WORKING PAPER

Eurofound

Living conditions and quality of life

Transitioning from interviewer-administered to self-administered designs in cross-sectional surveys

Author: Ayline Heller (Leibniz Institute for the Social Sciences, GESIS)

Research managers: Eszter Sandor, Daphne Ahrendt and Gijs van Houten

Eurofound reference number: WPEF24071

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

Executive Summary **1**
1. Introduction **3**
2. Problems and challenges of IAM **5**
2.1 Survey nonresponse 5
2.2 Increasing cost 5
2.3 Fieldwork difficulties 6
2.4 Changing role of the interviewer 6
2.5 Effects on data quality and comparability 9
3. Challenges and opportunities of mode transitions **10**
3.1 Mode effects and the Total Survey Error 12
3.2 Meta-analytic evidence on mode effects 14
3.3 Dealing with mode effects and effects of mode transitions 15
4. Practice and experiences from survey programs **19**
4.1 European Value Study (EVS) 22
4.2 European Social Survey (ESS) 24
4.3 German General Social Survey (ALLBUS) 25
4.4 European Health Interview Survey (EHIS) 26
4.5 Generations and Gender Survey (GGS) 27
5. Scenarios **28**
5.1 Fully face-to-face IAM 31
5.2 SAM with a single-mode, web-based design 31
5.3 SAM with a single-mode, paper-based design 32
5.4 SAM-MM with a choice of web- and paper-based participation 32
5.5 Parallel run of IAM and SAM-MM 33
5.6 Additional considerations 33
References **35**

Executive Summary

In recent years, there has been a shift away from the well-established, interviewer-administered mode of data collection (IAM), like face-to-face or telephone interviews, towards web- or paper-based self-administered survey modes (SAM). Mixed-mode designs combining different approaches of data collection are becoming a standard tool. The reasons for this shift are manifold: increasing survey nonresponse despite increasing fieldwork efforts, surging costs, and increasing uncertainties associated with data collection – many long-running international survey programs are calling into question whether the current IAM is still yielding the same, high-quality data at a reasonable cost and effort. They are thus experimenting with self-administered, mixed-mode survey designs (SAM-MM) to retain or even increase data quality, while reducing fieldwork efforts and risks associated with it. However, transitioning survey modes comes with specific challenges, especially in the context of international, repeated cross-sectional surveys: comparability across modes and thus across the time series may be impaired. Drawing from experiences of the European Value Survey (EVS), the European Social Survey (ESS), the German General Social Survey (ALLBUS), the European Health Interview Survey (EHIS) as well as the Generations and Gender Survey (GGS), this working paper aims to guide through the changing survey landscape, addressing some of the major challenges and opportunities of mode transitions for international survey programs.

In line with meta-analytic findings, evidence from these survey programs suggests that SAM-MM designs may be a promising alternative to (face-to-face) IAMs, potentially reducing coverage and nonresponse biases and thus increasing representativity of the sample at a lower cost and with less fieldwork effort. Considering errors of measurement, differences in (observed and latent) mean scores may be possible, especially regarding sensitive topics. However, first results from EVS, ESS, EHIS and GGS indicate that the relationships between indicators did not change as much. Differences could thus potentially be mitigated through ex post harmonization.

In addition to assessing and harmonizing data after it has been collected, many issues can be mitigated or even avoided through clever design choices. Practical considerations and country-specific customs have to be taken into account when deciding on the initial mode of contact, mode choice sequences, and an adequate incentivisation strategy. While some survey programs strive towards standardization, e.g., by providing guidelines for mode designs for all participating countries, country-specific adaptations may be necessary to ensure comparably high data quality, as not all mode designs work equally well in all countries (e.g., due to internet accessibility and penetration rates, if a web-based SAM component is being included).

Ideally, any changes in survey design should be accompanied by an experimental implementation to estimate and quantify effects on survey errors. In the case of mode transitions, a parallel run with separate samples of both the previous and the envisaged new mode design can be used to isolate mode effects, because time-related exogenous influences are excluded. Moreover, the random assignment to either the “old” or the “new” modes facilitates causal inferences on mode effects and finally, data may be used to establish a base line for harmonization in the future. If implementing a complete parallel run is not feasible, implementation in selected countries may be a viable option (as in the EVS). Moreover, experimental pilot studies (as in the GGS) or split-designs to test specific design features, like mode choice sequences or incentivisation strategies (as in ALLBUS or the GGS), can be useful tools.

As face-to-face IAMs are confronted with major issues, comparability over time is threatened, even if the same mode of data collection is being retained. Transitioning to a well-designed SAM-MM for data collection may thus be a promising avenue for the future, especially if the transition is accompanied by experimental implementation of certain design features.

1. Introduction

For the longest time, conducting large-scale, repeated cross-sectional surveys in a face-to-face, interviewer-administered mode (IAM) has been considered the gold standard when it comes to data quality. In recent years, however, studies have faced severe fieldwork difficulties, including higher refusal and noncontact rates and difficulties with interviewer recruitment and training. Not only are these developments severely impeding planning with regards to both scheduling and funding, but they also pose a serious threat to data quality and thus comparability of data over time.

As a result, many large-scale survey programs, such as the European Value Survey (EVS), the Generations and Gender Survey (GGS), the European Social Survey (ESS), or the European Health Interview Survey (EHIS) are allowing participating countries to experiment with or implement self-administered (SAM) study designs, oftentimes combining different self-administered modes of data collection, like web- and paper-based, in a mixed-mode design (SAM-MM)1. The EVS, for example, has been allowing experimental parallel runs of a SAM-MM design in addition to the regular face-to-face setting since 2017 (European Values Study [EVS], 2020), and the ESS is planning to do the same in 2025 with the aim of switching to a fully SAM-MM design by 2027. In case of the EHIS, most countries transitioned to some form of mixed-mode design (MM) by the 3rd wave: Most of them to a SAM-MM, but some of them to MM designs combining both IAM and SAM. These developments have sparked the interest of survey researchers worldwide: At the biannual conference of the European Survey Research Association (ESRA) in 2023, the directors of both the European Social Survey (ESS) and the General Social Survey (GSS) invited "data creators, data users, and survey practitioners to discuss methodological and statistical challenges for cross-sectional studies considering such a move." (O'Muircheartaigh et al., 2023). This led to a lively exchange of experiences and reinforced the urgency of developing frameworks to help aide mode transitions, as remaining in the IAM setting may not only become increasingly difficult, but also ever less advisable with reference to data quality.

While mode transitions from IAM to SAM are certainly accompanied by various challenges, they may also open up new opportunities: In SAM-MM designs, higher flexibility regarding time and location of survey completion as well as the option to choose mode based on personal preferences and targeting of specific groups may help improved coverage and reduce nonresponse errors (Stadtmüller et al., 2021). The absence of an interviewer in SAM may reduce social desirability biases especially with regards to sensitive topics (Bosnjak, 2017). In fact, first results from Germany suggest that SAM-MM surveys "shorten the fieldwork period and lead to higher response rates, while being more cost-efficient than face-to-face surveys" (Wolf et al., 2021). They may thus pose a "viable alternative for cross-sectional general population surveys" (Wolf et al., 2021).

Drawing from experiences of selected surveys, we aim to guide through the changing landscape of data collection in order to help large-scale, cross-sectional survey programs make an informed decision on survey mode (or mode combination) based on their needs, intentions, and financial resources. The working paper is structured as follows: Section 2 explores the problems and

[1] While mixing modes is possible at all stages of data collection (including recruitment and follow-up phase), we will focus on combining different modes in the immediate response phase. Differences in initial mode of contact and mode choice sequencing will be briefly discussed in section 5.6.

challenges of IAM surveys that now motivate an increasing number of survey programs to switch to SAM approaches, focusing on survey nonresponse, increasing costs, fieldwork difficulties, and the changing role of the interviewer. Section 3 draws attention to the challenges and opportunities associated with mode transitions. We will consider both errors of representation as well as errors of measurement within single-mode and mixed-mode designs and discuss them in the light of current meta-analytic evidence. A special focus of section 3 will further be comparability both across the mode change in the context of a time series and between the modes in an SAM-MM design. We will discuss possible comparability issues, ways of assessing comparability, and ways to prevent (ex-ante harmonization) or mitigate (ex-post harmonization) comparability issues. Section 4 then gives an overview of the practice and experiences of different large survey programs (EVS, ESS, the German General Social Survey (ALLBUS), Generations and Gender Survey (GGS)). All studies we draw from are conducted in a repeated cross-sectional setting – except the GGS which is a panel survey – with time series ranging back as far as 1980, in the case of ALLBUS. They all use a probability-based sampling approach and all of them have been facing fieldwork difficulties that have led them to an implementation of mode changes at least to some extent to reduce cost and maintain or increase data quality. Lastly, in section 5 we explore different mode-configuration scenarios, including further considerations on initial mode of contact, mode choice sequencing, incentive strategies, as well as the issue of standardization vs. country-specific adaptations. Specifically, we will include the following scenarios:

1. Fully face-to-face, IAM setting
2. Fully SAM using either
    a. A single-mode, web-based or
    b. A single-mode, paper-based design
3. SAM-MM setting using web- and paper-based modes
4. Parallel run of both 1. And 3.

For all scenarios, advantages and disadvantages are weighed up in order to enable readers make an informed decision on which survey mode to choose going forward. We will conclude with some additional considerations that should be taken into account before implementing a mode switch.

2. Problems and challenges of IAM

The landscape of data collection has undergone drastic changes in the last decades with survey programs being confronted with declining response rates coupled with increasing refusal of potential respondents to participate at all (Beullens et al., 2018; Leeuw et al., 2019). While this trend is not limited to the face-to-face setting, IAMs are especially vulnerable, as they depend on trained and experienced interviewers. Increasing fieldwork efforts to raise participation is thus connected to a sharper rise in survey cost compared to SAM as well as higher risks regarding the scheduling. Moreover, the role of interviewers is changing, becoming more multi-facetted and complex (Charman et al., 2024). As a result, interviewer recruitment is becoming more difficult and additional interviewer training and supervision may be needed, yet again adding to the financial burden of IAMs. While rising costs are a grave issue, especially in the face of shrinking budgets, the core problem lies elsewhere: all these issues may impact data quality and comparability of measurement over time. Thus, remaining in the same survey mode may not yield the same results.

This section will outline problems and challenges IAMs are facing, starting with the general trend of increasing survey nonresponse across all modes that has led to a striking rise in survey cost. We will then explain why the face-to-face IAM was especially affected by this development, giving some insight on fieldwork difficulties and the changing role of interviewers. This section concludes with a brief outlook on the effects these developments might have on data quality.

2.1 Survey nonresponse

Survey nonresponse is not a new phenomenon. While the COVID-19 pandemic has fuelled the trend, dropping response rates have been reported long before then. Using trend data obtained from the Labour Force Surveys of 16 different countries and covering the years of 1980 to 1997, Leeuw and de Heer (2002) analysed response rates as well as noncontact and refusal rates as three core components of nonresponse. For data collection, countries may choose between different survey modes: face-to-face or telephone-based IAM, web- or paper-based SAM, or different types of MM designs. Even though patterns differed between countries, an overall trend was visible: Declining response rates coupled with an increase in noncontact rates and refusals to participate. Building on their previous analyses and covering the years until 2015, Leeuw et al. (2018) found these trends to continue, with only a small deceleration in the rise of refusal rates.

2.2 Increasing cost

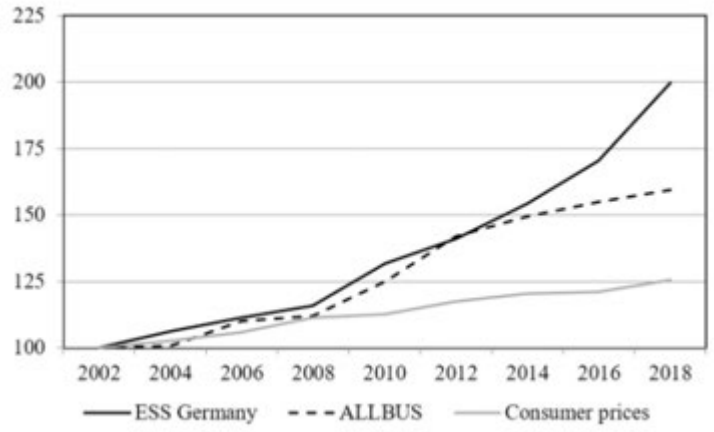
While evidence suggests that all survey modes are affected by this trend, the face-to-face IAM is particularly vulnerable, as it is more time-consuming and costly to begin with (fixed, set-up costs) and the cost per completed questionnaire (variable cost, depending on the size of the realized sample) tends to be higher. Consequently, costs are associated with more uncertainties and tend to increase at a higher rate compared to SAMs. Analysing the development of cost for ALLBUS and the ESS Germany between the years of 2002 and 2018, Wolf et al. (2021) observe an increase of up to 200% in survey costs compared to about 125% increase in consumer prices (see Figure 1).

2.3 Fieldwork difficulties

To address the issue of survey nonresponse, many survey programs have started to increase fieldwork efforts. Unfortunately, this while it may have stabilized response rates, it was not a necessarily a successful strategy to raise them: Results of the study by Wolf et al. (2021) confirm an ongoing trend of stagnating and even declining response rates for three large survey programs using a face-to-face IAM: ESS Germany, ALLBUS and the US-General Social Survey (US-GSS). In the case of ESS Germany, this was coupled with an increase in refusal rates, while an increase in noncontact rates could be observed in the ALLBUS. Results of the US-GSS further suggest that response rates are continually dropping even though fieldwork effort (as indicated by the number of fieldwork days) is steadily increasing.

This is confirmed by a study by Beullens et al. (2018), analysing data of the seven waves of the ESS (2002-2014) across the 36 EU countries. They conclude that response rates below 50% are quite common in Europe, despite an increase in fieldwork efforts: While noncontact rates were declining – which they took as an indicator for an increase in fieldwork efforts with (successful) multiple contact attempts – response rates were still dropping, and refusal rates rose (see Table 1 for the case of Germany). In addition the results by Wolf et al. (2021) presented above, evidence thus suggests that increasing fieldwork efforts will be successful only to a certain extent.

Figure 1. Costs for ALLBUS and ESS Germany as well as consumer prices, 2002 = 100.



Note: Price estimates are based on offers from field institutes to carry out the respective survey. Consumer price indices provided by German Federal Statistical Office

Source: Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys. Public Opinion Quarterly, 85(2), 623–648. https://doi.org/10.1093/poq/nfab03

2.4 Changing role of the interviewer

While data from the Labor Force Survey show that declining response rates are a problem across all survey modes, IAMs, once again, are particularly affected by these developments, also because they rely on trained and professional interviewers. During the COVID-19 pandemic, many surveys were forced to switch to a survey mode that did not rely on personal contact. Charman et al. (2024) point out that this led to a wave of retirements of some of the most experienced interviewers in the UK. Experience, however, is central to raise co-operation rates in IAM settings (Jäckle et al., 2013). At the

same time, the role of the interviewer is becoming more complex and multi-facetted. Multiple contact attempts and higher refusal rates call for more engaging, motivating, and resilient interviewers, who "now need to be persuaders as well as interviewers" (Charman et al., 2024, p. 13). As a consequence, extended training and fieldwork monitoring is necessary to retain standards and ultimately avoid interviewer effects. The greatest advantage of IAMs has been turning into a central weakness: the presence of an interviewer (Leeuw, 2008).

Table 1. Response rates, refusal rates, non-contact rates, and fieldwork duration for three social surveys, 2002 to 2018

Year	ALLBUS				ESS Germany				US-GSS			
	Response rate	Refusal rate	Non-contact rate	Field- work in days	Response rate	Refusal rate	Non-contact rate	Field- work in days	Response rate	Refusal rate	Non-contact rate	Field- work in days
2002	47.3	41.5	4.8	180	51.7	28.2	5.7	178	70.1	26.1	1.5	141
2004	45.7	42.4	6.7	134	51.0	32.8	7.0	144	70.4	22.5	2.4	140
2006	41.0	45.8	6.4	157	52.9	25.4	5.0	137	71.2	23.3	1.1	154
2008	40.3	48.2	6.6	176	42.7	32.6	6.4	158	70.4	24.1	1.2	150
2010	34.4	53.4	6.5	155	29.7	39.6	7.4	142	70.3	24.5	1.8	151
2012	37.6	49.2	6.5	161	33.7	45.9	4.7	139	71.4	21.0	2.3	170
2014	35.0	49.7	7.2	174	31.4	47.9	7.6	172	69.2	26.4	1.2	195
2016	34.9	47.0	10.3	167	30.6	48.0	3.3	216	61.3	32.7	2.0	229
2018	32.3	49.3	10.3	171	27.6	50.4	1.8	188	59.5	36.0	1.6	213

Note: Data were kindly provided by the respective teams; we thank Michael Blohm (ALLBUS) and René Bautista (US-GSS). Data for the ESS is based on Matsuo and Loosveldt (2013) for ESS 2002–2010, Beullens et al. (2014) for ESS 2012, Beullens and Loosveldt (2016) for ESS 2014, Wuyts and Loosveldt (2019) for ESS 2016, as well as own calculations for ESS 2018.

Source: Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys. Public Opinion Quarterly, 85(2), 623–648. https://doi.org/10.1093/poq/nfab03

2.5 Effects on data quality and comparability

In light of these developments, it is increasingly hard for face-to-face IAM to retain data quality, as limiting response bias and interviewer effects comes at a very high cost. Any changes in data quality may threaten comparability across time. The necessity to consider changing the mode of data collection is apparent, because remaining in IAM may pose as much of a threat to data quality and comparability as mode transitioning. In fact, SAM-MM designs are often viewed as a means to ensure data quality at a reasonable price. To assess the impact of switching survey modes from IAM to SAM settings, we will discuss differences between modes in, and the effect mode transitions may have on, data quality and comparability.

3. Challenges and opportunities of mode transitions

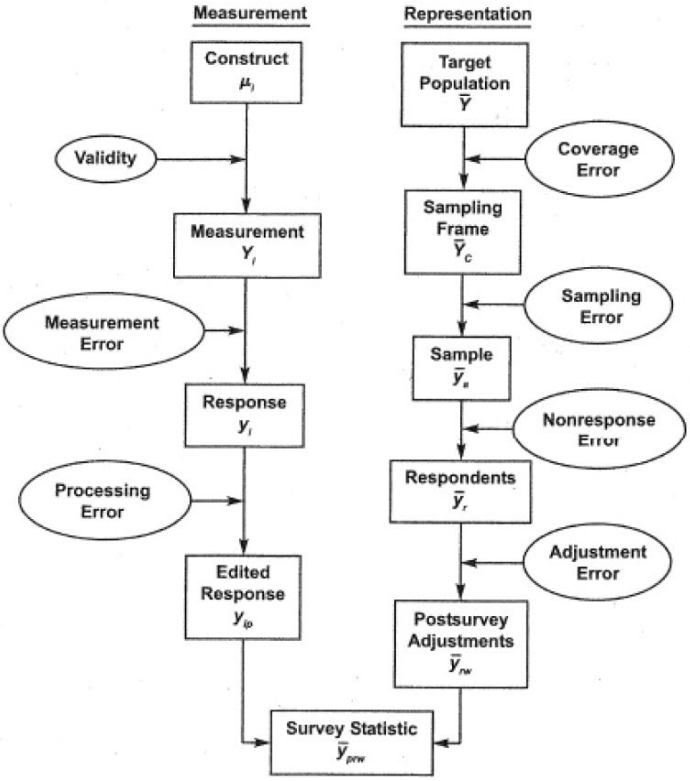
The issues outlined in the previous section show that the decision of survey mode not only affects timeliness, costs, and logistics, but may ultimately impair data quality as well. Data comparability is thus threatened, as it may decrease if the same mode is retained. While transitioning from IAM to a different form of data collection may offer the opportunity to retain or even increase data quality, the issue of comparability still has to be addressed, as any changes in quality may affect comparability.

The effects of mode on data quality and comparability can be structured along the lines to the Total Survey Error Framework (TSE; Groves et al., 2009). The TSE differentiates two types of errors that are relevant to survey statistics. While errors of representation are linked to coverage issues, i.e., differences "between the target population and the sampling frame" (p. 54), sampling, i.e., differences "between the sample and the sampling frame" (p. 57), and nonresponse (by affecting the probability of respondents to participate and complete the survey), errors of measurement are associated with the validity of the construct, the measure used to assess the construct and the response given by the participant (see Figure 2)2.

It is obvious that these two types of errors are interlinked, and survey mode can affect them in many different ways. To give an example, a well-trained interviewer can motivate reluctant respondents to participate at all, thus raising response rates, increasing coverage, and

[2] Both processing and adjustment errors are not as vulnerable to mode effects as they happenafter data collection. They are thus not the focus of the following sections.

Figure 2. The Total Survey Error Framework (TSE)



Source: Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey Methodology. Wiley Series in Survey Methodology. Wiley.

ultimately reducing nonresponse bias. This would be tapping into the representation side of the TSE. But they can also influence participants response behaviour during the interviewing process, thus possibly creating distortions in measurement, both in a negative or a positive way. On the one hand, they can serve as a motivation giving verbal and (in the case of face-to-face) nonverbal feedback (cf. Hope et al., 2022), they can minimize distractions and answer queries, thus reducing the cognitive burden and ultimately limiting satisficing behaviours. On the other hand, the presence of an interviewer may induce socially desirable response behaviour, interviewers could distort responses by adding interpretation through recording ambiguous answers as unambiguous, or through giving suggestive additional information. It is important to note that even though the former would increase data quality compared to self-administered survey modes and the latter would decrease it, any changes pose a challenge to comparability across modes in the context of mode transitions.

In order to make an educated decision when it comes to mode switches, a basic understanding of mode effects is thus necessary. In the following sections, we will outline different ways, survey mode can affect both errors of representation and errors of measurement in single-mode and mixed-mode surveys. We will give some empirical evidence on mode effects, drawing from meta-analyses and systematic reviews. Finally, we will explain, how issues arising from mode switches may be mitigated both ex ante via design as well as ex post by assessing and adjusting for them.

3.1 Mode effects and the Total Survey Error

The choice of survey mode should always be guided by the research questions, and different data collection methods may be more or less adequate given the certain concepts to be measured and the specified target population (Leeuw, 2005). Furthermore, (country-specific) restrictions and survey customs as well as practical considerations will always influence the decision-making process. While each mode has its strengths and weaknesses, a basic differentiation between IAM (e.g., face-to-face, telephone) and SAM (web- or paper-based) is useful. In the case of (SAM)-MM, many combinations of modes are possible. For example, the mode of recruitment may differ from the mode of data collection, as is the case when advance or follow-up letters are sent out (Leeuw, 2008). Different modes can be offered concurrently or sequentially, as in push-to-web designs, where respondents are invited to an online survey via mail and are only given the option to participate via paper-and-pencil upon a first or second reminder (Stadtmüller et al., 2021). Or different modes can be used for different parts of the survey, as in a self-administered part for sensitive topics during a face-to-face survey (de Leeuw, 2008). Focusing in on the two branches of the TSE (see Figure 2), we will describe how certain mode features may influence errors of representation and errors of measurement in both single-mode IAM/SAM or MM designs.

3.1.1 Errors of representation

Starting with the right branch of the TSE (see Figure 2), the choice of survey mode may impact errors of representation through errors of *coverage*, i.e., differences "between the target population and the sampling frame" (Groves et al., 2009, p. 54) that may arise due to media usage and practical consideration. For example, differences in internet accessibility may hinder participation in general or for specific subgroups in the web mode. Errors of *sampling*, i.e., differences "between the sample and the sampling frame" (Groves et al., 2009, p. 57), arise when not all potential respondents in the sampling frame have the same propensity of being part of the sample, e.g. in nonprobability samples like online panels. As we are assuming all samples to be probability-based, this error is not as relevant to our considerations. Finally, *nonresponse* error refers to differences "between the sample and the respondent pool", i.e., potential respondents systematically refusing to participate, e.g., because of a lack of motivation or due to discomfort with the mode choices offered.

In addition to the errors outlined above, *self-selection* may arise in MM designs, as participants may differ systematically in their mode preference and ability to participate in a certain mode. At the same time their responses may differ systematically depending on the mode of their participation, due to the characteristics of the respective mode. In the next section, we will thus give an overview of mode effects on measurement error to then turn to different ways of assessing, disentangling, and mitigating both types of errors.

3.1.2 Errors of measurement

Moving from the representation to the measurement side of the TSE, the focus shifts towards individual responses. Survey modes form the context for the response situation. As such, they can have subtle effects on any aspect of the response process. Modes may influence all stages of the response process, such as question comprehension, retrieval of information, judgement, and response selection (Tourangeau et al., 2000). They may also impact the difficulty perception and motivational components of the response process and thus increase or decrease respondents' tendency to answer carefully (optimizing) or carelessly (satisficing) (Krosnick, 1991). Looking at the consequences for the collected data, we discuss mode effects that change the data quality, mode

effects that change the scaling (i.e., the measurement units), and mode effects specific to aspects of socially desirable responding and sensitive questions.

Quality differs between modes when modes make questions easier or harder to understand, or when modes differ in how well they can maintain respondent motivation and focus. In other words, comprehension may depend on how we can provide information to respondents. De Leeuw (2008) states that in IAM, more complex structures can be used, as the interviewer can provide additional explanations and repair misunderstandings. At the same time, SAM, especially the web mode, allows for the integration of different media types, thus allowing to use multiple communication channels. As for motivation, interviewers in IAM may help with engagement, cushion frustrations, and minimize external distraction. At the same time, the greater temporal and spatial flexibility of SAM modes may have its own motivation benefits. Lastly, different modes may also reduce or increase extraneous respondent burden. For example, in telephone interviews respondents must keep many response options in their working memory while in personal interviews, web, and paper-based surveys they can relegate that to visual aids.

Scaling differs between modes when respondents translate their internal attributes differently into the response scheme depending on the mode. This is less of an issue for nominal, manifest concepts (such as education levels) but certainly an issue for latent constructs, such as subjective well-being, attitudes, values, or personality. The very same respondent who would choose “somewhat interested” in politics in one mode, might now choose “interested” in another mode. That is not necessarily an error. It is simply a different scaling in that “interested” now represents a different true interest level. Such scaling differences usually do not impact the data quality of the modes in separation, they only pose a problem if we want to compare responses across modes (e.g., along a time series with a switch from IAM to SAM). In essence, mode changes may lead to what psychometricians call a measurement invariance violation (see section 3.3.2 and Box 1).

Lastly, modes differ in socially desirable responding and in how respondents react to sensitive questions. We emphasize this topic separately from quality and scaling for two reasons. First, the impact of modes on socially desirable responding and sensitive question responses is a long-standing area of research with a focus on the differences between IAM and SAM. This is because modes with interviewers may induce respondents to respond desirably or to shy away from sensitive questions in ways which SAM modes do not. The second reason is that socially desirable responding may express itself both as a quality or a scaling problem. It becomes a quality problem if socially desirable responding becomes outright deceptive (positive, but factually wrong answers). However, more moderate forms of socially desirable responding still retain all the measured information but with a slightly shifted scale. If all respondents choose slightly more (or less) favourable responses, then the response distribution shifts its mean value, but the differences between respondents are retained.

3.1.3 Overlapping effects in mixed-mode designs

Many researchers have pointed out that the choice of survey mode is always a trade-off between cost and errors, and MM approaches give “an opportunity to compensate for the weaknesses of each individual mode at affordable cost” (Leeuw, 2005, p. 235). De Leeuw (2005) gives an overview of different rationales for the implementation of MM designs in different phases of data collection and their effects on survey quality: During the contact phase, advance notifications may be used to correct the sampling frame, raise response rates, and enhance credibility and trust, thus reducing coverage and nonresponse errors without any threats to measurement (given a single-mode

response phase). During the follow-up phase, reminders can be used to raise response rates, thus reducing nonresponse errors, but possibly affecting measurement, if combined with the offer of participating in a different mode of data collection. In the following, we will focus on combining different modes during the response phase, which is by far the most complex situation: While it may reduce costs, improve coverage, and improve privacy of measurement (when SAMs are being offered), comparability across modes and time may be impaired due to differences in representation and measurement. It is thus central to address mode effects both ex ante via design choices and ex post using existing methods to assess and mitigate errors (see section 3.3).

3.2 Meta-analytic evidence on mode effects

As the number of studies comparing different types of single-mode or mixed-mode designs has drastically increased in recent years, it makes sense to first take a look at meta-analytic evidence to get a general idea of mode effects on errors of representation and errors of measurement. Reviewing the results of several meta-analytic studies, the study by Bosnjak (2017) can serve as a valuable guide. With regards to errors of representation, "face-to-face surveys typically achieve the highest response rates, followed by [paper-based] mail surveys, and Web-based surveys at the bottom of the response rate league" (p. 19). However, all meta-analyses considered in this study were conducted in the 80s, 90s or early 2000s, with internet access and penetration rates being much lower (World Bank Open Data, 2024). In the following paragraphs, we will thus first discuss internet coverage as a predictor for response rates in single-mode web surveys. We will then draw attention to the connection of response rates and sample representativity, presenting evidence that mixing survey modes may in indeed improve representativity, while at the same time reducing survey cost, if implemented correctly. Finally, we will assess the (limited) meta-analytic evidence on errors of measurement that points towards differences in response behaviour on sensitive topics. Strategies to potentially mitigate these differences exists and will be discussed in section 3.3.2.

Regarding single-mode web surveys, a meta-analysis by Daikeler et al. (2020) focused on the moderators of mode performance in 114 experimental comparisons between web and other survey modes. Even though internet accessibility has improved during that time frame, web surveys still fared worse with regards to response rates compared to other survey modes (paper-based mail, telephone or other). Several moderators could be identified, including the mode of initial contact, the number of contact attempts, the use of prenotifications, and the survey country. Results from another meta-analysis elaborate on the last issue suggesting "web surveys achieve high response rates in countries with high population growth, high internet coverage, and a high survey participation propensity. On the other hand, web surveys are at a disadvantage [with respect to differences in response rates to comparison survey modes] countries with a high population age and high cell phone coverage." (Daikeler et al., 2022). With regards to cost-efficiency, this may not be as much of a problem in the web-based SAM as the variable cost is comparatively lower, depending on the specific design. However, survey nonresponse can be a risk to data quality as well, if sample composition is systematically distorted due to nonresponse bias. To this end, web-based SAM are often combined with other (more costly) surveying strategies, like reminders, prenotifications, or incentives to raise response rates (see section 5.6).

In a meta-analysis on 69 studies comparing representativity and nonresponse bias associated with different survey modes, Cornesse and Bosnjak (2018) find that single-mode web surveys are indeed "less representative than other single-mode surveys" (p. 9) when considering external benchmark

variables. Furthermore, they observe a positive association between using benchmark variables, R-Indicators, and response rates as different indicators for sample representativity and nonresponse bias[3], suggesting that decreasing response rates may generally be connected to issues of sample composition. Finally, they find that MM surveys tend to fare better regarding R-Indicators compared to single-mode surveys, indicating that mixing survey modes may indeed improve representativity of the sample.

According to Rybak (2023) mixing survey modes was also a successful strategy to reduce survey costs, if conducted correctly. In a meta-analysis of the ISSP waves 1996-2018 and the ESS rounds 1 to 9 (2002-2018), he found that both single-mode (paper-based) mail and MM surveys lead to lower nonresponse bias when using internal criteria of representativeness. Contrary to the analysis by Cornesse and Bosnjak (2018) there was no clear connection between response rates and nonresponse bias: while leading to the highest response rates, the face-to-face mode did not show the lowest bias. Rybak concludes that "the switch from face-to-face to other modes should be considered a great opportunity to raise research quality, even considering possible risks, mainly some kinds of mode effects".

Concerning errors of measurement, meta-analytic evidence is still scarce. The abovementioned review by Bosnjak (2018) finds "that within each of the two classes [SAM and IAM], there are no substantial differences in terms of social desirability bias, except for the propensity to report sensitive behaviors" (p. 18f.), with computerized SAMs indicating larger prevalence rates for sensitive behaviours, followed by paper-and-pencil SAMs and IAMs. As we will see in section 3.3.2, these differences can potentially be mitigated through ex post harmonization. Other indicators of measurement quality "point to comparable valid answers to factual questions between [paper-based] mail, face-to-face, and telephone modes". As meta-analytic evidence is still limited and does not yet include effects of mode transitions, we will draw attention to experiences of those survey programs that already started to implement changes in survey modes in section 4. Those survey programs that conducted a parallel run (or are planning to do so) are of special interested, as the data from these random experiments facilitate causal inferences on the effects of mode and mode switches. Before that, we will now present some general means of assessing and addressing mode effects as well as effects of mode transitions.

3.3 Dealing with mode effects and effects of mode transitions

Mode effects can be addressed at two different stages in the surveying process. First, before data is being collected, specific design choices may help increase data quality and/or comparability – with those two goals interfering with one another in some instances. Second, mode differences can (and should) be assessed after the collection phase, and some strategies exist to help mitigate errors ex post.

3.3.1 Ex ante: Design choices

If mode differences are addressed early in the survey planning process, many issues can be avoided, and errors can be mitigated through clever design choices. Many survey programs strive towards standardization wherever possible, but idiosyncrasies of certain modes (and countries) may demand

[3] See section 3.3.2 for an explanation of different indicators to assess sample composition and representativity.

some degree of flexibility. In ex ante design choices, standardization, and adaptation have to be weighed up to ensure high data quality that is comparable across time, mode, and country. Note that granting countries degrees of freedom in choosing a mode might be easier for programs with dedicated national teams in those countries than for programs without such national teams.

For example, regarding errors of representation, country-specific differences with regard to the availability of addresses as well as web coverage and other survey customs can be explored beforehand and help decrease coverage error. Using the optimal incentivisation strategy can help reduce nonresponse errors and specific mode choice sequences can be used to decrease cost while increasing response rates.

Regarding errors of measurement, questionnaires may have to be adapted to the mode of presentation. Survey complexity may have to be reconsidered to reduce respondent burden if no interviewer is present to answer queries[4]. Harmonization may be especially challenging in the case of sociodemographic variables, as their assessment tends to be a lot more detailed in IAM. The general layout may have to be optimized for the specific mode, e.g., if mobile devices are to be used. For MM, the rule of thumb states that in order to ensure comparability across modes, questionnaire design has to follow the restrictions of the most restrictive mode (Stadtmüller et al., 2021). While all these adjustments to the questionnaire are central to ensure data quality (and comparability across modes in the case of MM designs), they might, at the same time, threaten comparability of data over time, as any changes in data quality pose a threat to comparability, even if data quality is increasing. It is thus always necessary to assess relevant aspects of the TSE to mitigate errors were possible and ensure comparability.

3.3.2 Ex post: Assessment and harmonization

Because not all mode effects can – or should, e.g., in the case of a reduction of social desirability bias – be avoided through design, it is necessary to assess the quality and the extent of comparability issues and to use ex-post harmonization strategies to mitigate comparability issues after data have been collected.

With regard to errors of representation, response rates are usually used as an indicator for systematic unit nonresponse. An analysis of sample composition and a comparison with external benchmark criteria can further shed light on errors of coverage and nonresponse. This is usually done using variables that are not expected to show much variation across modes, such as sociodemographic variables. Otherwise, deviations between survey and benchmark might be representation or measurement related. In addition to reporting such differences in percentage point, R-Indicators, describing the propensity to respond to a survey, are often used as indicators. When no or insufficient external or benchmark criteria are available, internal criteria may be used instead, e.g., suggest using the gender balance in heterosexual couples living together in two-person households, as it should be 50:50 per definition (cf. Rybak, 2023). If assessment reveals that mode choice does not cause differences in representation, then we can apply the same mitigation strategies as for representation issues in general, such as adjustment weights or propensity score-based approaches.

[4] While respondent burden is generally considered a larger problem in long SAM surveys, recent results from the EVS suggest that there are little differences between longer and shorter SAM-MM questionnaires regarding measurement quality (Cernat et al., 2022).

Errors of measurement are usually estimated using indicators for response biases, like item nonresponse, nondifferentiation or straightlining, extreme answers or acquiescence. Response styles associated with social desirability can either be assessed using specific questionnaires or inferred by comparing answers in different modes. While this is usually done by evaluating observed mean differences between modes, methods to investigate measurement invariance can give more insights on the level at which problems of comparability may occur. Assuming a survey instrument measures a latent construct (e.g., subjective well-being, depression, optimism, etc.), we would want it to measure the same substantive construct with the same precision using the same units of measurement across modes and across time (in repeated cross-sectional surveys). Different levels of invariance are usually differentiated for multi-item constructs: If *configural* measurement invariance is met, latent constructs do not differ in their structure across modes. *Metric* invariance is the prerequisite for the analysis of structural relationships (variances and covariances) such as regression coefficients, and thus serves as the basis for analyses for reliability of scales. Finally, at least partial *scalar* invariance is necessary to make meaningful comparisons of latent mean scores.

If scalar invariance is met, but latent means differ systematically between modes, ex-post harmonization strategies can be used to transform and adjust scores. If scalar invariance is not met, methods of partial or approximate invariance (alignment, Bayesian approaches; Leitgöb et al., 2023) can be used to mitigate errors and achieve comparability of latent means. For pseudo-metric (often latent) concepts, such as values, interests, personality, or attitudes, this allows us to mitigate comparability issues with ex-post harmonization (e.g., observed score equating in a random groups design; Singh, 2022; Kolen & Brennan, 2014). Notably, this approach can even be applied to single-item measures.

Box 1. Measurement Invariance

Comparability, or formally speaking measurement invariance , across measurement contexts (between modes, across time, or across countries) is a multi-layered issue: Are we measuring the same substantive construct (*configural* invariance), with the same measurement precision (*metric* invariance), and the same units of measurement (*scalar* invariance)? In instruments for latent constructs, respondents may change a different response option based on the presentation of the response scales (e.g., verbally vs. visually or horizontally vs. vertically). Such violations of (scalar) invariance may lead to methodological artifacts in the data, like spurious mean differences between modes. Compared to fundamental differences in the construct, these scaling differences can be rectified through ex post harmonization.

As for mode switches in the context of long-running cross-sectional survey programs, parallel runs of both the “old” IAM and the anticipated “new” SAM or MM design provide the best basis to assess mode differences and establish a base line for harmonization. In these random experiments, time-related exogenous influences are excluded and the random assignment to either the “old” or the “new” modes facilitates causal inferences on mode effects. This means that we can directly compare the response distribution for each variable between IAM and SAM. Due to the random experimental structure, we would expect the same distributions in both modes. Deviations in distributions then

quantify the joint effect of different errors of representation and different errors on measurement between IAM and SAM. If errors of representation have been found to be comparable or sufficiently mitigated with the approaches detailed above, this implies we can even isolate different errors of measurement and mitigate them using ex-post harmonization strategies. Parallel runs are usually implemented once as a transitional stage before a complete switch is being performed. They can provide quantitative evidence of the effects of changes in mode and/or design. By preceding or accompanying them with specific experimental settings, like pilot studies or split-sample experiments, feasibility and impact of certain design choices can be systematically explored.

4. Practice and experiences from survey programs

Next, we take a look at the practices of several selected survey programs to understand how these long standing repeated cross-sectional studies have dealt with the changes in survey landscape and draw from their experiences with mode switches. As mentioned above, all studies use a probability-based sampling approach and implemented mode switches at least to some extent to address the difficulties they have been facing. We recommend looking at Table 2 for an overview. Below, we then address each study in greater detail, drawing attention to the survey type, the initial survey mode, the (planned or implemented) mode switch, and whether or not they intend to or have successfully implemented a parallel run of both SAM and IAM survey designs. As all surveys considered an MM design as most fitting for (experimenting with) mode transitions, we will further describe their respective mode of initial contact, the mode choice sequence that was used, the incentivisation strategy as well as other adjustments that were made. We will not describe every aspect of all surveys and due to the scope of this working paper, we will not be able to present results from all methodological experiments that were conducted, but rather draw attention to experiences that are especially relevant to considerations on mode transitions.

Table 2. Overview of selected surveys

Survey	Survey Type	Initial survey mode	Mode switch	Parallel run	Within MM			
					Initial contact	Mode choice sequence	Incentives	Other adjustments
EVS	Cross-sectional, every 9 years since 1981 including up to 47 countries	Face-to-face IAM	Optional SAM-MM alongside the IAM in 2017	Six countries in 2017	Postal (recommended)	Web first, paper pencil offered after one reminder (recommended), country-specific	Unconditional, prepaid (recommended), country-specific	Matrix design for web survey; country-specific adaptations
ESS	Cross-sectional, biannually since 2002 including up to 36 countries	Face-to-face IAM	Temporary SAM-MM in some countries during COVID, aim of switching to SAM-MM by 2027	Planned for 2025	2025: postal first or fieldworker assisted first	Experimental implementation in some countries in some waves	Country-specific	Country-specific
ALLBUS	Cross-sectional, biannually since 1980 (West Germany) /1991 (East Germany)	Face-to-face IAM	Temporary SAM-MM during COVID, aim of switching to SAM-MM by 2025	2023	Postal	2020: concurrent; 2021/2023: experimental tailored sequential	2023: Unconditional, prepaid + postpaid	Split survey for SAM-MM to reduce length
EHIS	Cross-sectional, three waves since 2006, including all EU member states	Country-specific, IAM recommended	Mostly MM by 3rd wave	-	Country-specific	Country-specific	Country-specific	Country-specific

Survey	Survey Type	Initial survey mode	Mode switch	Parallel run	Within MM			
					Initial contact	Mode choice sequence	Incentives	Other adjustments
GGS	panel, two rounds with three waves each, since 2004, first round in EU, second round included Central and East Asia and South America	Face-to-face	MM (Web-based SAM + additional mode; preferably face-to-face IAM), country specific adjustments	experimental pilot study in three countries	Country-specific	Push-to-web experiments in three countries	Experimental implementation in some countries	Further experiments regarding reminders and respondent selection in some countries

Note: EVS = European Value Survey, ALLBUS = German General Social Survey, ESS = European Social Survey, EHIS = European Health Interview Survey, GGS = Generations and Gender Survey, IAM = interviewer-administered mode, SAM = self-administered mode, MM = mixed-mode, SAM-MM = self-administered, mixed mode (web- and paper-based)

4.1 European Value Study (EVS)

The European Value Study is a repeated cross-sectional study that has been conducted every nine years since 1981 and included up to 47 European countries. In 2017, an option to use a SAM or SAM-MM design alongside the usual face-to-face IAM setting was added along with specific implementation guidelines to ensure comparability while allowing for necessary country-specific adjustments. This was accompanied by several methodological experiments. Overall, the implementation was evaluated positively, and SAM was endorsed as a complement or even a "viable alternative to face-to-face surveys when conducting general population studies (in Germany)" (Wolf et al., 2021, p. 644). As multi-national parallel runs like these are rare despite their usefulness when it comes to an assessment of the effects of mode switches, a closer look at the implementation of the 2017 EVS survey and the results available so far is warranted.

In the methodological guidelines, a web component with an optional sequential design to "increase the response rate and the diversity of respondents" (EVS, 2020, p. 33) is endorsed. It is stated that the sampling frame should be the same for both IAM and SAM, the initial contact was recommended as mail mode, and an unconditional prepaid incentive was endorsed. None of the six countries carrying out this parallel run implemented it exactly as stated in the guidelines, as country-specific adjustments were necessary, and several methodological experiments were conducted to assess different implementation strategies.

Table 3 gives an overview of the design features of the surveys in participating countries (Denmark, Finland, Germany, Iceland, Switzerland, and the Netherlands). All countries except the Netherlands used both web and paper modes, the Netherlands opted for a web only mode using an existing longitudinal panel as a sampling frame. Of the remaining countries, all used a push-to-web, sequential design, with Germany implementing methodological experiments for both mode choice sequence and incentivisation. Only the Swiss sample and parts of the German sample were implemented with the recommended unconditional incentive; the others used some form of conditional incentivisation strategy. To reduce overall response burden, four countries utilized shorter matrix questionnaires (i.e., respondents only saw different, systematically varied selections of questions) with either a follow-up survey to complement the matrix design (Iceland, the Netherlands, and Switzerland) and/or a parallel full-length questionnaire (Germany, Iceland, and Switzerland).

Table 3. Main country-specific design features of the EVS 2017 mixed-mode field

Design Feature	DK	FI	DE	IS	NL	CH
Mode(s)	Web and paper	Web and paper	Web and paper	Web and paper	Web only	Web and paper
Contact design	Invitation by letter; push-to-web: paper with 1st reminder	Invitation by letter, push-to-web, paper on request.	Invitation by letter; *MM matrix:* simultaneous (paper, web) vs. sequential (push-to-web) 3/4 contacts (phase 1/2) *MM full:* simultaneous (paper, web) 3 contacts	Invitation by letter; paper only on request	Invitation by e-mail;	Invitation by letter, push-to-web: paper with 2nd reminder
Type of sample	Random, individuals, register data	Random individuals, register data	Random, individuals, register data	Random, individuals, register data	Random selection among LISS panel participants	Random, individuals, register data
Incentive	Conditional monetary incentive	Conditional lottery of gift cards (10 × €100 and 1 × €500)	*MM matrix:* 5€ unconditional vs. 10€ conditional *MM full:* 5€ unconditional	Conditional lottery (10 × 10,000 ISK/63€ and 1 × 100,000ISK/635€)	Standard LISS panel (15€ per hour of survey completion)	9€ (10 CHF) unconditional + conditional lottery for FU (3 iPads)

Source: Luijkx, R., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. Þ., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., . . . Wolf, C. (2021). The European Values Study 2017: On the Way to the Future Using Mixed-Modes. European Sociological Review, 37(2), 330–346. https://doi.org/10.1093/esr/jcaa049

Analysing sample composition and response rates (as indicators for representativity of the sample), and item nonresponse (as an indicator for accuracy of measurement) in all six countries, Luijkx et al. (2021) conclude that SAM "was successful in most of the participating countries" and "[SAMs] can complement the traditional F2F mode in large-scale population-wide surveys; especially if it is possible to reduce bias and further increase data quality" (p. 343). In fact, SAM outperformed IAM with regards to response rates in Germany and Iceland with results from Germany suggesting remarkably lower costs (savings of over 50 percent) and shorter fieldwork time (6 to 8 weeks compared to 6 months for the IAM; Wolf et al., 2021). With regards to sample composition (compared to official population data) and item nonresponse, however, IAM performed slightly better, with both IAM and SAM yielding "acceptable data quality" (Luijkx et al., 2021) in all six countries.

Further investigating the effects of survey length and mode on data quality, and errors of measurement specifically, Cernat et al. (2022) analysed the German sample of the EVS comparing three implementation strategies (IAM, SAM-MM with full-length, and SAM-MM with a shorter, matrix questionnaire). Taking into consideration over 100 items in more than 24 measures, the authors looked at differences in point estimates, distribution of variables, and other indictors of data quality (e.g., item nonresponse and response style indicators) across these groups. Furthermore, they investigated measurement invariance (see section 3.1.2). Results suggest "only few systematic differences in data quality at the item level" (p. 32) between IAM and SAM-MM designs, including differences in point estimates, SAM-MM showing lower variation, and slightly higher item nonresponse (as already presented above). Regarding measurement invariance, 21 out of 24 measures were able to reach metric invariance and the majority (15 out of the 24) also reached scalar invariance, thus allowing for meaningful comparisons of latent means and variances between modes for most measures. Similar results were achieved regarding survey length, with 22 out 24 reaching metric and 19 out of 24 reaching scalar invariance.

Parallel runs of both matrix design and full-length questionnaire further suggest that survey length does not lower response rates in SAM, but may come with considerably less fieldwork efforts and may increase reliability of measures (Cernat et al., 2022). Wolf et al. (2021) thus recommend "fielding a single but longer self-administered survey instead of multiple shorter ones". Interestingly, if being offered a choice between web- and paper-based participation in SAM-MM, the share of paper-based respondents largely differed between countries, ranging from 5% in Iceland to 74% in Germany (Luijkx et al., 2021). This is likely to be attributed to specific survey climates and practices. In Germany, the share of paper-based respondents also differed between experimental designs: while the concurrent design had the largest share and was thus more expensive than the sequential design (e.g., through higher data encoding costs; Wolf et al., 2021), it was also most successful at reaching participants.

4.2 European Social Survey (ESS)

The European Social Survey has been conducted biannually since 2002 and included up to 36 European countries. Due to the COVID-19 pandemic, nine countries (Austria, Cyprus, Germany, Israel, Latvia, Poland, Serbia, Spain, and Sweden) switched to SAM-MM in Round 10, while 22 countries used the previous face-to-face IAM mode. In Round 11 (2023), the ESS returned to the previous design, but participating countries are highly encouraged to conduct a parallel run in Round 12 (2025) with the goal of switching to a fully SAM-MM design by Round 13 (2027; European Social

Survey [ESS], 2024). This transition was preceded by six experimental studies between 2003 and 2012 (Villar & Fitzgerald), and is assisted by an international team of four experts.

Looking at the data of Round 10, Koch and Briceno-Rosas (2023) report sociodemographic sample composition, comparing to the previous, fully IAM Round. While response rates dropped an average of 9.6 percentage points in those countries that used an SAM-MM (compared to 4.8 percentage points for those countries continued in IAM), the authors conclude that overall results "did not indicate that sample composition became worse among the countries fielding ESS 10 as a self-completion survey due to the Covid situation" (Koch & Briceno-Rosas, 2023, p. 22).

Analyses of the previous methodological experiments yield some additional insights on selection effects and comparability of measurement quality. Already in Round 4 (2008), an MM experiment was implemented in the Netherlands considering different sequences of web-based SAM, face-to-face IAM, and telephone-based IAM, alongside the regular, face-to-face IAM (as a separate sample with a single-mode design). Investigating selected items on media usage, life satisfaction, social and political trust, Revilla (2010) find little differences between the face-to-face IAM and the MM designs with regards to data quality. Interestingly, largest differences in both sample composition and measurement arise between telephone-based IAM and both face-to-face IAM and web-based SAM within the MM design, leading the author to the overall conclusion that "a mixed-mode using only CAPI [face-to-face IAM] and CAWI [web-based SAM] should not be problematic in terms of quality comparisons. Adding CATI [telephone-based IAM] however may be an issue if the difference between CATI and the two other modes comes from differential measurement and not from differential selection" (p. 163). The latter could not be differentiated given the design of the experiment.

In Round 8 (2016/17), as a result of the series of experimental studies, a cross-national online survey panel (CRONOS) was designed and implemented alongside the IAM to further investigate feasibility of probability-based web designs (ESS, 2024). Respondents of the regular survey were used as a recruitment base. Investigating differences in measurement quality, Cernat and Revilla (2020) find higher item nonresponse and higher levels of primacy effects in the CRONOS panel compared to the ESS, but similar levels of nondifferentiation. Furthermore, analyses of measurement invariance suggest that metric and scalar invariance can be assumed for four out of the five measures under investigation. By and large, these findings regarding errors of measurement are in line with those of the EVS and indicate that while there may be differences in measurement between IAM and SAM for some measures, they may be attributed to shifts in mean scores to an extent and may thus successfully mitigated through ex post harmonization (see section 3.3.2).

4.3 German General Social Survey (ALLBUS)

ALLBUS is among the oldest general social surveys in Europe as it has been running biannually since 1980 in West Germany and 1991 in unified Germany respectively. It is based on a stratified random sample of the German population aged 18 years and older with an oversampling of the East German population. ALLBUS, like other survey programs, has been dealing with decreasing response rates leading them to increase fieldwork efforts and implement several measures to address and help overcome this issue.

Already in 2010, a random experiment on conditional incentivisation was implemented with a random subsample receiving 10€ upon completion of the interview. This led to an increase in

response rates while also slightly reducing fieldwork efforts, as cooperation increased and contact attempts could be reduced (Blohm & Koch, 2017). Sample composition, on the other hand, did not show any major differences in the incentivized compared to the non-incentivised groups suggesting that errors of representation did not change due to the incentivisation strategy.

In 2021, due to the COVID-19 pandemic, ALLBUS switched to SAM-MM design with a web- and paper-based design as modes of data collection and using postal addresses obtained from official registries (mail mode) as an initial mode of contact. A push-to-web experiment was implemented, i.e., a randomized group of participants were offered the cheaper, web-based mode first and the option to participate on paper was only offered upon sending a reminder. Comparing the sequential and the concurrent mode sequencing groups, Asimov and Blohm (2024) find similar sample compositions. They further identified age as a candidate variable for a tailored mode sequencing approach: concurrently offering both modes to older target persons while using a sequential design for younger target persons slightly improved sample composition.

In 2023, a parallel run of both face-to-face IAM and SAM-MM with a web- and a paper-based component was conducted including another push-to-web experiment as well as an experiment with unconditional, prepaid as well as conditional, postpaid incentives. In 2025, ALLBUS will switch to a full SAM-MM design (Hochman et al., forthcoming).

4.4 European Health Interview Survey (EHIS)

The European Health Interview Survey (EHIS) has been collecting data on health status, health care use, and health determinants in three data waves starting 2006. While only 17 EU countries participated in the first wave, all EU member states as well as Iceland and Norway took part in the second wave (2013-2015) and Servia, Albania, and Turkey were added in the third wave (2019). While IAM is recommended in the survey guidelines, most countries implemented some sort of MM design by the third wave (cf. European Commission, 2020, 2022).

As outlined in the section 3, sensitive topics, like health status, are especially prone to measurement errors, as the presence of an interviewer may induce socially desirable response behaviour. Braekman et al. (2020) investigated this issue using data from the Belgian Health Interview Survey 2018 that used a face-to-face IAM (with a paper-based SAM component for sensitive topics). They compared it to data obtained through a web-based SAM survey that was administered at the same time but differed in other design features as well (e.g., incentives, recruitment, sampling). Even when adjusting for sociodemographic differences in net samples, logistic regressions showed significant differences in prevalence rates of 9 of the 18 indicators collected in IAM compared to the web-based SAM. Once again, results were more comparable between the two paper-based SAM component of the IAM and the web-based SAM. While this suggests differences in observed means, it does not necessarily mean, that variances and covariances between modes differ. In fact, Zager Kocjan et al. (2022) find evidence for scalar invariance for three common measures for psychological functioning[5] in the third wave of the Slovenian sample of EHIS after controlling for self-selection propensities to either the face-to-face IAM or the web-based SAM, web design. In line with previous findings, only latent means systematically differed with respondents reporting better psychological

[5] The Patient Health Questionnaire Depression Scale (PHQ-8), the Satisfaction with Life Scale (SWLS), and the Mental Health Continuum - Short Form (MHC-SF)

functioning in the face-to-face setting, indicating a social desirability bias. Again, this issue can be addressed through ex-post harmonization.

4.5 Generations and Gender Survey (GGS)

The Generations and Gender Programme (GGP) is a social science infrastructure that was implemented to provide internationally comparable data on family and demography in 2000. Its centrepiece is the Generations and Gender Survey (GGS), a longitudinal panel survey that has fielded two rounds with three waves since 2004 in up to 19 EU countries. In the second round (GGS-II), a completely new panel was set up and additional countries in Central and East Asia as well as South America were added. As a panel study, it observes the same respondents over time and thus faces slightly different challenges with regards to errors or representation and errors of measurement compared to cross-sectional survey programs, e.g., panel attrition and recruitment of new participants to ensure representativity of sample composition, or having to separate within-person changes from between-person changes to disentangle different kinds of temporal changes and ensure comparability (Cernat & Sakshaug, 2021). Only one wave of the GGS-II has been conducted so far, so we will be treating it as a cross-sectional survey.

During the first round of the survey (GGS-I) two central limitations were identified: first, the decentralized structure, which resulted in large cross-country variation regarding fieldwork and questionnaire design and necessitated extensive harmonization efforts, and second, the face-to-face, IAM of data collection. Due to “the increasing amount of labor costs”, “interviewer effects and the stagnant or even declining response rate” (cf. Gauthier et al., 2023, p. 2), a pilot study of a MM design was implemented in three countries (Croatia, Portugal, and Germany) in 2017. Based on the results, the second round was implemented using an MM design including a web component and one additional mode, with face-to-face IAM being recommended. Moreover, to increase cooperation, unconditional, prepaid as well as conditional, postpaid incentives were recommended, but due to the cost, country-specific deviations were allowed. Many countries decided to field pilot studies to experiment different incentive schemes. For example, Hong Kong randomly assigned participants of a pilot study into nine different incentive groups and the group with a pre- as well as a postpaid incentive was most successful regarding response rates (Gauthier et al., 2023).

Of the 20 countries participating in the first wave of GGS-II, nine used a fully web-based SAM, six decided to implement some MM design: In Germany and Sweden, a paper-based, mail SAM was used “as a fallback plan to reduce survey nonresponse and potential recruitment or selection bias” (Gauthier et al., 2023, p. 9). In France, a telephone-based IAM was used alongside the web, and Uruguay and the Czech Republic added a face-to-face IAM.

Some evidence on mode related measurement effects was obtained using data from the push-to-web experiment in Croatia and Germany. Piccitto et al. (2022) investigated different measures of subjective well-being, their associations among each other and with different sociodemographic variables as well as the stability of these associations across modes. While mean differences were detected between the web-based SAM and face-to-face IAM, suggesting a social desirability bias, the relationships between indicators did not change. Based on these results, the authors thus conclude that a ”change from F2F to WEB mode will not lead to a need to rethink our causal or associational models” (p. 3458).

5. Scenarios

Survey programs are facing increasing difficulties to ensure data quality, especially in the face-to-face IAM due to declining response rates despite increased fieldwork efforts. This led to soaring survey costs and increasing risks associated with data collection. As a result, many survey programs started to implement experiments with SAM, mainly MM designs, or switched to these modes entirely in the hopes of reducing risks, costs, and efforts associated with data collection. This development raised serious questions about data comparability across these modes as well as across time, as most survey programs have established long-running, valuable timelines and put a lot of effort in retaining them.

Before we present different scenarios of how face-to-face IAM surveys can deal with the issues arising from this trend, we want to stress once again that both remaining in the current, face-to-face IAM as well as switching to a different mode will be associated with certain challenges and the choice of survey mode will always be a trade-off between data quality (i.e., aspects of the TSE as well as comparability across both time and mode) and practical considerations (i.e., feasibility and costs).

All mode transitions will come with (a certain degree of) increased efforts of ex ante harmonization. Furthermore, assessment and possibly mitigation of differences in errors will be necessary, even if the same mode is being retained. All things considered, these increased efforts may prove beneficial in the long run though, at least in some scenarios.

This being said, Table 4 gives an overview of the scenarios we will present in the following sections: first, remaining in a fully face-to-face IAM, second, switching to a single-mode, web-based SAM, third, switching to a single-mode, paper-based SAM, fourth, using an SAM-MM approach with a web- and paper-based component, and last, a parallel run of two independent samples, one using a face-to-face IAM and the other using the outlined SAM-MM approach. For each scenario, we will be pointing to some of the major advantages and disadvantages of the respective mode or mode combinations.

We conclude this section with some additional considerations that affect all survey modes in probability-based, international survey programs. These issues are closely linked to practical considerations of survey design. We will discuss modes of initial contact, mode choice sequences, and the use of incentives, and we will conclude with considerations on the issue of standardization vs. adaptation in the context of international survey programs.

Table 4. Advantages and disadvantages associated with certain modes as well as mode transitions.

Scenario	Advantages	Disadvantages	
		Of the mode itself	Of mode transition
Face-to-face IAM	- Established modes of contact facilitate implementation - Direct contact between **I** and **R** - Use of supporting material possible → higher complexity of questionnaire - Verbal and nonverbal feedback - Motivation may be increased - Distractions may be avoided	- High and rising cost, time consuming - **I** recruitment, training, supervision, and monitoring as a scheduling risk - Increasing fieldwork efforts necessary to retain data quality - Low flexibility for **R** regarding timing and location of response - **I**-effects and social desirability bias possible	
Web-based SAM	- Lowest cost - Fastest data acquisition - Process can be automated and documented more easily compared to IAM - High flexibility in questionnaire design regarding media use - Collection of meta-data - High flexibility for R regarding timing and location of response - Less intrusive: Reduced social desirability (especially with regards to sensitive topics) compared to IAM	- Representativity / coverage → internet accessibility/penetration rates may differ between countries - Technical difficulties may occur - Little to no control over survey situation (e.g., distractions, queries, etc.) - Lack of I support: Questionnaire length may have to be adjusted to reduce R burden - Reduced motivation may lead to satisficing	- First implementation increases efforts of ex ante harmonization - Ex post assessment of errors and harmonization to ensure comparability
paper-based SAM	- Less costly than IAM, but more fieldwork effort compared to web-based SAM - Shorter turn-over than IAM, but longer than web-based IAM - Higher flexibility for R regarding timing or response compared to IAM, not as flexible as web-based SAM - Less intrusive than IAM, similar to web-based SAM	- Little to no control over survey situation (e.g., distractions, queries, etc.) - Lack of I support: Questionnaire complexity may have to be adjusted - Reduced motivation may lead to satisficing	- First implementation increases efforts of ex ante harmonization - Ex post assessment of errors and harmonization to ensure comparability

Scenario	Advantages	Disadvantages	
		Of the mode itself	Of mode transition
SAM-MM	- Clever design and choices regarding mode of initial contact, mode choice sequence and incentives can reduce cost - More fieldwork effort and time needed compared to single-mode SAMs - Timely, as people tend to use different media → mode choice preferences may be met, increasing coverage - Little differences in measurement expected between SAM modes	- Complexity of questionnaire design has to follow most restrictive mode - Little to no control over survey situation (e.g., distractions, queries, etc.) - Lack of I support: Questionnaire length and complexity may have to be adjusted	- First implementation increases efforts of ex ante harmonization both across time and between modes - Ex post assessment of errors and harmonization to ensure comparability
Parallel run	- "Best of both worlds" regarding representation and coverage - Exogenous, time-related influences excluded - Direct comparison of modes possible - Harmonization efforts may establish a baseline for the future	- Highest cost, most time consuming - Most fieldwork efforts - Difficulties regarding implementation of SAM-MM	- First implementation increases efforts of ex ante harmonization similar to SAM-MM - Ex post assessment of errors and harmonization to ensure comparability

*Note: **I** = Interviewer, **R** = Respondent, IAM = interviewer-administered mode, SAM = self-administered mode, SAM-MM = self-administered, mixed mode, with a web- and paper-based component*

5.1 Fully face-to-face IAM

Remaining in an already established mode comes with the advantage of an easier implementation, as sampling modes and procedures are already established. With regards to fieldwork efforts, this is only true to an extent though: as we have outlined in section 2, increasing fieldwork efforts might be necessary to achieve comparable data quality with regards to coverage and response biases. Additional interviewer training, supervision and monitoring might be necessary to increase response rates by helping to motivate reluctant respondents to participate in the survey.

With regards to errors of measurement, IAMs, particularly face-to-face, have been considered the gold standard, because the direct contact between interviewer and respondent allows for verbal- and nonverbal feedback. On the one hand, this can lead to increased motivation and distraction can be mitigated, reducing cognitive burden and satisficing behaviours. On the other hand, evidence suggest that IAMs are associated with more social desirability bias compared to SAM (see section 3.1 and 3.2). Establishing good rapport between interviewer and respondent is thus central to measurement quality. This can only be achieved if interviewers are well-trained. Additional interviewer training may be needed to reduce interviewer effects, further increasing costs, and leading to risks associated with survey scheduling.

Overall, securing comparability with regards to both data quality as well as comparability over time will come at a relatively high cost in the IAM compared to single-mode SAM as well as some MM designs. Due to the changes in survey landscape, comparability over time should be assessed after data have been collected.

5.2 SAM with a single-mode, web-based design

The main advantage of the single-mode, web-based approach is the low fixed (up front) and variable (per completed case) costs as well as the fast data acquisition. Actual costs may vary, depending on the initial mode of contact, the use and mode of prenotifications and reminders as well as the incentivisation strategy. While implementing a new design and new procedures will always be associated with increased efforts, especially regarding ex ante harmonization strategies (see section 3.2.1), these processes can easily be automated and documented for future reference in a web-only approach. Regarding errors of representation, differing internet accessibility and penetration rates may lead to issues of cross-country comparability. Even though web-based SAM gives respondents the most freedoms regarding timing and location of responses, meta-analytic evidence reveals lowest response rates for this mode (see section 4.1). If a web-based SAM is considered for the data collection phase, it may be necessary to combine it with a different mode of initial contact and/or an adequate incentivisation strategy to increase response rates.

Looking at effects on errors of measurement, differences arise due to the absence of the interviewer, as outlined in section 3.1.2. While the high flexibility for respondents to choose both timing and location of their response may increase motivation, there is also no way of controlling the survey situations. Distractions may occur, leading to breakoffs or satisficing behaviours. Furthermore, queries and misunderstandings cannot be addressed in this setting. This may especially be a problem if technical difficulties or incompatibility with certain devices occur. Some authors suggested reducing survey length in (web-based) SAM to reduce respondent burden and thus increase cooperation. Results from the EVS indicate, however, that survey length may not be as

much of a problem (see section 4.2.1). Finally, SAM can be considered less intrusive and is thus associated with a reduced social desirability bias, especially regarding sensitive topics (see section 4). Overall, increased ex post harmonization efforts may be expected if a change to (web-based) SAM is being performed to assess and mitigate differences in errors over time.

These efforts may be helpful in the future though, as implementing at least a web-based component allows for more flexibility on many different levels, including the questionnaire design with the option of adding different types of media content and obtaining detailed information on meta-data.

5.3 SAM with a single-mode, paper-based design

As in web-based SAM, the paper-based design comes with increased implementation efforts and efforts of ex ante harmonization upon first execution. Overall, paper-based SAM is less costly compared to IAM but entails more fieldwork effort and longer-turn turn-over compared to web-based SAM.

Similarly, paper-based SAM shows higher flexibility for respondents to choose the timing of responses compared to IAM, but it is not as flexible as web-based SAM in that having to mail back the questionnaire may be an additional obstacle to some participants, and respondents cannot choose the location of response as freely. The latter may be of help for errors of measurement though, as it may reduce distractions compared to web-based SAM. However, the lack of interviewer support may be even graver for single-mode, paper-based SAM, as filters and other more complex designs may not be feasible on paper, or they may lead to misunderstandings and satisficing behaviours. Complexity of questionnaires may thus have to be reduced, without having the advantage of including more modern types of media (e.g., sounds and videos) that comes with the web-based SAM or the flexibility of interviewer materials in IAM. Just as in the web-based SAM, paper-based SAM may come with an increase in ex post harmonization efforts to ensure comparability over time.

5.4 SAM-MM with a choice of web- and paper-based participation

As there are several different ways of conducting a mixed-mode survey (see section 3.1.2), differentiating the effects of every possible scenario would go beyond the scope of this paper. This scenario will thus focus on some of the most relevant considerations when switching to an SAM-MM design with a web- and a paper-based component for data collection.

Mixing data collection modes can be considered timely, as people tend to use different types of media and mode choices of different (groups of) participants may be met, thus leading to higher coverage and lower unit nonresponse. A clever design and choices regarding mode of initial contact, mode choice sequence, and incentivisation strategies may even achieve this, while at the same time reducing cost compared to IAM. In fact, Wolf et al. (2021) documented savings of over 50% in the MM compared to the IAM in the parallel run conducted in the EVS in Germany 2017 (see section 4.2.1). As with the implementation of any SAM in a long-running survey program, preparation and ex ante harmonization efforts will increase. However, fieldwork time may decrease: Wolf et al. (2021) documented a reduction from 6 months in the IAM down to 6 to 8 weeks in the MM design.

While little differences are to be expected regarding measurement between the two SAM in SAM-MM, the absence of the interviewer and the resulting changes in response behaviours outlined in the previous sections still apply, making ex post assessment and harmonization of measurement

inevitable to ensure comparability over time. First results from EVS, EHIS, and GGS suggest, that ex post harmonization is possible if differences are due to differences in means only.

5.5 Parallel run of IAM and SAM-MM

A parallel run is used as a transitional mode when switching from one survey design to another to empirically quantify effects of design changes. In a full parallel run, two completely separate samples are drawn, the first using the established face-to-face IAM, the second using the envisaged new design. As both samples are drawn at the same time, exogenous, time-related influences are excluded and causal inferences on mode effects are facilitated.

However, this scenario is by far the most time-consuming and associated with the highest costs, as is involves both the IAM with all the issues regarding increased fieldwork efforts, soaring costs outlined in the previous section, and the hurdles of implementing a new mode. Furthermore, both ex ante and ex post harmonization efforts will increase in this scenario. The trade-off of these increase efforts is a more direct comparability of modes that may be used to establish (country-specific) baselines for "converting" or harmonizing IAM and SAM data in the future.

In fact, any changes in mode design should ideally be accompanied by an experimental implementation to assess possible effects. This can also be achieved using other experimental settings, e.g., pilot studies or split-designs. A parallel run or experimental pilot study can also be implemented in selected countries only (as conducted in the EVS and GGS respectively) to reduce efforts and still obtain empirical evidence on feasibility and effects of mode switches.

5.6 Additional considerations

Clever design choices may help survey programs achieve high data quality at a reasonable cost. As outlined in section 3.3.1, standardization and (country-specific) adaptations have to be weighed up to ensure data quality is comparable across time, mode, and country. A design choice that leads to high data quality in one country may not be as successful in another country, e.g., due to internet coverage or specific survey climates and practices. For example, in the EVS, when given the option between paper-based and web-based participation, more participants opted for the paper-based in Germany compared to Iceland. However, it might not be possible to implement country-specific "ideal" designs, due to limitations of the contractor (especially in the case of centralized survey programs). Ideally, these practices are investigated and/or experimented with before the launch of a new mode, as they also affect the overall survey cost. In the following, we will point to some central considerations that should be addressed before transitioning to a new mode.

Different *modes of initial contact* can be used to increase response rates and reduce errors of measurement, namely coverage, sample composition, and unit nonresponse. While errors of measurement are theoretically possible due to differences in perceived credibility and legitimacy of surveys based on the mode of contact, these effects have not been the focus of research so far.

Practical considerations play a large role in the mode choice for initial contact, mainly the availability of certain sampling frames. While using web-based recruitment strategies would be the most cost-effective, drawing probability-based samples is still almost impossible, as random selection of potential participants cannot be realized (i.e., there is no list of email addresses of all residents in a country and/or not every household has internet access). The mail mode is thus oftentimes the preferred option, but even obtaining addresses from local registries poses a challenge in many

countries. Regarding mode of initial contact, a list of addresses is needed in every participating country and postal service has to be reliable to limit both coverage and sampling biases. Initial contact via mail can be associated with more credibility compared to web-based SAM.

To address this issue, many survey programs make use of a *push-to-web design*, i.e., using an already established mode of contact (like phone calls or existing address lists) to invite selected participants to a cheaper data collection mode (or to establish a web-based panel, as in the ESS and CRONOS). This approach may reduce fieldwork efforts by using existing channels. Furthermore, using IAMs in these settings may increase perceived legitimacy and importance of the survey (while also increasing cost).

Oftentimes, the push-to-web design is combined with a certain *mode choice sequence*, with the more expensive mode being offered only upon a first or second reminder, to increase coverage and response rates while keeping costs low. Based on meta-analytic evidence, Bosnjak (2017) concludes, that this sequential design should be preferred over a concurrent design, as evidence suggest that offering paper-based mail and web participation simultaneously may actually decrease response rates and is associated with higher costs (Stadtmüller et al., 2021). Results from ALLBUS suggest, that tailored approaches to mode sequencing based on age may help improve sample composition, i.e., offering older target persons both paper- and web-based mode at the same time while using a sequential design for younger participants (see section 4.3).

Another consideration regards the *incentivisation strategy*. While meta-analytic evidence suggests that unconditional, prepaid incentives should be preferred over promised ones (Bosnjak, 2017), experiences from ALLBUS show, that conditional, postpaid incentives can also be successful strategy to increase response rates and reduce fieldwork efforts (see section 4.3). A pilot study implemented in the GGS-II in Hongkong found a combination of both unconditional, prepaid as well as conditional, postpaid monetary to be achieve the highest response rates (see section 4.5).

References

Asimov, A., & Blohm, M. (2024). Sequential and Concurrent Mixed-Mode Designs: A Tailored Approach. *Journal of Survey Statistics and Methodology*, smae016. https://doi.org/10.1093/jssam/smae016

Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts? *Survey Methods: Insights from the Field (SMIF)*. https://surveyinsights.org/?p=9673

Blohm, M., & Koch, A. (2017). Respondent Incentives in a National Face-to-Face Survey: Effects on Outcome Rates, Sample Composition and Fieldwork Efforts, *7*(1), 89–122. https://doi.org/10.12758/mda.2013.004

Bosnjak, M. (2017). Mixed-Mode Surveys and Data Quality. In S. Eifler & F. Faulbaum (Eds.), *Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung* (pp. 11–25). Springer VS. https://doi.org/10.1007/978-3-658-15834-7_1

Braekman, E., Charafeddine, R., Demarest, S., Drieskens, S., Berete, F., Gisle, L., van der Heyden, J., & van Hal, G. (2020). Comparing web-based versus face-to-face and paper-and-pencil questionnaire data collected through two Belgian health surveys. *International Journal of Public Health*, *65*(1), 5–16. https://doi.org/10.1007/s00038-019-01327-9

Cernat, A., & Revilla, M. (2020). Moving from Face-to-Face to a Web Panel: Impacts on Measurement Quality. *Journal of Survey Statistics and Methodology*, *9*(4), 745–763. https://doi.org/10.1093/jssam/smaa007

Cernat, A., & Sakshaug, J. W. (Eds.). (2021). *Measurement Error in Longitudinal Data*. Oxford University Press.

Cernat, A., Sakshaug, J., Christmann, P., & Gummer, T. (2022). The Impact of Survey Mode Design and Questionnaire Length on Measurement Quality. *Sociological Methods & Research*. https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-2162147

Charman, C., Mesplie-Cowan, S., & Collins, D. (2024). *The post-pandemic role of face-to-face fieldworkers*. National Centre for Social Research. https://natcen.ac.uk/sites/default/files/2024-02/Survey%20future%20report%20v4.1.pdf

Cornesse, C., & Bosnjak, M. (2018). Is there an association between survey characteristics and representativeness? A meta-analysis. *Survey Research Methods*, *12*(1), 1–13. https://doi.org/10.18148/srm/2018.v12i1.7205

Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology*, *8*(3), 513–539. https://doi.org/10.1093/jssam/smz008

Daikeler, J., Silber, H., & Bošnjak, M. (2022). A Meta-Analysis of How Country-Level Factors Affect Web Survey Response Rates. *International Journal of Market Research*, *64*(3), 306–333. https://doi.org/10.1177/14707853211050916

European Commission, E. (2020). *European Health Interview Survey (EHIS wave 3): Methodological manual*. 2020 edition (re-edition). Publications Office of the European Union. https://data.europa.eu/doi/10.2785/135920

European Commission, E. (2022). *Quality report of the third wave of the European health interview survey: 2022 edition*. *Statistical reports*. Publications Office of the European Union. https://data.europa.eu/doi/10.2785/081621

European Social Survey. (2024, April 30). *Modes of Data Collection | European Social Survey*. https://www.europeansocialsurvey.org/methodology/methodological-research/modes-data-collection

European Values Study. (2020). *European Values Study (EVS) 2017: Methodological Guidelines*. *GESIS Papers: 2020/13*. https://doi.org/10.21241/ssoar.70110

Gauthier, A. H., Kong, S., Grünwald, O., Bujard, M., Caporali, A., Deimantas, V. J., Emery, T., Jablonski, W., Koops, J., Rijken, A., & Schumann, A. (2023). *Data Brief: The Generations and Gender Survey second round (GGS-II)*. GGP Technical Paper Series. https://zenodo.org/records/10220746 https://doi.org/10.5281/zenodo.10220746

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. *Wiley Series in Survey Methodology*. Wiley. https://books.google.de/books?id=HXoSpXvo3s4C

Hochman, O., Stanciu, A., & Hadjar, A. (Eds.). (forthcoming). *40 Jahre ALLBUS - Die deutsche Gesellschaft im Wandel*. Springer VS.

Hope, S., Campanelli, P., Nicolaas, G., Lynn, P., & Jäckle, A. (2022). The Role of the Interviewer in Producing Mode Effects: Results From a Mixed Modes Experiment Comparing Face-to-Face, Telephone and Web Administration. *Survey Research Methods*, *16*(2), 207–226. https://doi.org/10.18148/srm/2022.v16i2.7771

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-operation with Face-to-Face Surveys. *Survey Research Methods*, *7*(1), 1–15. https://doi.org/10.18148/srm/2013.v7i1.4736

Koch, A., & Briceno-Rosas, R. (2023). *Assessment of socio-demographic sample composition in ESS Round 10.* European Social Survey, GESIS.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. *Statistics for Social and Behavioral Sciences*. Springer.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Leeuw, E. D. de (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, *21*(5), 233–255. https://typeset.io/papers/to-mix-or-not-to-mix-data-collection-modes-in-surveys-2o6nvk404u

Leeuw, E. D. de. (2008). Choosing the method of data collection. In E. D. de Leeuw, D. A. Dillman, & J. J. Hox (Eds.), *EAM book series. International handbook of survey methodology* (Reprinted., pp. 113–135). Psychology Press.

Leeuw, E. D. de, & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, J. Eltinge, & R. Little (Eds.), *Survey Nonresponse* (pp. 41–54). John Wiley & Sons.

Leeuw, E. D. de, Hox, J., & Luiten, A. (2018). International Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data. *Survey Methods: Insights from the Field (SMIF)*. https://surveyinsights.org/?p=10452

Leeuw, E. D. de, Hox, J., Silber, H., Struminskaya, B., & Vis, C. (2019). Development of an international survey attitude scale: measurement equivalence, reliability, and predictive validity. *Measurement Instruments for the Social Sciences*, *1*(1). https://doi.org/10.1186/s42409-019-0012-x

Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., Roover, K. de, Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of

the art, and future perspectives. *Social Science Research*, *110*, 102805. https://doi.org/10.1016/j.ssresearch.2022.102805

Luijkx, R., Jónsdóttir, G. A., Gummer, T., Ernst Stähli, M., Frederiksen, M., Ketola, K., Reeskens, T., Brislinger, E., Christmann, P., Gunnarsson, S. Þ., Hjaltason, Á. B., Joye, D., Lomazzi, V., Maineri, A. M., Milbert, P., Ochsner, M., Pollien, A., Sapin, M., Solanes, I., . . . Wolf, C. (2021). The European Values Study 2017: On the Way to the Future Using Mixed-Modes. *European Sociological Review*, *37*(2), 330–346. https://doi.org/10.1093/esr/jcaa049

O'Muircheartaigh, S., Hanson, T., Hochman, O., Bautista, R., & Fitzgerald, R. (2023, June 18). *Adapting survey mode in a changing survey landscape: Experiences from repeat cross-national, cross-sectional, and general social surveys 1.* ESRA, Milan, Italy. https://www.europeansurveyresearch.org/conf2023/prog.php?sess=93#main

Piccitto, G., Liefbroer, A. C., & Emery, T. (2022). Does the Survey Mode Affect the Association Between Subjective Well-being and its Determinants? An Experimental Comparison Between Face-to-Face and Web Mode. *Journal of Happiness Studies*, *23*(7), 3441–3461. https://doi.org/10.1007/s10902-022-00553-y

Revilla, M. (2010). Quality in Unimode and Mixed-Mode designs: A Multitrait-Multimethod approach. *Survey Research Methods*, *4*(3), 151–164. https://doi.org/10.18148/srm/2010.v4i3.4278

Rybak, A. (2023). Survey mode and nonresponse bias: A meta-analysis based on the data from the international social survey programme waves 1996-2018 and the European social survey rounds 1 to 9. *PloS One*, *18*(3), e0283092. https://doi.org/10.1371/journal.pone.0283092

Singh, R. K. (2022). Harmonizing Single-Question Instruments for Latent Constructs With Equating Using Political Interest as an Example. *Survey Research Methods*, *16*(3), 353–369. https://doi.org/10.18148/srm/2022.v16i3.7916

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

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Villar, A., & Fitzgerald, R. Using mixed modes in survey research: Evidence from six experiments in the ESS. In M. Breen (Ed.), *Values and Identities in Europe: Evidence from the European Social Survey* (259-293). Routledge.

Wolf, C., Christmann, P., Gummer, T., Schnaudt, C., & Verhoeven, S. (2021). Conducting General Social Surveys as Self-Administered Mixed-Mode Surveys. *Public Opinion Quarterly*, *85*(2), 623–648. https://doi.org/10.1093/poq/nfab039

World Bank Open Data. (2024, April 26). *World Bank Open Data*. https://data.worldbank.org/indicator/IT.NET.USER.ZS?end=2021&locations=OE&start=1990&view=chart

Zager Kocjan, G., Lavtar, D., & Sočan, G. (2022). The effects of survey mode on self-reported psychological functioning: Measurement invariance and latent mean comparison across face-to-face and web modes. *Behavior Research Methods.* Advance online publication. https://doi.org/10.3758/s13428-022-01867-8

WPEF24071

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.