For the EWCS, for example, we recommend that for respondents with multiple jobs, the words "main" or "all" be highlighted, depending on whether respondents should answer only for their main job or consider all jobs. Fourth, introductory sections that are sentence fragments in CAPI have to be turned into full sentences in the adaptation. We recommend, for instance, to change the phrase "Now thinking about your business…" to "The following questions are about your business."

Show cards and instructions

Another common example of necessary adjustments to question wording is when the CAPI questionnaire uses show cards to facilitate data collection, which are sometimes referred to in the

question text. Normally, these show cards can simply be replaced with a corresponding single-choice questions 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 (e.g., “Please select one answer” or “Please select all answers that apply to you”) must be inserted in self-administered questionnaires. In all cases, the question or item text should be adapted as minimally as possible.

## Presentation of the questionnaire

### Device types

Unlike paper-and-pencil surveys and surveys conducted by interviewers, respondents to online-based surveys use different device types: desktop devices (computers, laptops/notebooks) and mobile devices (tablets, smartphones). It has been shown that the shorter the survey, the more respondents use their smartphone to complete it (Toepoel & Lugtig, 2018). These different devices can affect data quality and cause measurement errors because the different display of the survey can have different effects on the respondents (Olson et al., 2019) or, in the case of sensitive questions, because respondents are less willing to share sensitive information when using mobile device features (Mavletova & Couper, 2013). This is, therefore, an important aspect in questionnaire design when transitioning to a web-based 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 primacy effects, socially desirable, and non-substantive responses (e.g., Mavletova, 2013). Furthermore, less straight-lining behaviour has been found among respondents using mobile devices than among those using a desktop device (Keusch & Yan, 2017).

Findings on missing data rates and length of open-ended answers varied in previous research. Some found more missing data (e.g., Keusch & Yan, 2017) and shorter responses on mobile devices (e.g., Mavletova, 2013), whereas other researchers 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 the survey 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, web survey results showed biased estimates of mobile devices compared to 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).

Specific person characteristics have been associated with the tendency to answer surveys using mobile devices: younger and more educated respondents, females, and those more familiar with web surveys (e.g., Keusch & Yan, 2017; Sommer et al., 2017). Thus, it is strongly recommended to collect and provide information about the specific device type used and the number of logins to the survey as well as screenshots of the questionnaire format of both versions in order to understand and identify potential differences in measurement quality between different devices and formats within web-based surveys (Olson et al., 2019).

Mobile-first questionnaire layout

Because mobile screens are, first, smaller than desktop screens and, second, 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), and ensure that the survey is displayed in the intended manner on all possible 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 for this, we recommend choosing a mobile-first layout. In doing so, aspects like the number of items that are presented on one screen must be determined. 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. Therefore, visual features such as graphics and animations should only be used sparingly and when necessary (Dillman & Smyth, 2007).

Presenting item batteries in a mobile-first format requires adaptation to several parts of the question structure. Items must be presented separately with vertically shown response options. This is known as an item-by-item format. In consequence, the introductory question text is only visible to respondents while they fill out the first item(s), but not as they scroll down to answer later items of an item battery (at least in standard online survey software). In general, regardless of device type, grid formats have been found to slightly increase straight-line responding, measurement error, and technical difficulty, and decrease concurrent validity and survey evaluation, compared to item-by-item formats (Liu & Cernat, 2018; Mavletova et al., 2018).

In the EWCS, for instance, the CAPI question text read “Regarding your business, do you…?,” followed by the items read out by the interviewer, for instance, “… have the authority to hire or dismiss employees.” In the adapted online questionnaire, the question text and items should consist of full sentences: “Regarding your business, do the following statements apply to you?” and “I have the authority to hire or dismiss employees.” In addition, for the EWCS adapted CAWI questionnaire, 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). Furthermore, a randomisation of items is recommended whenever possible. In general, items should be randomised within survey pages, for instance, to prevent order effects (Dillman et al., 2014).

# Specific survey questions and issues

In addition to these general decisions of question design, specific survey questions require special adaptations to a self-administered format, leading to changes in question wording or even question type. Sample questions that are especially relevant in the EWCS refer to (a) respondent’s occupation and industry. Moreover, adaptation must consider aspects of (b) gender mainstreaming.

## Respondent’s occupation and industry

Occupation

In interviewer-administered modes, such as CAPI and CATI (computer-assisted telephone interviewing), it is common that respondents are asked to name their job title and describe their occupation in an open-ended format. Interviewers are trained to probe for as much information

(e.g., by means of follow-up questions) as is necessary to enable a 4-digit ISCO level classification (e.g., Belloni et al., 2014; Conrad et al., 2016; Schierholz et al., 2018). The reason for the open format is that it is hardly possible to capture all possible answers and provide them in a list of predefined options or to narrow them down (Züll, 2016). In self-administered surveys, retaining the format of an open-ended question is accompanied by post hoc increased costs and resources (i.e., coding time; Olson et al., 2019). In addition, higher item non-responses or too vague answers that are unlikely to allow for such detailed coding afterwards are to be expected. However, in self-administered mixed-mode surveys such as the GESIS Panel, where respondents are asked about their occupation in an open format, the number of non-codable answers is not particularly high (e.g., GESIS Panel: 5.9–7.1% of the answers; personal communication). Previous research indicates that the length of an occupation description is not a strong determinant of post-hoc codability (Conrad et al., 2016; Massing et al., 2019).

If the data quality provided by open-ended answers proves to be too low, the open-ended format can be replaced or supplemented with four alternatives: First, a series of successive predefined closed questions can be asked to capture details on the occupation of the respondent. For instance, these questions can be used to capture relevant dimensions for coding, such as whether a respondent has personnel responsibility or an executive function. These questions can also be used to determine the economic sector, with questions being displayed via skip-pattern categories, which continuously narrow the occupational field until the exact occupation remains (Olson et al., 2019). Second, search trees can be implemented, which gradually narrow down the specific occupation. Third, so-called look-up databases can be used, which search for matching occupation titles in a deposited databank while respondents are typing. Moreover, for 3MC settings, coding databases covering occupational titles exist that facilitate the development of closed questions (see also Tijdens, 2014, 2015).

A fourth possibility is using a sequential mixed-mode approach. In a study reported by Peycheva et al. (2021), respondents were asked to self-code their occupation in the initially offered web mode using a look-up database. Respondents entered key words, and the coding tool searched for relevant job titles, offering a list of corresponding codes. If the look-up method was not successful in identifying a code, interviewers were tasked with identifying the relevant occupation code in follow-up telephone or face-to-face interviews, employing an in-interview code assignment. Results showed that respondents' ethnic background, university participation, and cohabitation status affected their successful use of the look-up database, indicating that self-coding is problematic for certain respondent groups. However, the interviewer-administered follow-ups indicated that interviewer characteristics, such as interviewers' sex, age, and years of interviewing experience, strongly impacted the likelihood of assigning an occupation code during the interview. Thus, both self- and interviewer-administered coding have caveats, indicating that a multi-mode occupation assessment for all respondents may be the optimal route.

To remain consistent between survey modes, for the EWCS, we recommend using the same question text in the CAWI mode for the job title and job activity questions as in the CAPI mode, requiring that respondents enter their answers into open-ended text fields. Fill-in instructions should be included that name examples of the level of detail required, the overarching goal being to obtain answers that allow for an accurate identification of the occupation at the 4-digit ISCO level.

Industry

Similar issues apply to the respondent's industry and subsequent "Nomenclature of Economic Activities" (NACE classification). Previous work on the industry question in the web-based WageIndicator Survey has shown that respondents tend to skip the question about the industry, presumably because it is deemed too cognitively demanding (https://www.surveycodings.org/industry). Within the project "Synergies for Europe's Research Infrastructures in the Social Sciences" (SERISS), an occupation-based sector prediction has been developed, which provides respondents with a limited set of industries most likely for their occupation (https://seriss.eu/resources/deliverables/).

In the EWCS, respondents are asked to describe the main activity of the organization they work for, with the goal of coding at a 2- or 3-digit NACE level. In the CAPI format, the question text was kept relatively simple but also vague ("What is the main activity of the company or organisation where you work?"), and interviewers were trained to obtain the relevant information. In the online adaptation, we recommend wording the question text more precisely ("What sector of business or industry is the company or organisation where you work mainly active in?") and adding instructions that explain the level of detail required and giving examples (e.g., "software development (not: IT), manufacture of cutlery (not: factory)").

## Gender mainstreaming

As a final aspect, questionnaire adaption must consider gender mainstreaming. This means that the adaptation (and all translations thereof, see also the section 'Questionnaire translation') must consider how a survey question best addresses female and male respondents alike. The EWCS strives towards a gender-sensitive rather than a gender-neutral approach (Nicot & Houtman, 2006). The American Psychological Association (APA) gives guidance on how to write about people and their personal characteristics without bias and about gender with inclusivity and respect (APA, 2020). The minimum requirement for a gender-sensitive language in surveys is presenting both the female and male form where applicable (e.g., for occupations such as "waitress" or "waiter").

References to gender can occur in a survey in several ways: when a question asks about the respondent's gender, in other questions about respondent characteristics (e.g., a question about the respondent's work status may be gender-aligned if the respondent's gender is known or gender-inclusive if it is not known), and in questions that use gender-specific nouns or pronouns (e.g., references to occupations or references to the respondent's boss, colleagues, clients, or employer).

Unlike other self-administered modes, such as paper-and-pencil, online questionnaires benefit from the possibility that they can be programmed to show only the gender-aligned version in such instances (e.g., showing only "waitress" as an occupation to female respondents and only "waiter" to male respondents). This, however, results in additional translation and programming challenges. Moreover, the question on gender is generally asked towards the end of a survey. Programming gender-aligned language in a survey requires changing this order and collecting information on gender at the beginning.

Finally, both the possibilities and necessity of gender-specific modifications differ markedly by language. Provided that it is possible in the language, it is also conceivable to avoid gendered endings (e.g., use "police officer" instead of "policeman" in occupational titles) and to use inclusive terms (e.g., humans, people) as recommended by the APA (2020).

Gender identity

Eurofound considers the online questionnaire to be the beginning of a new trend regarding gender sensitivity. The adapted questions are intended to capture gender identity in a consistent way and include a non-binary option. Questions about the gender of the respondent and the proxy should ideally address all respondents who identify as male or female in the respective form. The question text asks how the respondents would describe themselves or another member of their household. The non-binary response option "other" should be chosen to be as all-encompassing and inclusive as possible.

There are three instances where the EWCS questionnaire records the sex or gender of a person: the immediate boss, the respondent, and other household member(s). Each of these three questions is followed by questions about the respective person. We recommend wording these subsequent questions in a gender-aligned way. For instance, the items about the immediate boss should be worded in a male (i.e., "He respects me as a person.") or female form (i.e., "She respects me as a person.") if the boss was classified as male or female and if a gender-aligned approach is possible in the respective language. If respondents choose the non-binary option "other" or leave the question unanswered, the items should contain both forms ("He/She respects me as a person.").

Following the question about the gender of the respondent or proxy, the question about the respective work status should also be gender-aligned. This simplifies the question text in languages that use different male and female forms for terms such as employee, employer, or homemaker. Moreover, it also allows countries that offer different child-leave policies for men and women to use the respective terms. In addition, the question about the proxy's relationship to the respondent profits strongly from gender alignment, as the response options become more specific. For instance, the response option "My child or the son/daughter of my cohabiting partner" should be presented as "My son or the son of my cohabiting partner" for male proxies, as "My daughter or the daughter of my cohabiting partner" for female proxies, and neutrally as "My child or the child of my cohabiting partner" when the gender is declared as "other" or unknown.

# Aspects of questionnaire administration for successive interviews and modularisation

This section is divided into three parts. First, it addresses issues related to the length and different possibilities to shorten the questionnaire. Second, recruitment procedures are presented. Third, the handling of missing data is addressed.

## Length of the questionnaire

Another important aspect of questionnaire administration when moving to a self-administered online mode relates to the appropriate length of the questionnaire.

Interview duration is determined by both survey and respondent characteristics. In terms of question characteristics, the number of clauses, words within clauses, and answer categories impact response times (Yan & Tourangeau, 2008). However, the larger share of the variation in interview duration is explained by the characteristics of the respondents (Gummer & Rossmann, 2015). For instance, respondents with higher cognitive abilities generally need less time to fill in surveys; highly motivated respondents (i.e., respondents with higher topic interest) spend more time answering complex survey questions, such as open-ended questions. Respondents from opt-in online panels are generally more experienced at answering survey questions, reducing survey time (Gummer & Rossmann, 2015).

Prior research indicates that the effect of switching from a CAPI to an online questionnaire on interview duration is minimal. For the EVS, the average interview duration for a face-to-face survey was 59 minutes, and 55 minutes for a web survey of the same content (Wolf et al., 2021).

The ideal length of an online survey is between ten and 15 min and the maximum survey length between 20 and 28 min (e.g., Revilla & Höhne, 2020). Therefore, it is common that longitudinal surveys shorten their questionnaire when they transition from an interviewer-administered to a self-administered mode (Olson et al., 2019). Because especially survey questions asked at the end of a questionnaire receive less attention than those asked at the beginning – and measurement error increases (Peytchev & Peytcheva, 2017) – it is necessary to maintain respondents' motivation and attention throughout the whole survey by keeping a questionnaire short (Neuert, 2021). Consequently, respondents' effort and fatigue are reduced, on the one hand, and response rates and data quality increase, on the other hand (e.g., Littvay, 2009).

However, there are several things to consider. One option is to reduce the questionnaire for all respondents from, for example, 45 min to 30 min and only cover the most relevant questions. This procedure is preferred for surveys where large sample sizes are not available and where programming-intensive and time-consuming post-hoc estimation of missing data is too costly or complex (see, e.g., Peytchev et al., 2020). In such a case, it is advisable to give respondents the option to pause participation and continue later. An option to keep a survey short is to split the questionnaire across waves (panel design): After completing the initial survey, respondents are invited to a further round of questions. The disadvantage of this approach is that it increases fieldwork costs, and because not all respondents will participate in multiple surveys, incomplete data occur in many cases (Wolf et al., 2021; see also the following section).

## Planned missingness design

Another option is to use a so-called planned missingness design. There a several ways to implement a planned missingness design (for an overview, see Little & Rhemtulla, 2013): The first is the split questionnaire (also called multi-form) design, the second is the 2-method measurement design, and the third the wave missing design.

In the **split questionnaire design**, for example, two-thirds or half of the entire questionnaire is completed in each form, resulting in the post-hoc estimated scores and actual assessed scores being very similar and comparable (e.g., Smits & Vorst, 2007). There are two ways to design a split questionnaire. The first option is to divide a, for example, 45-min questionnaire into several shorter ones such as three 15-min questionnaires. Multi-item instruments are not split in the process, but only a portion of respondents receive them. This option should be selected primarily if the composite score of the instrument is important for further analysis (Raghunathan & Grizzle, 1995). In other cases, however, the second option is preferable.

The second option is that respondents receive a core questionnaire (e.g., asking for sociodemographic characteristics or other variables with high predictive ability) and a specific item subset of each multi-item instrument, to which participants are randomly assigned. Thereby, it is recommended that item combinations be chosen so that every conceivable pair of questions is captured in at least one form (Peytchev & Peytcheva, 2017). Furthermore, each facet of a construct should be represented in each form as well as a few items in all forms (Moore et al., 2020). From a statistical point of view, the variables across the subsets should be correlated with each other. This ensures that the observed elements have prediction power for the missing elements. Within a subset, the variables should not be highly correlated. The purpose of this is to ensure that responses are not acquired to an item that does not provide any additional information (Raghunathan & Grizzle, 1995).

Next, it must be decided whether respondents receive all modules with some time in between, whether they receive only one module, or whether they receive a core questionnaire and, for example, one further module. Presenting all split forms to the participants with some time in between has been shown to result in higher dropout rates, irrespective of the number of modules (i.e., three vs. 10), compared to the condition in which participants received the entire questionnaire at once (Toepoel & Lugtig, 2018). However, even though the respondents also received all parts of the questionnaire, just not the entire questionnaire at once, there were some considerable benefits of split questionnaires: fewer item missing and satisficing response behaviour, better survey evaluation (Toepoel & Lugtig, 2018), and lower measurement error (Peytchev & Peytcheva, 2017). Moreover, there was no impact on responding tendencies and data quality (Toepoel & Lugtig, 2018).

If not all modules are presented to each respondent, it must be assured that respondents are randomly assigned to the modules (missing completely at random [MCAR]) to yield unbiased estimates in modern missing data estimation and to overcome the power that is lost due to the reduced data associated with the design (Rhemtulla & Little, 2012). Simulation studies have demonstrated that the higher the correlations between the variables across different parts/forms/modules/splits, the less efficiency loss this approach entails after data collection when imputing missing data (Raghunathan & Grizzle, 1995).

In the **2-method measurement design**, most (or all) respondents receive a cheaper, less valid measurement of a construct, while a few respondents receive the cheaper version and additionally a

more expensive, highly valid measurement, leading to lower standard errors and a higher sample size for chosen constructs (e.g., Graham et al., 2006). The 2-method measurement design is suitable for situations in which there is a choice between two different measures of a construct that differ significantly in terms of cost and effort. In the social sciences, this is usually the case whenever self-administered measures show weaknesses in terms of validity (e.g., when asking adolescents about their smoking behaviour). In this case, objective, more expensive measures such as biological markers or, depending on the construct, face-to-face interviews may be used for a subset of the sample (Graham et al., 2006; Little & Rhemtulla, 2013).

In the **wave missing design**, participants omit one or more complete measurement occasions (Little & Rhemtulla, 2013). As an example, in a design with monthly measurements across 12 months, each participant could be assigned two missing measurements. While the first two planned missingness designs can be applied in cross-sectional and longitudinal studies, the wave missing design can only be used in longitudinal studies (Little & Rhemtulla, 2013; Wood et al., 2019). Indeed, many longitudinal studies already employ this design: Multiple age cohorts are assessed, while making use of data from one cohort to impute values for another (Duncan et al., 2013; Little, 2013; Wood et al., 2019).

Which design is to be preferred must be decided on a survey-specific basis. For longitudinal studies, a split-form missing design is recommended in preference to a wave missing design, because it has been shown to be more effective in capturing the multilevel model (Wood et al., 2019).

To conclude, planned missingness design has several advantages: First, the number of measures collected from each respondent can be increased while decreasing respondents' burden and fatigue. Second, due to lower survey costs, researchers can use larger sample sizes, thus, improving data quality and validity (Rhemtulla & Little, 2012). Third, the likelihood of non-response and missing values that are not random (MNAR) decreases, leading to multiple imputation and maximum likelihood methods being more valid (e.g., Imbriano, 2018). That is, the least efficiency loss can be achieved with the planned missingness design.

In a recent mode experiment within the German part of the EVS 2017/2018, the average length of the face-to-face interviews was 59 minutes, whereas the survey length was reduced to 38 minutes for the web interview using a matrix design (Wolf et al., 2021).

# Recruitment procedures

There are two cases in which it is necessary to think about recruitment procedures: in longitudinal surveys with multiple waves and in split questionnaire designs if all respondents shall receive multiple or all modules. In these cases, it must be ensured that the respondents participate in all waves/parts of the survey.

## Survey retention

There are several strategies for retaining participants in surveys, including incentives, sending newsletter or reminder letters, providing feedback options or summaries of study results, or offering alternative methods of data collection (e.g., Watson & Wooden, 2009). In their methodological report, the German Family Panel (Panel Analysis of Intimate Relationships and Family Dynamics [pairfam], continued as Family Research and Demographic Analysis [FReDA-pairfam] from 2022 onwards) describes sending an information brochure and updating the project homepage with

important information, for example, on data protection or study results, as strategies. They also established a hotline for respondents and allowed respondents to update their address on the website to avoid panel attrition (Brüderl et al., 2021).

Willingness to participate in further survey waves also depends on the way surveys are designed and administered (Watson & Wooden, 2009). Particularly in web-based surveys, the building of trust is an important factor. Participants must be given the feeling that the survey is legitimate (Dillman et al., 2014) and trustworthy and that their personal data will be protected and handled with confidence.

Additionally, cooperation has been found to be influenced by the use of *incentives* (Watson & Wooden, 2009). Survey panels (e.g., the GESIS Panel or the Longitudinal Internet studies for the Social Sciences [LISS] Panel) use incentives (monetary or non-monetary) as a strategy for panel maintenance. To increase response rates and retention in panel surveys or subsequent modules, it is recommended to provide the greater share of the incentive directly at the beginning (i.e., when assessing the first survey and when asking for participation in a follow-up survey/wave; Peytchev et al., 2020). For panel surveys or surveys with multiple parts, a strategy might be to provide an additional incentive for participation in case respondents have not participated in a consecutive (part of the) survey or to incentivise respondents with lower likelihoods of participating in later parts of the survey (an approach currently tested within FReDA; Schneider et al., 2021). For the 3MC context, it is important to note that there may be different country-specific regulations regarding incentives (Sommer, 2017).

## Sociodemographic variables

Sociodemographic variables can also be used as determinants to predict and compensate for nonparticipation (Radler & Ryff, 2010). For example, it is more difficult to retain individuals with lower levels of education, with lower health status, men, and the unmarried in longitudinal studies (Radler & Ryff, 2010). Considering such information collected in an initial survey, helps match the recruitment strategy to respondents with certain characteristics, such as higher incentives for those with lower levels of education (tailored design, see Dillman et al., 2014; Radler & Ryff, 2010). Previous research has shown that an incentivization might increase response rates for often under-represented groups in surveys (e.g., younger respondents, respondents from minority ethnic groups; Singer et al., 1999) However, how strong the impact of the incentive will be and how respondents react to it, is usually not known. Moreover, incentivization may be seen as ethically questionable, in particular if respondents receive differential incentives.

## Re-contact data

According to the American Association for Public Opinion Research (AAPOR; Olson et al., 2019) report on transitions from telephone surveys to self-administered and mixed-mode surveys, the most common mode used to recruit respondents for surveys that have transitioned to self-administered or mixed modes is by mail. Mail is also often used to provide a URL and access code for the online survey. Moreover, invitations by mail allow to include study-relevant information materials as well as prepaid incentives. If email addresses or mobile phone numbers are available and consent to re-contact has been obtained, recruitment via email or text message is also possible. It can be assumed that these are also the most important channels for re-contacting respondents

(Olson et al., 2019). It is advisable to use multiple (re-)contact modes, regardless of the mode of the actual survey (Dillman et al.,2014).

Re-contact data in terms of "nonresponse follow-ups:" If all efforts to recruit respondents or to motivate them to participate in multiple survey waves fail, it is possible to collect so-called re-contact data after the deadline to participate in a module or after the survey has closed. That is, non-participants are contacted again and asked to respond to a small questionnaire or to provide some information about themselves and the reasons for their non-participation (e.g., Karvanen et al., 2016; Kopra et al., 2018). Respondents who comply with this re-contact request can be used as a proxy for all non-participants and help reduce selection bias by adjusting for this non-participation bias in the population estimates (e.g., Karvanen et al., 2016; Kopra et al., 2018). Moreover, consideration may be given to recording the intention to drop out from the survey for respondents who participate in the survey (Graham, 2009). However, it needs to be noted that re-contacting might only be permitted if respondents have not rejected participation in the study in general.

## Handling of missing data

When applying a planned missingness design, one must handle missing data (i.e., the scores of items not unadministered) after data collection. There are two different methods of appropriate missing data procedures: multiple imputation (e.g., Kaplan & Su, 2016; Raghunathan & Grizzle, 1995) and (full information) maximum likelihood (for an overview of the vast number of different estimators for both methods, see Little & Rhemtulla, 2013). Both methods can produce similar results when implemented in comparable ways (e.g., Collins et al., 2001). The additional inclusion of auxiliary variables strongly related to the items of interest is recommended to improve missing data estimation, for example, by reducing estimation bias due to MNAR missingness (e.g., due to further unplanned missing data in addition to the planned missing data) and compensating for lost power (e.g., Collins et al., 2001; Graham, 2009). Because it is easier to include auxiliary variables in multiple imputation, this method is preferable compared to maximum likelihood estimation (Collins et al., 2001).

To minimize the loss of information in planned missingness designs, the optimal pattern of allowed and planned missingness should be determined in advance. In general, a high number of split questionnaires or questionnaire forms is recommended to obtain "the greatest uniformity in the information matrix" (Littvay, 2009, p. 112). At the same time, it is essential to ensure that there are common observations and enough participants so that each combination can be assigned. That is, the number of splits must be chosen so that there are enough participants per split questionnaire form. Search algorithms can help to find the optimal split solution (i.e., the way the instruments and variables are distributed among the modules/splits/forms) and to indicate the required sample sizes (e.g., Ioannidis et al., 2016).

# Aspects of questionnaire administration in multi-language questionnaires

Generally, there are some challenges with any 3MC survey. The first issue concerns the measurement invariance of the survey across languages and countries. It has been shown that it is not so much the question mode and format that are not equivalent in different countries, but rather the non-equivalence is due to not comparable constructs, translation errors and the different languages and cultures, even in, for example, three neighbouring European countries (Germany, France, Switzerland; Roberts et al., 2020). Furthermore, the meaning and understanding of some questions or features may vary in different cultures (Hibben & de Jong, 2020), including design and visual components, which are culturally embedded, such as unlucky numbers (e.g., 4 in China and Japan, 17 in Italy and Brazil, 39 in Afghanistan), colours with negative or political symbolism, and culturally specific symbols/icons, or with which some cultures are (un)familiar (e.g., thermometers or response scales with faces/smileys; e.g., Hansen et al., 2016; Mohler et al., 2016).

Apart from that, in some cultures, it is necessary to make adaptions that, first, improve understanding to maintain conceptual coverage, second, reduce difficulty to maintain the same level of difficulty across populations, or third, conform to cultural discourse conventions or sensitivities (for an overview of common forms of adaptation required, see Mohler et al., 2016). Therefore, it is of great importance to (pre)test both the source and translated and adapted questionnaires (inclusive visual designs) to ensure both measurement invariance between the language versions and validity and cross-cultural comparability (Goerman & Caspar, 2010; Hibben & de Jong, 2020; Schoua-Glusberg & Villar, 2014).

## Questionnaire translation

*Good* questionnaire translations need to meet specifications and requirements as formulated for a specific translation project, such as the needs of the target group, of the survey mode, and of the study with regard to consistency across waves, etc. (Behr, 2018). When a project transitions from a face-to-face survey to a self-administered online survey, new specifications need to make explicit to all translation teams involved what kind of work will be required from them, amongst others:

- When shifting an *existing* instrument from face-to-face to the online mode, in principle, the changes implemented in the source questionnaire should equally be implemented in the translations. Beyond these source changes, however, the entire translation of a questionnaire needs to be carefully checked for further wordings that are not suitable for the online mode.
- With regard to *gender mainstreaming,* teams need to decide how gender should be addressed, for example, by using separate programming for male or female versions, using slashes or brackets, for instance, *satisfait(e)* (Behr, 2018). In the online mode, reliance on an interviewer to use the appropriate wording is no longer possible.

Transitioning to the online mode means also that new programming options may become available, for instance, placeholders (or so-called fills). However, one needs to be careful of what these may mean for multilingual implementation. Placeholders that work in the English language may

challenging or not be suitable at all for multilingual implementation (e.g., fills for date or tense may result in inadequate linguistic structure; Behr, in press; Wang et al., 2017).

Online surveys require an appropriate translation of interface and navigation elements; these translations need to be produced with usability in mind so as not to confuse respondents in their attempt to access and navigate through the survey (Wang et al., 2017).

If the online survey is based on one programming (i.e., a kind of template for the source questionnaire) that serves as the basis for translations teams, the topic of potential country-specific adaptations (e.g., country-specific questions, country-specific formats such as date formats) should be addressed early on during questionnaire development with question developers and survey programmers (Behr, in press).

In general, recommendations for technical instrument design should be heeded when transitioning to the self-administered online mode. Hansen et al. (2016) provide a valuable source of inspiration, including pointers to differences in multicultural implementation, for example, different time formats, different associations with colours, different scripts, and different text directionality.

When changing the survey mode, the attention typically paid to translation should not be reduced. This means, also for self-administered online surveys, best practice procedures such as the team-based interdisciplinary TRAPD model should be implemented. In TRAPD, two translators produce independent translations of a questionnaire (*Translation*); subsequently, they meet with substantive and survey experts to reconcile these translations (*Review*); the final translations are signed off (*Adjudication*) before a pretest (*Pretest*) is implemented. The entire procedure is documented (*Documentation*; Harkness, 2003). Team-based translations apply in particular to new items in the online setting and to items with major changes due to the mode switch. Smaller wording changes may not require double translations, even though they should also be carefully vetted by different parties.

Technical pretests – beyond substantive pretests – on a questionnaire translation in the online mode should be implemented with a threefold aim: a) *linguistic* testing (e.g., is all correctly translated, including error messages; do the different language scripts display correctly); b) *interface* testing (e.g., is all text visible, also in drop-box design or error messages); and c) *functionality* testing (e.g., does the instrument work as intended, including correct filtering and sending respondents through the survey). Compatibility testing – that is, testing whether the translated instruments work with different devices, web browsers, etc. – should equally be conducted (Behr, in press) to ensure that the target group can correctly see and read the survey.

Last but not last, (multilingual) survey entry needs to be considered, too. Compared to interviewer-based surveys, an advantage of web-based surveys is that they can be programmed to allow participants to select their preferred language (Olson et al., 2019) – preferably at the beginning of the survey so that all subsequent questions can be displayed in the selected language. A distinction has to be made, on the one hand, between a survey that automatically recognizes which country someone is accessing from and automatically offers the corresponding language, and, on the other hand, and a language selection by respondents within a multilingual country. For example, a survey in Switzerland should always be provided in German, French, and Italian; languages, from which the survey respondents should then be able to choose their preferred survey language. The way how respondents navigate to the correct survey entry page needs to be carefully considered, too. Sha et

al. (2018) describe, for instance, how limited English-language speakers in the US can optimally access translated surveys in the US context.

## Infrastructural burdens

Administration of web-based surveys can come with administrative hurdles and infrastructural constraints, especially in non-Western countries or rural areas. Amongst these are insufficient internet connection or penetration, lack of suitable software, internet censorship, lack of computer skills, lack of expertise to conduct and program computerized surveys (e.g., de Jong, 2016; Hansen et al., 2016). Therefore, it is necessary to know in advance the technological infrastructure and inappropriate or unauthorized survey topics of each target country to ensure that participants can easily access the survey link (de Jong, 2016).

# Conclusion

To sum up, when moving from an interviewer-based to a self-administered web-based large-scale survey, a number of issues need to be carefully considered, especially for 3MC surveys that deal with multiple languages and cultures. Among other things, decisions and adjustments must be made regarding (a) response options and formats (i.e., open-ended numerical questions, multiple-response questions, and the inclusion of non-substantive response options for certain questions, such as sensitive or knowledge questions); b) the presentation of additional information such as definitions and explanations (for all respondents or only upon request); c) question wording and instructions (general, but also gender- and culturally-sensitive language); d) the layout of the questionnaire (which should be compatible and easy to answer with all possible types of devices); e) the handling of specific survey questions (e.g., occupation and industry); (f) length of the questionnaire (i.e., shortening, split questionnaire design, or planned missingness design); and (g) recruitment strategies (including re-contact data).

For all these aspects, we have provided general and EWCS-specific recommendations and examples based on findings from previous literature as well as experience from survey practitioners and methodological experts – always taking into account the greatest possible equivalence to the CAPI questionnaire.

Cross-cultural cognitive interviews and web probing were used to examine how the EWCS CAPI questionnaire can best be transitioned to a self-administered online questionnaire, verifying many of the question design decisions from the literature (see Final Report). However, questions specific to the EWCS require further attention and research in a self-administered or multi-mode questionnaire. Both the existing literature and the findings from cognitive pretesting indicate that transitioning questions on occupation and sector may lead to a systematic bias in reporting, with respondents with lower education and/or literacy being less likely to provide codable responses. We recommend research on two-step study designs in which respondents provide their responses in a self-administered format in one step and with the help of an interviewer in the other step. This can be done using open-ended questions in both steps and/or testing combination of open-ended questions, search trees and look-up databases.

The next task is to determine whether the recommended adaptations for the EWCS questionnaire are optimal and produce the intended results, ideally using mixed-method data collection. In general, applying mixed-mode designs is always a good strategy because they shorten and facilitate fieldwork, lead to higher response rates, and better represent the general population of a country (e.g., Wolf et al., 2021). To ensure high data quality and equivalence in mixed-mode designs, it is advisable to control for selection, demographic variables, and responses from previous waves (Hox et al., 2015). For the EWCS as a 3CM survey, possible mode effects should be examined in all countries and language versions. Internet penetration, computer skills and literacy may vary across countries and respondent groups. Therefore, mixed-mode data collection and analysis are necessary to examine possible effects of the mode on response behaviour.

Finally, the effect of a planned missingness design on data quality, such as reliability and validity, should be examined. Currently, there is little empirical data on this from large-scale surveys (for first results from the EVS, see Wolf et al., 2021). Research projects on the statistical modelling of planned missingness are ongoing (i.e., DFG project 407454818 “Modular Questionnaire Designs for Social

Surveys: Statistical Modelling of Design Missingness"). For the EWCS, we recommend an in-depth questionnaire evaluation under the aspect of modularization prior to choosing a planned missingness design. Finally, an empirical study should compare the data quality of a full online questionnaire in comparison to a planned missingness design.

# References

Al Baghal, T. and Lynn, P. (2015), 'Using motivational statements in web-instrument design to reduce item-missing rates in a mixed-mode context', *Public Opinion Quarterly*, Vol 79, No. 2, pp. 568–579.

Antoun, C., Conrad, F. G., Couper, M. P. and West, B. T. (2019), 'Simultaneous estimation of multiple sources of error in a smartphone-based survey', *Journal of Survey Statistics and Methodology,* Vol 7, No. 1, pp. 93–117.

Antoun, C., Katz, J., Argueta, J. and Wang, L. (2018), 'Design heuristics for effective smartphone questionnaires', *Social Science Computer Review,* Vol 36, No. 5, pp. 557–574.

American Psychological Association (2020), *Publication manual of the American Psychological Association* (7th ed.), Washington, DC.

Behr, D. (2018), 'Translating questionnaires for cross-national surveys: A description of a genre and its particularities based on the ISO 17100 categorization of translator competences', *Translation & Interpreting*, Vol 10, No. 2, pp. 5–20.

Behr, D. (in press), 'Computer-assisted migration research: What we can learn from the field of software localization for source questionnaire design and translation', in Pötzschke, S. and Rinken, S. (eds.), *Migration research in a digitized world: Using innovative technology to tackle methodological challenges*, IMISCOE Research Series, Springer.

Belloni, M., Brugiavini, A., Maschi, E. and Tijdens, K. (2014), 'Measurement error in occupational coding: An analysis on SHARE data', University Ca' Foscari of Venice, Deptartment of Economics Working Paper Series No. 24/WP/2014.

Bishop, G. F., Hippler, H.-J., Schwarz, N., Strack, F., Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholls, W. L. and Waksberg, J. (1998), 'A comparison of response effects in self-administered and telephone surveys', *Telephone Survey Methodology*, Wiley, New York, NY.

Blasius, J. (2012), 'Comparing ranking techniques in web surveys', *Field Methods*, Vol 24, No. 4, pp. 382–398.

Breton, C., Cutler, F., Lachance, S. and 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,* Vol 50, No. 4, pp. 1005–1036.

Brüderl, J., Schmiedeberg, C., Castiglioni, L., Becker, O. A., Buhr, P., Fuß, D., Ludwig, V., Schröder, J. and Schumann, N. (2021), 'The German Family Panel: Study design and cumulated field report (Waves 1 to 12)', Pairfam Technical Paper 01.

Buskirk, T. D. and Andrus, C. H. (2014), 'Making mobile browser surveys smarter: Results from a randomized experiment comparing online surveys completed via computer or smartphone', *Field Methods*, Vol 26, No. 4, pp. 322–342.

Callegaro, M., Murakami, M. H., Tepman, Z. and Henderson, V. (2015), 'Yes-no answers versus check-all in self-administered modes', *International Journal of Market Research*, Vol 57, No. 2, pp. 203–223.

Cernat, A., Couper, M. P. and Ofstedal, M. B. (2016), 'Estimation of mode effects in the health and retirement study using measurement models', *Journal of Survey Statistics and Methodology*, Vol 4, No. 4, pp. 501–524.

Chambers, C. T., Giesbrecht, K., Craig, K. D., Bennett, S. M. and Huntsman, E. (1999), 'A comparison of faces scales for the measurement of pediatric pain: Children's and parents' ratings', *Pain*, Vol 83, No. 1, pp. 25–35.

Chang, L. and Krosnick, J. A. (2010), 'Comparing oral interviewing with self-administered computerized questionnaires. An experiment', *Public Opinion Quarterly*, Vol 74, No. 1, pp. 154–167.

Christian, L. M. and Dillman, D. A. (2004), 'The influence of symbolic and graphical language manipulations on answers to paper self-administered questionnaires', *Public Opinion Quarterly,* Vol 68, No. 1, pp. 57–80.

Christian, L. M., Dillman, D. A. and 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*, Vol 71, No. 1, pp. 113–125.

Christian, L. M., Dillman, D. A. and Smyth, J. D. (2008), 'The effects of mode and format on answers to scalar questions in telephone and web surveys', in Lepkowski, J., Tucker, C., Brick, M., DeLeeuw, E., Japec, L., Lavrakas, P., Link, M. and Sangster, R. (eds.), *Advances in telephone survey methodology*, Wiley, Hoboken, NJ, pp. 250–275.

Chyung, S. Y., Kennedy, M. and Campbell, I. (2018), 'Evidence-based survey design: The use of ascending or descending order of Likert-type response options', *Performance Improvement*, Vol 57, No. 9, pp. 9–16.

Clifford, S. and Jerit, J. (2016), 'Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions', *Public Opinion Quarterly*, Vol 80, No. 4, pp. 858–887.

Collins, L. M., Schafer, J. L. and Kam, C.-M. (2001), 'A comparison of inclusive and restrictive strategies in modern missing data procedures', *Psychological Methods,* Vol 6, No. 4, pp. 330–351.

Conrad, F. G., Couper, M. P. and Sakshaug, J. W. (2016), 'Classifying open-ended reports: Factors affecting the reliability of occupation codes', *Journal of Official Statistics*, Vol 32, No. 1, pp. 75–92.

Conrad F. G., Couper M. P., Tourangeau R. and Galesic M. (2005), 'Interactive feedback can improve quality of responses in web surveys', *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 3835–3840.

Conrad, F. G., Couper, M. P., Tourangeau, R. and Peytchev, A. (2006), 'Use and non-use of clarification features in web surveys', *Journal of Official Statistics,* Vol 22, No. 2, pp. 245–269.

Couper, M. P., Kennedy, C., Conrad, F. G. and Tourangeau, R. (2011), 'Designing input fields for non-narrative open-ended responses in web surveys', *Journal of Official Statistics*, Vol 27, No. 1, pp. 65–85.

Couper, M. P., Tourangeau, R., Conrad, F. G., and Zhang, C. (2013), 'The Design of Grids in Web Surveys', *Social Science Computer Review*, Vol 31, No. 3, pp. 322–345.

de Jong, J. (2016), 'Self-administered surveys', *Guidelines for best practice in cross-cultural surveys, Survey Research Center*, Institute for Social Research, University of Michigan, Ann Arbor, MI, available at https://ccsg.isr.umich.edu/chapters/data-collection/self-administered-surveys/

DeCastellarnau, A. (2018), 'A classification of response scale characteristics that affect data quality: a literature review', *Quality & Quantity*, Vol. 52, pp. 1523–1559.

DeRouvray, C. and Couper, M. P. (2002), 'Designing a strategy for reducing "no opinion" responses in web-based surveys', *Social Science Computer Review*, Vol 20, No. 1, pp. 3–9.

Dillman, D. A. (2017), 'The promise and challenge of pushing respondents to the web in mixed-mode surveys', *Survey Methodology,* Vol 43, No. 1, pp. 3–31.

Dillman, D. A. and Christian, L. M. (2005), 'Survey mode as a source of instability in responses across surveys', *Field Methods,* Vol 17, No. 1, pp. 30–52.

Dillman, D. A. and Edwards, M. L. (2016), 'Designing a mixed-mode survey', in Wolf, C., Joye, D., Smith, T. W. and Fu, Y. (eds.), *The Sage handbook of survey methodology*, Sage, London, pp. 255–268.

Dillman, D. A. and Smyth, J. D. (2007), 'Design effects in the transition to web-based surveys', *American Journal of Preventive Medicine*, Vol 32, No. 5, pp. S90–S96.

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

Domnich, A., Panatto, D., Signori, A., Bragazzi, N. L., Cristina, M. L., Amicizia, D. and 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*, Vol 17, No. 4, e94.

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

Duffy, B., Smith, K., Terhanian, G. and Bremer, J. (2005), 'Comparing data from online and face-to-face surveys', *International Journal of Market Research*, Vol 47, No. 6, pp. 615–639.

Duncan, T. E., Duncan, S. C. and Strycker, L. A. (2013), *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*, New York, NY.

Emde, M. and Fuchs, M. (2012), 'Exploring animated faces scales in web surveys: Drawbacks and prospects', *Survey Practice*, Vol 5, No. 1.

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*, Vol 34, No. 2, pp. 244–254.

Funke, F., Reips, U. D. and Thomas, R. K. (2011), 'Sliders for the smart: Type of rating scale on the web interacts with educational level', *Social Science Computer Review*, Vol 29, No. 2, pp. 221–231.

Galesic, M., Tourangeau, R., Couper, M. P. and Conrad, F. G. (2007), 'Using change to improve navigation in grid questions', Paper presented at the General Online Research Conference (GOR 07), Leipzig, Germany.

Galesic, M., Tourangeau, R., Couper, M. P. and Conrad, F. G. (2008), 'Eye-tracking data: New insights on response order effects and other cognitive shortcuts in survey responding', *Public Opinion Quarterly*, Vol 72, No. 5, pp. 892–913.

Geisen, E. and Bergstrom, J. R. (2017), *Usability testing for survey research*, Boston, MA.

Goerman, P. and Caspar, R. (2010), 'A preferred approach for the cognitive testing of translated materials: Testing the source version as a basis for comparison', *International Journal of Social Research Methodology*, Vol 13, No. 4, pp. 303–316.

Grady, R. H., Greenspan, R. L. and Liu, M. (2019), 'What is the best size for matrix-style questions in online surveys?', *Social Science Computer Review*, Vol 37, No. 3, pp. 435–445.

Graham, J. W. (2009), 'Missing data analysis: Making it work in the real world', *Annual Review of Psychology*, Vol 60, pp. 549–576.

Graham, J. W., Taylor, B. J., Olchowski, A. E. and Cumsille, P. E. (2006), 'Planned missing data designs in psychological research', *Psychological Methods*, Vol 11, No. 4, pp. 323–343.

Gummer, T. and Rossmann, J. (2015), 'Explaining interview duration in web surveys: A multilevel approach', *Social Science Computer Review*, Vol 33, No. 2, pp. 217–234.

Gummer, T., Vogel, V., Kunz, T. and Roßmann, J. (2020), 'Let's put a smile on that scale: Findings from three web survey experiments', *International Journal of Market Research*, Vol 62, No. 1, pp. 18–26.

Hansen, S. E., Lee, H. J., Lin, Y. and McMillan, A. (2016), 'Instrument technical design', *Guidelines for best practice in cross-cultural surveys*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, available at https://ccsg.isr.umich.edu/chapters/instrument-technical-design/

Harkness, J. A. (2003), 'Questionnaire translation', in Harkness, J. A., van de Vijver, F. and Mohler, P. Ph. (eds.), *Cross-cultural survey methods*, John Wiley & Sons, Hoboken, NJ, pp. 35–56.

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*, Vol 21, No. 1, pp. 111–121.

Heerwegh, D. and Loosveldt, G. (2008), 'Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality', *Public Opinion Quarterly*, Vol 72, No. 5, pp. 836–846.

Hibben, K. C. and de Jong, J. (2020), 'Cross-cultural survey guidelines: Pretesting – Cognitive interviewing', *Guidelines for best practice in cross-cultural surveys*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, available at https://ccsg.isr.umich.edu/chapters/pretesting/cognitive-interviewing/

Höhne, J. K. and Lenzner, T. (2015). 'Investigating response order effects in web surveys using eye tracking', *Psihologija*, Vol 48, No. 4, pp. 361–377.

Hox, J. J., De Leeuw, E. D. and Zijlmans, E. A. O. (2015), 'Measurement equivalence in mixed mode surveys', *Frontiers in Psychology*, Vol 6, Article 87.

Imbriano, P. (2018), 'Methods for improving efficiency of planned missing data designs', Doctoral Dissertation.

Ioannidis, E., Merkouris, T., Zhang, L. C., Karlberg, M., Petrakos, M., Reis, F. and Stavropoulos, P. (2016), 'On a modular approach to the design of integrated social surveys', *Journal of Official Statistics*, Vol 32, No. 2, pp. 259–286.

Kaplan, D. and Su, D. (2016), 'On matrix sampling and imputation of context questionnaires with implications for the generation of plausible values in large-scale assessments', *Journal of Educational and Behavioral Statistics*, Vol 41, No. 1, pp. 57–80.

Karvanen, J., Tolonen, H., Härkänen, T., Jousilahti, P. and Kuulasmaa, K. (2016), 'Selection bias was reduced by recontacting nonparticipants', *Journal of Clinical Epidemiology*, Vol 76, pp. 209–217.

Keusch, F. and Yan, T. (2017), 'Web versus mobile web: An experimental study of device effects and self-selection effects', *Social Science Computer Review*, Vol 35, No. 6, pp. 751–769.

Klausch, T., Hox, J. J. and Schouten, B. (2013), 'Measurement effects of survey mode on the equivalence of attitudinal rating scale questions', *Sociological Methods & Research*, Vol 42, No. 3, pp. 227–263.

Kopra, J., Härkänen, T., Tolonen, H., Jousilahti, P., Kuulasmaa, K., Reinikainen, J. and Karvanen, J. (2018), 'Adjusting for selective non-participation with re-contact data in the FINRISK 2012 survey', *Scandinavian Journal of Public Health*, Vol 46, No. 7, pp. 758–766.

Lambert, A. D. and Miller, A.L. (2015), 'Living with smartphones: Does completion device affect survey responses?', *Research in Higher Education*, Vol 56, No. 2, pp. 166–177.

Little, T. D. (2013), *Longitudinal structural equation modeling,* New York, NY.

Little, T. D. and Rhemtulla, M. (2013), 'Planned missing data designs for developmental researchers', *Child Development Perspectives*, Vol 7, No. 4, pp. 199–204.

Littvay, L. (2009), 'Questionnaire design considerations with planned missing data', *Review of Psychology*, Vol 16, No. 2, pp. 103–114.

Liu, M. and Cernat, A. (2018), 'Item-by-item versus matrix questions: A web survey experiment', *Social Science Computer Review*, Vol. 36, No. 6, pp. 690–706.

Liu, M. and Keusch, F. (2017), 'Effects of scale direction on response style of ordinal rating scales', *Journal of Official Statistics*, Vol 33, No 1, pp. 137–154.

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

Massing, N., Wasmer, M., Wolf, C. and Zuell, C. (2019), 'How standardized is occupational coding? A comparison of results from different coding agencies in Germany', *Journal of Official Statistics*, Vol 35, No. 1, pp. 167–187.

Mavletova, A. (2013), 'Data quality in PC and mobile web surveys', *Social Science Computer Review*, Vol 31, No. 6, pp. 725–743.

Mavletova, A. and Couper, M. P. (2013), 'Sensitive topics in PC web and mobile web surveys: Is there a difference?', *Survey Research Methods,* Vol 7, No. 3, pp. 191–205.

Mavletova, A., Couper, M. P. and Lebedev, D. (2018), 'Grid and item-by-item formats in PC and mobile web surveys', *Social Science Computer Review*, Vol 36, No. 6, pp. 647–668.

Metzler, A., Kunz, T. and Fuchs, M (2015), 'The use and positioning of clarification features in web surveys', *Psihologija*, Vol 48, No. 4, pp. 379–408.

Miniukovich, A., Scaltritti, M., Sulpizio, S. and De Angeli, A. (2019), 'Guideline-based evaluation of web read-ability', paper presented at CHI Conference on Human Factors in Computing Systems

Proceedings (CHI 2019), Glasgow, Scotland, UK, available at https://doi.org/10.1145/3290605.3300738

Mohler, P., Dorer, B., de Jong, J. and Hu, M. (2016), 'Adaptation', *Guidelines for best practice in cross-cultural surveys, Survey Research Center*, Institute for Social Research, University of Michigan, Ann Arbor, MI, available at https://ccsg.isr.umich.edu/chapters/adaptation/

Moore, E. W. G., Lang, K. M. and Grandfield, E. M. (2020), 'Maximizing data quality and shortening survey time: Three-form planned missing data survey design', *Psychology of Sport and Exercise*, Vol 51, 101701.

Neuert, C. E. (2017), 'Processing forced-choice versus check-all-that-apply question formats: Evidence from eye tracking', *Field Methods*, Vol 29, No. 4, pp. 383–394.

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*, Vol 8, No. 4, pp. 617–635.

Neuert, C. E. (2021), 'The effect of question positioning on data quality in web surveys', *Sociological Methods & Research*, online first publication, available at https://doi.org/10.1177/0049124120986207

Nicolaas, G., Campanelli, P., Hope, S., Jäckle, A. and 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 2011-31.

Nicot, A.-M. and Houtman, I. (2006), 'Gender mainstreaming in surveys', *European Foundation for the Improvement of Living and Working Conditions*, available at https://www.eurofound.europa.eu/publications/report/2006/gender-mainstreaming-in-surveys

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. and 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.

Peytchev, A., Couper, M. P., McCabe, S. E. and Crawford, S. D. (2006), 'Web survey design: Paging versus scrolling', *Public Opinion Quarterly,* Vol 70, No. 4, pp. 596–607.

Peytchev, A. and Peytcheva, E. (2017), 'Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach', *Survey Research Methods,* Vol 11, No. 4, pp. 361–368.

Peytchev, A., Peytcheva, E., Conzelmann, J. G., Wilson, A. and Wine, J. (2020), 'Modular survey design: Experimental manipulation of survey length and monetary incentive structure', *Journal of Survey Statistics and Methodology*, Vol 8, No. 2, pp. 370–384.

Peycheva, D. N., Sakshaug, J. W. and Calderwood, L. (2021), 'Occupation coding during the interview in a web-first sequential mixed-mode survey', *Journal of Official Statistics*, Vol 37, No. 4, pp. 981–1007.

Radler, B. T. and Ryff, C. D. (2010), 'Who participates? Accounting for longitudinal retention in the MIDUS national study of health and well-being', *Journal of Aging and Health*, Vol 22, No. 3, pp. 307–331.

Raghunathan, T. E. and Grizzle, J. E. (1995), 'A split questionnaire survey design', *Journal of the American Statistical Association*, Vol 90, No. 429, pp. 54–63.

Redline, C. (2013), 'Clarifying categorical concepts in a web survey', *Public Opinion Quarterly*, Vol 77, No. S1, pp. 89–105.

Redline, C., Dillman, D. A., Dajani, A. N. and 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,* Vol 19, No. 4, pp. 403–420.

Revilla, M. and 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*, Vol 62, No. 5, pp. 538–545.

Rhemtulla, M. and Little, T. D. (2012), 'Planned missing data designs for research in cognitive development', *Journal of Cognition and Development*, Vol 13, No. 4, pp. 425–438.

Roberts, C., Sarrasin, O. and Stähli, M. E. (2020), 'Investigating the relative impact of different sources of measurement non-equivalence in comparative surveys', *Survey Research Methods*, Vol 14, No. 4, pp. 399–415.

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

Schierholz, M., Gensicke, M., Tschersich, N. and Kreuter, F. (2018), 'Occupation coding during the interview', *Journal of the Royal Statistical Society: Series A*, Vol 181, No. 2, pp. 379–407.

Schneider, N. F., Bujard, M., Wolf, C., Gummer, T., Hank, K. and Neyer, F. J. (2021), 'Family Research and Demographic Analysis (FReDA): Evolution, framework, objectives, and design of "The German Family-Demographic Panel Study"', *Comparative Population Studies*, Vol. 46, pp. 146–186.

Schoua-Glusberg, A. and Villar, A. (2014), 'Assessing translated questions via cognitive testing', in Miller, K., Willson, S., Chepp, V. and Padilla, J. L. (eds.), *Cognitive interviewing methodology*, John Wiley & Sons, Hoboken, NJ, pp. 51–67.

Sha, M., Hsieh, Y. and Goerman, P. (2018), 'Translation and visual cues: Towards creating a road map for limited English speakers to access translated Internet surveys in the United States', *Translation & Interpreting,* Vol 10, No. 2, pp. 142–158.

Singer, E., Van Hoewyk, J., Gebler, N., Trivellore, R. and McGonagle, K. (1999), 'The effect of incentives on response rates in interviewer-mediated surveys', *Journal of Official Statistics*, Vol 15, No. 2, pp. 217–230.

Smits, N. and Vorst, H. C. (2007), 'Reducing the length of questionnaires through structurally incomplete designs: An illustration', *Learning and Individual Differences*, Vol 17, No. 1, pp. 25–34.

Smyth, J. D., Christian, L. M., and 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*, Vol 72, No. 1, pp. 103–113.

Smyth, J. D., Dillman, D. A., Christian, L. M. and Stern, M. J. (2006), 'Comparing check-all and forced-choice question formats in web surveys', *Public Opinion Quarterly*, Vol 70, No. 1, pp. 66–77.

Sommer, E. (2017), 'Summary of existing web panel strategies', Deliverable 7.1 of the Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS) project), available at https://seriss.eu/wp-content/uploads/2017/12/SERISS_Deliverable_7.1-Ammended.pdf

Sommer, J., Diedenhofen, B. and Musch, J. (2017), 'Not to be considered harmful: Mobile-device users do not spoil data quality in web surveys', *Social Science Computer Review*, Vol 35, No. 3, pp. 378–387.

Terentev, E. and 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*, Vol 22, No. 2, pp. 179–198.

Tijdens, K. (2014), 'Dropout rates and response times of an occupation search tree in a web survey', *Journal of Official Statistics,* Vol 30, No. 1, pp. 23–43.

Tijdens, K. (2015), 'Self-identification of occupation in web surveys: Requirements for search trees and look-up tables', *Survey Methods: Insights from the Field*, pp. 1–11.

Toepoel, V. (2017), 'Online survey design', in Fielding, N. G., Lee, R., M. and Blank, G. (eds.), *The SAGE handbook of online research methods*, SAGE Publications Ltd., pp. 184–202.

Toepoel, V., Das, M. and Van Soest, A. (2009), 'Design of web questionnaires: The effects of the number of items per screen', *Field Methods*, Vol 21, No. 2, pp. 200–213.

Toepoel, V. and 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*, 0894439318784882.

Tourangeau, R., Brick, J. M., Lohr, S. and Li, J. (2017), 'Adaptive and responsive survey designs: A review and assessment', *Journal of the Royal Statistical Society: Series A*, Vol 180, No. 1, pp. 203–223.

Tourangeau, R., Couper, M. P. and Conrad, F. G. (2004), 'Spacing, position, and order: Interpretive heuristics for visual features of survey questions', *Public Opinion Quarterly,* Vol 68, No. 3, pp. 368–393.

Tourangeau, R., Couper, M. P. and Conrad, F. G. (2007), 'Color, labels, and interpretive heuristics for response scales', *Public Opinion Quarterly*, Vol 71, No. 1, pp. 91–112.

Tourangeau, R., Maitland, A., Rivero, G., Sun, H., Williams, D. and Yan, T. (2017), 'Web surveys by smartphone and tablets: Effects on survey responses', *Public Opinion Quarterly*, Vol 81, No. 4, pp. 896–929.

Tourangeau, R. and Yan, T. (2007), 'Sensitive questions in surveys', *Psychological Bulletin,* Vol 133, No. 5, pp. 859–883.

Wang, L., Sha, M. and Yuan, M. (2017), 'Cultural fitness in the usability of U.S. Census Internet Survey in Chinese language', *Survey Practice*, Vol 10, No. 3, pp. 1–8.

Watson, N. and Wooden, M. (2009), 'Identifying factors affecting longitudinal survey response', in Lynn, P. (ed.), *Methodology of longitudinal surveys*, John Wiley & Sons, Hoboken, NJ, pp. 157–182.

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C. and Verhoeven, S. (2021), 'Conducting general social surveys as self-administered mixed-mode surveys', *Public Opinion Quarterly*, Vol 85, No. 2, pp. 623–648.

Wood, J., Matthews, G. J., Pellowski, J. and Harel, O. (2019), 'Comparing different planned missingness designs in longitudinal studies', *Sankhya: The Indian Journal of Statistics*, Vol 81, pp. 226–250.

Yan, T. and Tourangeau, R. (2008), 'Fast times and easy questions: The effects of age, experience and question complexity on web survey response times', *Applied Cognitive Psychology*, Vol 22, pp. 51–68.

Züll, C. (2016), 'Open-ended questions', *GESIS Survey Guidelines*, GESIS – Leibniz Institute for the Social Sciences, available at https://doi.org/10.15465/gesis-sg_en_002

**The European Foundation for the Improvement of Living and Working Conditions (Eurofound) is a tripartite European Union Agency established in 1975. Its role is to provide knowledge in the area of social, employment and work-related policies according to Regulation (EU) 2019/127.**