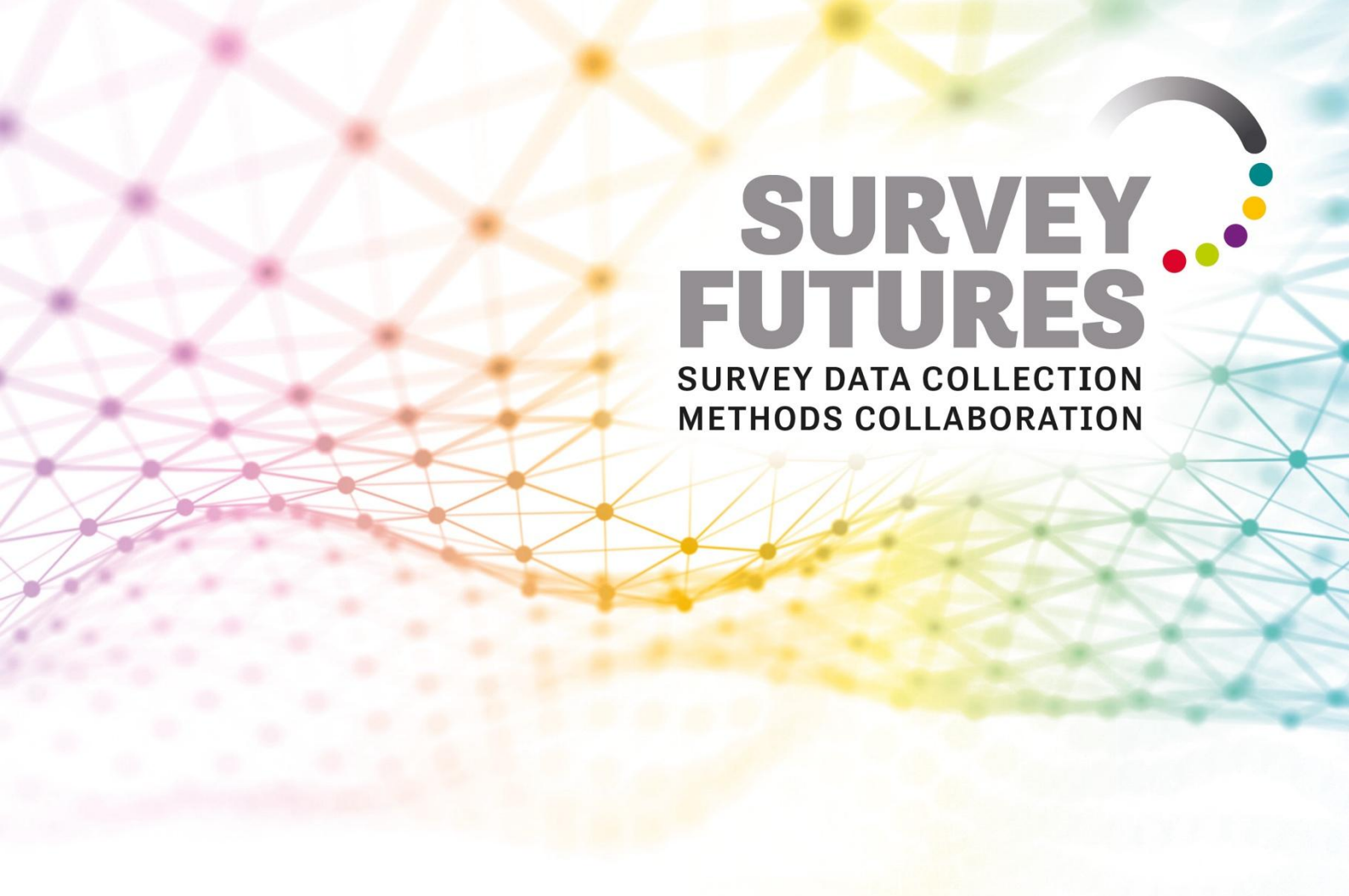




SURVEY FUTURES

**SURVEY DATA COLLECTION
METHODS COLLABORATION**

Working Paper 6:

**Still the gold standard in survey research?
Comparing face-to-face and self-completion
data collection in a repeat cross-sectional
general social survey in Great Britain**

Rory Fitzgerald¹, Nhlanhla Ndebele¹, Olga Maslovskaya², Cristian Domarchi², Peter Lynn³, Tim Hanson¹, Ruxandra Comanaru¹

¹City St George's, University of London; ²University of Southampton; ³University of Essex

October 2025

www.surveyfutures.net

Contents

Abstract	1
1. Introduction	2
2. Background and literature review	3
Response rates	3
Sample representativeness	5
Data quality	7
3. Research questions	8
4. Data	9
Methods	13
5. Results	16
RQ1: Response rates	16
RQ2: Sample representativeness	16
Data quality	20
6. Conclusions and discussion	23
Acknowledgements	24
References	25

Still the gold standard in survey research? Comparing face-to-face and self-completion data collection in a repeat cross-sectional general social survey in Great Britain

Abstract

Surveys aim to provide estimates of the behaviour, social conditions, or attitudes for the population they seek to represent. Since modern surveys of the general population were first established, the best way to collect high quality data was felt to be via face-to-face interviews amongst probability samples of households or individuals. However, more recently, face-to-face data collection in Great Britain has been impacted by declining response rates, increasing evidence of interviewer effects, rising costs and a reduction in the number of providers. At the same time, self-completion surveys in Great Britain offer an increasingly convincing alternative to face-to-face data collection, with higher levels of web penetration and digital literacy, zero interviewer effects, relative cost efficiency, as well as promising response rates and representativeness. Together these changes call into question whether the face-to-face method truly remains the 'gold standard' for surveys of the British population.

This paper compares face-to-face data collection on the 10th round of the European Social Survey in Great Britain with an experimental self-completion survey (sequential web to paper) conducted at the same time, using the same questionnaire. The self-completion approach achieved a considerably higher response rate than the face-to-face survey, slightly better representativeness, a much shorter data collection period and substantially lower cost per interview. At the same time, it was found that the self-completion survey had slightly inferior data quality on some measures. The paper concludes that self-completion data collection offers a high-quality alternative to face-to-face data collection in Great Britain, potentially becoming the new 'gold standard' in the near future for surveys that can be conducted by web and paper modes in combination.

1. Introduction

Face-to-face interviewing is one of the oldest modes of survey data collection and is often referred to as the 'gold standard' in survey methodology, due to the benefits of direct in-person interaction between the interviewer and the respondent (de Leeuw, 1992, 2008; Loosveldt, 2008; Neuman, 2012). Interviewers can play a helpful role in recruiting and motivating sampled members, administering the questionnaire according to the principles of standardised interviewing, clarifying queries, and probing inadequate responses (de Leeuw, 1992, 2008; Groves et al., 2009; Loosveldt, 2008). Compared to other modes without direct in-person contact, face-to-face interviews offer some unique advantages for collecting high-quality data.

Where appropriate individual-level sampling frames are unavailable, like in Great Britain (GB), face-to-face contact was also thought to offer the greatest potential for reducing coverage and sampling errors (de Leeuw, 2008; Groves et al., 2009). Using address-based samples, interviewers visit households, and apply procedures to randomly select eligible respondents within households (Gaziano, 2005). Interviewers can also encourage participation in lengthy interviews (de Leeuw, 2008; Neuman, 2012), guide respondents through complex questionnaires, and utilise visual and auditory stimuli, thereby reducing item non-response and premature terminations of interviews.

At the same time, the greatest asset of face-to-face interviews can also be its most significant weakness (de Leeuw, 2008). The presence of an interviewer can influence respondent behaviour when formulating their responses (Krosnick, 1991) – for example, by overreporting socially desirable behaviours and underreporting undesirable ones (de Leeuw, 2008; Dillman et al., 2014; Groves et al., 2009). Variations among interviewers in the application of standardised interviewing may further introduce response variance and affect survey statistics (Dillman et al., 2014; West & Blom, 2017). In their study on interviewer effects within the context of the European Social Survey (ESS), Beullens and Loosveldt (2016) found that failing to account for these effects led to an overestimation of effect sizes in relationships between survey variables and an underestimation of standard errors. Furthermore, face-to-face interviews are often more expensive, time-consuming and have a greater environmental impact than surveys administered in other modes, with interviewer training and travel as well as the actual interview time escalating costs, especially when there are geographically dispersed samples (Ibid).

Different intertwined factors, reflecting broader methodological, societal, and technological changes, have likely contributed to declining response rates in social surveys (Maslovskaya et al., 2025). Societal changes have introduced challenges that have exacerbated non-response in face-to-face data collection. A trend towards smaller households (Dixon & Tucker, 2010; Stoop, 2005), increased workforce participation – particularly among women (Tucker & Lepkowski, 2007) – and a rise in single individuals living and working in urban areas (Goyder, 1987) have all reduced the likelihood of someone being at home for an interview, thereby increasing non-contact rates (Durrant & Steele, 2009). Although computer-assisted interviewing has improved questionnaire administration, these technological innovations require interviewers to learn new technical skills, and can make it harder to retain experienced

interviewers who may be less technically minded (Beullens et al., 2018; Dixon & Tucker, 2010; Schaeffer et al., 2010). More recently, the COVID-19 pandemic significantly disrupted face-to-face data collection. It accelerated the decline in response rates; for example, ESS face-to-face response rates in the UK fell to 27% in Round 11 (2023) (Fitzgerald, 2024), while in GB, after returning to face-to-face data collection in late 2023, the Labour Force Survey (LFS) response rates at Wave 1 had fallen to 34%. In GB, the pandemic led to the retirement of many experienced interviewers and agencies have been struggling to recruit replacements (Charman, Mesplie-Cowan & Collins, 2024). The strengthening of casual workers' rights is also posing a challenge to a sector that has long relied on freelancer employment and paying interviewers per completed interview. In addition, public attitudes towards doorstep visits may have changed irreversibly, with people becoming more socially averse, which has implications for face-to-face fieldwork (Charman et al., 2025; Smith, 2020). These factors have led to a diminishing pool of face-to-face interviewers in GB that have to be shared amongst the agencies still offering this mode of collection (Charman et al., 2025). Furthermore, the COVID-19 pandemic has driven large parts of the population to become more accustomed to online activities, and the net effect may be an acceleration of the ongoing shift from offline to web-based methodologies (Smith, 2020).

2. Background and literature review

The challenges to face-to-face surveys outlined above led a number of major UK surveys to consider moving to self-completion as either the predominant or the sole mode of data collection (Brown & Hancock, 2015; Cabinet Office, 2016; Jäckle et al., 2015; NatCen Social Research, 2022; ONS, 2025). This in turn has spawned significant experimentation that has enabled comparisons of the relative success of the two approaches. As a result, the data collection infrastructure in GB has undergone a paradigm shift. Social surveys are experiencing major transformations in their design and implementation.

Response rates

Historically, face-to-face surveys have achieved higher response rates than other data collection modes (de Leeuw, 1992, 2008; Dillman et al., 2014; Groves et al., 2009). However, face-to-face response rates have declined in recent times both in GB and internationally. In the United States, Williams and Brick (2018) observed the rate with which face-to-face response rates were declining increased from 0.5 percentage points annually between 2000 and 2005 to 1 percentage point annually between 2006 and 2014, on average, across several face-to-face surveys. In Europe, Beullens et al. (2018) analysed data from 35 countries across seven face-to-face rounds of the ESS between 2002 and 2014, identifying a trend of decreasing average response rates. This trend has continued, with the UK response rates for the face-to-face ESS falling to 41% in Round 9 (2019) – the last round before the COVID-19 pandemic – compared to 51% in Round 2 (2004) (Fitzgerald, 2024). Similarly, de Leeuw et al. (2018) reported an average annual decline of 0.73 percentage points in the Labour Force Survey (LFS) response rates across 27 countries between 1980 and 2015. In GB specifically, the LFS face-to-face response rate dropped from 75% in 2004 to below 55% in 2018 (ONS, 2020a) and by 2020

had fallen to 45% at Wave 1, before face-to-face data collection was suspended temporarily due to the COVID-19 pandemic (ONS, 2020b).

Even prior to the COVID-19 pandemic, surveys had begun moving either towards online-only or mixed-mode data collection, as a response to the increasingly rapid decline in response rates in face-to-face surveys, as well as the surge in their fieldwork costs. This move was possible due to growing internet penetration and use. The pandemic then created an urgent necessity to move some surveys to self-completion modes. The surveys which were already preparing for the transition expedited existing transformation plans and implemented the new methodology for data collection to circumvent the restrictions imposed by the pandemic. Several large-scale surveys in the UK such as the LFS, the National Survey for Wales, and the British Social Attitudes (BSA) Survey, made the transition using telephone interviewing during the early stages of the pandemic and later moved to mixed-mode strategies, which included a large component of online self-administration. Despite decreasing face-to-face response rates, several studies found that even lower rates were experienced after switching to these other approaches. For example, the BSA decreased from a 44.3% response rate in its last face-to-face data collection operation in 2019 (NatCen Social Research, 2020) to 16.1% in their survey conducted in 2023, that primarily used the online mode (NatCen Social Research, 2024). Another UK survey, the Community Life Survey, conducted an experimental comparison of a self-completion (web plus paper) protocol with the previous face-to-face design of the survey in 2012. Response rates for the self-completion survey ranged from 27% with no incentive to 39% with a £5 unconditional incentive, while the face-to-face survey achieved 60% response (TNS BMRB, 2013). A larger-scale test in 2013-14, with paper questionnaires sent only upon request, achieved 27.6% response (TNS BMRB, 2014b), compared to 61% achieved by the main face-to-face survey in the same year (TNS BMRB, 2014a).

In 2015, the Office for National Statistics (ONS) embarked on the Census and Data Collection Transformation Programme (CDCTP), which aimed to make data collection more dynamic and efficient by increasing the use of non-survey data sources and by adopting digital-by-default data collection. A centrepiece of this programme was an ambition to turn the LFS, which in 2015 consisted of 75,000 face-to-face interviews, into a primarily online survey. Two large-scale tests of an online-first design were conducted in 2017. The first (Ipsos MORI, 2018a) achieved an overall response rate of 19.9% of households (17.2% of households gave a full response where responses to demographic and employment questions were provided for all household members). This test included built-in experiments with various design features expected to affect response rates, (e.g., the combination of invitation/reminder letters, day of mailing, envelope colour, country-specific branding in Scotland and Wales), but any effects found were small: the response rate did not exceed 22% with any combination of the features. The second test (Ipsos MORI, 2018b), with a revised questionnaire, experimented with different forms of incentives. It gained responses from 25.2% of households overall, ranging from 19.4% without any incentive to 27.0% with a combination of £5 unconditional and £10 conditional incentive upon the whole household completing. At this time, the equivalent response rate for first wave interviews on the LFS, carried out face-to-face, was 57% (ONS, 2017), despite the interview being longer than on the experimental online-first version of the survey. Thus, at that time – even with incentives, which the face-to-face LFS did not previously

use – the online approach appeared to perform much worse than face-to-face in terms of response rate.

Drawing on lessons from these two LFS tests and other development work, ONS conducted the Labour Market Test Survey in 2018-19, with a new mixed-mode design in which a push-to-web approach was followed-up with face-to-face contacts with the remaining non-respondents. The push-to-web phase achieved 28.4% household response (ONS, 2020c) compared to 55.0% household response to the face-to-face LFS in approximately the same period (ONS, 2019).

A recent review of cross-sectional general population surveys with a self-completion component conducted between 2018 and 2024 in the UK indicated that response rates for self-completion only surveys ranged from 7% to 36% (Domarchi et al., forthcoming). In addition, most ESS countries that have implemented the ESS self-completion approach in preparation for the full mode switch planned for 2027, have achieved response rates of between 30% and 40% (Fitzgerald, 2024). This evidence might be an indication of a paradigm shift, but more corroborating studies are needed.

Given the literature, *we expect the face-to-face response rate in our GB comparison to be higher than for the self-completion survey.*

Sample representativeness

When evaluating the quality of a social survey, methodologists are not only concerned with the response rates achieved, but also with the representativeness of the achieved sample (Dillman et al., 2014; Groves et al., 2009). In the context of social surveys, *representativeness* refers to the degree to which a survey sample accurately reflects the characteristics of the target population and is vital for ensuring that it is possible to generalise findings (Babbie, 2016; Ormston et al., 2024). There is little published information in technical reports comparing the unweighted sample composition of face-to-face and self-completion surveys conducted within the same studies in GB or the UK. However, some insights are available in the literature. For example, in an experimental study evaluating the push-to-web methodology for mixed-mode surveys using address-based samples in the UK, Lynn (2020) compared the composition of a self-completion sample in a web-first protocol with those who participated in a CAPI interview. This study used data from the recruitment wave for a new sample to refresh the Innovation Panel, part of Understanding Society: The UK Household Longitudinal Study, which is designed for methodological testing and development. The results suggested that older respondents (aged 60 or above) were less likely to participate online than in face-to-face interviews, compared to those under 60. Similarly, respondents with no educational qualifications or less than degree-level education were less likely to respond online than those with degree-level education or higher. These findings were supported by Ormston et al. (2024) who conducted a review of key UK surveys – though some international surveys were also included – examining the impact of mode changes on non-response bias. The evidence suggested that self-completion surveys were more likely to underrepresent older individuals and those with lower levels of education compared to face-

to-face surveys (Ormston et al., 2024). In terms of sex and urbanicity, Lynn (2020) found no significant differences in participation by mode.

Wolf et al. (2021) compared the sample composition of face-to-face and self-completion mixed-mode (web and mail) surveys conducted as part of the 2017/18 European Values Study in Germany. They found no evidence of underrepresentation by sex in either mode when compared with population statistics. On the other hand, for age, the self-completion mode underrepresented respondents aged 18-29 and overrepresented those aged 60-69, while face-to-face overrepresented respondents aged 60-69 (to a higher extent than in self-completion) and underrepresented those over 70. Wolf et al. (2021) also found that both modes significantly underrepresented the less educated (perhaps surprisingly more pronounced in face-to-face) and overrepresented the highly educated (more pronounced in self-completion). Finally, Wolf et al. (2021) found that the face-to-face mode performed slightly worse than the self-completion mode by underrepresenting the urban population. This finding aligns with other evidence suggesting that residents of large cities may be harder to reach through face-to-face methods than through self-completion surveys (Haan et al., 2014).

Although males and young people tend to be underrepresented in social surveys, this pattern is generally consistent across survey modes (Haan et al., 2014; Ormston et al., 2024). Studies have found no significant differences in sample composition by sex across modes (Lynn, 2020; Wolf et al., 2021). In contrast, evidence suggests young people are more underrepresented in self-completion surveys than in face-to-face surveys (Wolf et al., 2021), older people are more likely to participate in face-to-face than in self-completion surveys (Lynn, 2020; Ormston et al., 2024) and respondents with lower levels of education are less likely to participate in self-completion surveys (Lynn, 2020; Ormston et al., 2024). Evidence on mode-related effects for urbanicity is generally quite mixed (Lynn, 2020; Wolf et al., 2021).

Based on this literature, we expect that the *overall sample representativeness in the face-to-face survey will be better than in self-completion*. Specifically, we expect:

- *there will be no mode differences in representativeness by sex.*
- *the face-to-face survey will overrepresent older respondents and underrepresent younger respondents, but the self-completion survey will underrepresent younger respondents even more.*
- *the self-completion survey will underrepresent the lower educated population compared to the face-to-face survey.*
- *the face-to-face survey will underrepresent respondents in large cities compared to self-completion surveys.*

Data quality

When it comes to data quality little information is provided in the published technical reports of social surveys in GB. In the literature there are mixed findings in terms of the effects of survey mode on *item non-response*. It is worth noting that the way in which item non-response is handled may differ between face-to-face and self-completion surveys. In the face-to-face surveys, it is typical for ‘*don’t know*’ and ‘*refusal*’ codes to be available to interviewers but they are rarely visible or prompted to respondents. This is the approach used in the face-to-face ESS. In self-completion surveys, these response options may appear on the questionnaire (visible to respondents), be prompted when respondents try to leave a question blank (in web surveys only), or not appear at all, with respondents sometimes informed that they can leave any question blank if they prefer not to answer (this latter approach is used on the self-completion ESS). These variations in presentation can lead to differences in item non-response levels.

A systematic review and meta-analysis reported in Čehovin et al. (2023) found significantly lower mean item non-response rates for web surveys compared with paper-based self-completion surveys – however, they did not detect any significant differences in item non-response in web surveys compared to face-to-face interviews. Conversely, Jäckle et al. (2015) found significantly higher item non-response rates in a mixed-mode design including both online self-completion and face-to-face interviews, compared with a single-mode face-to-face survey. These findings originate from an experiment in the context of a longitudinal survey in which participants had previously been interviewed face-to-face. Meanwhile Jäckle et al. (2015, p. 58), found it would generally be expected that item non-response rates will be higher in self-completion web surveys compared to interviewer-administered surveys, “*unless the combination of the nature of the question and the design of the web instrument is particularly favourable*”, in which case the rates may approach those achieved in face-to-face surveys. Given the literature and the nature of the ESS questionnaires discussed in this paper *we expect to observe higher item non-response rates for the self-completion survey compared to the face-to-face one*. We examine item non-response in the education variable separately due to the complexity of its categorisation. Since prior research has not analysed item non-response in this variable, nor specifically examined mode differences for education questions, this paper provides unique evidence on whether such differences can be expected when collecting education information.

Some studies in the literature find that *straightlining* (Krosnick & Alwin, 1988; Reuning & Plutzer, 2020) is more common in self-completion web surveys than face-to-face surveys (e.g. Heerwegh & Loosveldt, 2006) as it is related to respondents speeding through questions in web surveys (Zhang & Conrad, 2014). However, there are also mixed and opposing findings. One of the experiments reported in Villar and Fitzgerald (2017) compared the implementation of a mixed-mode design in the ESS including web-based interviewing as one of the modes, to a conventional face-to-face interview in three countries (Estonia, UK, and Sweden). The authors found that the proportion of straightlining respondents was, in general, similar or lower in the mixed-mode design than in the standard ESS face-to-face interview. When looking at respondents by mode, the authors found no evidence of higher straightlining among web

respondents when compared to face-to-face mode. Given the literature *we do not expect differences in straightlining between the face-to-face and self-completion ESS surveys in GB.*

The limited literature comparing *internal consistency* across survey modes suggests minimal differences (Revilla and Saris 2013; Revilla 2015). Considering this limited evidence, *we do not expect significant differences in consistency scores between the face-to-face and self-completion surveys.*

Surveys typically collect information about *occupation* through open-ended questions asking participants to provide their job title and describe their work. These questions are designed to gather sufficient data for post-interview coding by specialists using either manual or computer-aided coding procedures (Peycheva et al., 2021). Such open-ended questions can be administered in both interviewer- and self-completion survey modes. Previous research indicates that coding reliability depends primarily on two factors: (1) the quality of the information provided by the respondent in the open-ended questions (Conrad et al., 2016) and (2) the coding methods employed (Kocar et al., 2023). However, the literature provides no evidence that the mode of survey administration significantly affects the quality of occupation codes. While previous studies have noted that occupation descriptions may be shorter in self-administered modes, potentially influencing coding rates (Conrad et al., 2016), emerging evidence suggests that office coding can achieve comparable results across modes, with no significant differences observed (Kocar et al., 2023). This paper contributes to that growing body of evidence. *We therefore expect no meaningful differences between the self-completion and face-to-face as identical coding procedures are applied to the occupation variable* (Conrad et al., 2016; Kocar et al., 2023).

3. Research Questions

The following three research questions (RQs) will guide our comparative analysis of face-to-face and self-completion data collection among a probability sample of GB respondents in a general population social survey. Based on evidence from literature and practice, as reported above in section 2, we have outlined below our hypotheses for each RQ.

RQ1: Can a self-completion general social survey in GB now achieve a *response rate* as high as a face-to-face survey?

H1: The response rate for the face-to-face survey will be higher than the self-completion survey response rate.

RQ2: Can a self-completion general social survey in GB using probability sampling be as *representative* as a face-to-face survey?

H2: Overall, the face-to-face representativeness will be better than the self-completion.

H2.1: No mode-related differences will be found in representativeness by sex.

H2.2: The face-to-face survey will over-represent older respondents; younger respondents will be under-represented in both, but more so in the self-completion survey.

H2.3: Lower educated people will be under-represented in the self-completion survey compared to the face-to-face survey

H2.4: Face-to-face surveys will under-represent respondents in large cities compared to self-completion surveys.

RQ3: Does *data quality* differ between the self-completion and face-to-face modes and if so, how?

H3: The overall data quality in self-completion will not differ from face-to-face.

H3.1a: The item non-response for face-to-face will be lower than for self-completion.

H3.1b: Regarding the education question, there will be no difference between the two modes in item non-response.

H3.2: There will be no difference in straightlining between face-to-face and self-completion.

H3.3: There will be no difference in the proportion of open responses that can be coded to the occupation variable between face-to-face and self-completion modes.

H3.4: There will be no difference in the internal consistency of items between face-to-face and self-completion.

4. Data

We use the European Social Survey (ESS) Round 10 sample from GB, collected face-to-face (European Social Survey European Research Infrastructure Consortium (ESS ERIC), 2023) and compare that to a self-completion survey conducted for part of the same fieldwork period (Hanson et al., 2024). The two surveys used an almost identical questionnaire. The ESS is a biennial cross-national general social survey conducted in over 30 countries in Europe. The self-completion data were collected as part of an experiment, investigating the practicalities of transitioning the ESS from a face-to-face to a self-completion survey (Hanson, 2023) which built upon earlier development of a self-completion ESS, triggered by the restrictions on in-person contact due to the COVID-19 pandemic (Fitzgerald & Sibley, 2021).

The samples used for both surveys were drawn from the UK postal address file (PAF). However, the sampling frame for the face-to-face survey included addresses from Northern Ireland, whereas the self-completion survey was restricted to addresses in GB (England, Scotland, and Wales). Furthermore, the target population for the face-to-face survey was those aged 15 years old or over, while for the self-completion survey it was those aged 18 years and over. Therefore, for the analysis in this article, respondents from Northern Ireland and those under 18 years old were excluded from the face-to-face achieved samples¹.

There were differences in the sampling and study designs between the face-to-face and self-completions surveys. Both surveys used stratified random sampling designs; however, the self-completion survey employed an un-clustered design, whereas the face-to-face survey used a clustered design, to minimise interviewer travel time. The table below shows the design effects (DEFF) for the two surveys.

Table 1 : Design effects for face-to-face and self-completion surveys

Mode	<i>n</i>	<i>DEFF_p</i>	<i>DEFF_c</i>	<i>DEFF</i>	Effective <i>n</i>	Efficiency (%)
					<i>n</i> / <i>DEFF</i>	Effective <i>n</i> / <i>n</i>
Face-to-face	1125	1.193	1.091	1.302	864	77
Self-completion	2908	1.263	1.000	1.263	2302	79

For the UK, the ESS aims to achieve an effective sample size of 1500. Based on *overall DEFF* the face-to-face survey would require a sample of 1953 (i.e. $1500 * 1.302$) compared to 1895 (i.e. $1500 * 1.263$) for the self-completion survey.

Although both designs showed some loss of precision, the self-completion survey was marginally more efficient (79% v 77%), with the reduction in precision attributable mainly to unequal weighting. In contrast, the loss of efficiency in the face-to-face survey reflected the combined effects of clustering and weighting. *Overall, the results suggest that the self-completion survey was slightly more efficient than the face-to-face survey.*

For the face-to-face survey, field interviewers recruited participants at the sampled addresses once contact was made. In contrast, the self-completion survey used postal invitations sent to the sampled addresses. This also influenced the within-household selection methods used in each survey. In both cases, a single respondent was selected from each sampled address. For the face-to-face survey, the Kish grid method was implemented by the interviewer in households with more than one eligible person. For the self-completion, survey households were asked to employ the next birthday method, to identify one adult to participate. An

¹ For response rates calculations addresses in Northern Ireland were removed from the gross data however it was not possible to remove sample units where those aged 15-17 were the target respondent as this information is not included on the sample frame.

instruction was included in the invitation and reminder letters that the person aged 18+ with the next birthday in the household should complete the survey. There were also checks at the start of the questionnaire that the person with the next birthday was responding. A key advantage of the Kish grid is its potential to minimise selection bias (Díaz de Rada, 2021; Olson et al., 2019). However, there is little information in the literature on how frequently interviewers implement this correctly. In contrast, the next-birthday method is prone to inaccuracies and may introduce selection bias in households with multiple eligible members, particularly when variables of interest are related to birth dates or seasons of the year (Gaziano, 2005; Smyth et al., 2019). Existing literature suggests that rates of accurate selection for birthday methods, excluding single-person households, ranged from the mid-50% to mid-70% (Olson et al., 2014; Olson & Smyth, 2017; Smyth et al., 2019; Stange et al., 2016; J. Williams, 2015).

Fieldwork for the face-to-face survey was conducted by a large UK social research agency, combining CAPI and live video interviewing (LVI) (used as a back-up). Most questions made use of showcards, which did not include '*don't know*' or '*refusal*' codes. Respondents were informed at the start of the interview that they did not need to answer questions they did not want to. All sampled addresses were sent an advance letter to notify them that an interviewer would soon visit, which included a £5 unconditional gift voucher. In the last few months of fieldwork, due to the low response rate achieved, interviewers offered discretionary £25 gift vouchers to all cases where contact was made.

In contrast, the self-completion data collection was conducted by a small UK research agency in collaboration with ESS ERIC Headquarters. Sampled addresses received up to four mailings, starting with an invitation letter to complete a web questionnaire, which included a £5 unconditional cash incentive (bank note). Subsequent mailings were sent only to non-respondents. A first reminder was sent one week after the main launch invitation letter, a second reminder sent two weeks later (including a paper questionnaire), and a survey relaunch letter (which did not refer to the earlier mailings) sent six weeks after the second reminder. Whilst efforts were made to optimise the web questionnaire for mobile use where possible and to ensure the paper questionnaire was easy to use, comparability with the face-to-face questionnaire was also prioritised. Respondents were informed at the start of the interview that they did not have to answer questions they were not comfortable answering and could leave them blank (i.e., there were no specific '*don't know*' and '*refusal*' codes). Those who had not responded by the time of the survey relaunch received an additional unconditional £5 gift voucher. A three-group incentive experiment was incorporated into the design of the study, with each group receiving a conditional gift voucher of £10, £5 or no incentive, respectively.

In terms of questionnaire length, the face-to-face questionnaire included a section with the 21-item Human Value Scale, which the self-completion questionnaire excluded. When this section was excluded from the analysis, respondents spent approximately the same amount of time on average completing the questionnaire, regardless of whether it was administered by an interviewer or self-completed online. Specifically, the median completion time for the face-to-face interview was 51 minutes (95% CI [50.02, 51.98]), compared to 50 minutes (95% CI [49.26, 51.41]) for the web self-completion questionnaire.

ESS Round 10 was conducted after the easing of COVID-19 pandemic restrictions and end of national lockdowns (beginning in August 2021). The timing of this round had a significant impact on the effectiveness of the face-to-face approach. The fieldwork period was considerably longer for the face-to-face survey (lasting over 54 weeks) compared to the self-completion survey (nearly 14 weeks). Issues that affected the fieldwork were interviewer capacity, as well as respondents' reluctance to participate in an in-home interview. A contingency plan was instated to offer Live Video Interviewing (LVI).

The table below compares fieldwork characteristics for ESS Rounds 9–11 in the UK, all conducted face-to-face. It provides evidence that while Round 10 was a particularly challenging round, it was not a complete outlier. For example, the response rate for Round 11 was only six percentage points higher and the proportion of non-contacts to receive the required four visits was lower at Round 11 compared to Round 10. This suggests the cooperation rate had recovered slightly but underlines that capacity issues delivering face-to-face interviewing remain.

Table 2: ESS UK^a response rates Rounds 9–11 face-to-face

	Round 9	Round 10	Round 11
Fieldwork dates	31/08/2018 to 22/02/2019	15/08/2021 to 02/09/22	03/07/2023 to 09/12/2023
Fieldwork duration	25 weeks	54 weeks	22 weeks
Number of interviewers used	312	233	228
Non-contact rate (%)	7.2	11.4	8.9
% of non-contacts to receive 4 visits	71.8	65.6	56.8
Response rate ^b (%)	41.0	20.9	27.0

Notes:

a: We use UK rather than GB figures here since the ESS publishes UK rates, and Northern Ireland makes up a very small part of the sample, meaning that GB rates would be very similar.

b: The published ESS response rates differ slightly from AAPOR definitions but are closely aligned with AAPOR RR2.

Both surveys employed mixed-mode designs for data collection, to ensure better data quality and representativeness of the sample. In the face-to-face survey, data were primarily collected through CAPI, with a small proportion obtained via LVI. For the self-completion survey, around three quarters of responses were submitted through a web questionnaire, while a around a quarter of respondents used a paper questionnaire (see table below).

Table 3: Response by mode to ESS Round 10 face-to-face and experimental self-completion surveys (Great Britain)

Mode	Response mode								Total valid	
	CAPI		LVI		Web		Paper			
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Face-to-face	1082	96	43	4					1125	100
Self-completion					2120	73	788	27	2908	100

Notes: Unweighted estimates.

Methods

RQ1: Response rates

The response rates presented in this paper were calculated based on the American Association of Public Opinion Research (AAPOR) definitions. Response Rate 4 (*RR4*) was used as it accounts for complete and partially complete responses while also incorporating an estimate of unknown eligibility (eligibility rate) for nonrespondents (AAPOR, 2023). The eligibility rate is a property of the sampling frame and, under expectation, therefore takes the same value for both data collection modes. As substantial amounts of unknown eligibility are a common feature of self-completion surveys, it is the most appropriate AAPOR rate to use, whilst also providing the best basis for comparison to the face-to-face survey. Based on data from the ESS Round 10 face-to-face survey for GB, the eligibility rate was estimated at $e = 0.93$.

There were differences between the face-to-face and self-completion surveys in how partially complete responses were defined. For the self-completion survey, a response was considered usable if at least 75% of the ‘ask all’ questions were answered (Hanson, 2023). In contrast, there was no fixed rule on inclusion of usable cases for the face-to-face survey. However, only 2 cases in the dataset answered fewer than 75% of ‘ask all’ questions.

RQ2: Sample representativeness

To assess the representativeness of the face-to-face and self-completion samples, demographic characteristics (sex, age group, education, and region) were compared with external population benchmarks. For sex, age group, and region, the benchmarks were based on the GB mid-year population estimates for June 2021 (ONS, 2024). For education, the comparison was limited to respondents in England and Wales, using population estimates from the 2021 Census for people aged 18 and over (ONS, 2022). This excluded respondents from Scotland as the measurement of the highest level of education differed from that used

in England and Wales and could not be easily mapped to the International Standard Classification of Education (ISCED). The year of the population benchmark estimates was chosen because it most closely aligned with the data collection period for both the face-to-face and self-completion surveys.

Estimates for each survey mode accounted for the sampling design, specifically clustering and stratification in the face-to-face survey, and stratification in the self-completion survey, and the application of design weights to adjust for unequal probabilities of selection. Each estimate included 95% confidence intervals, which were used to assess whether differences from the population benchmarks were statistically significant.

RQ3: Data quality

We assessed data quality across four dimensions: item non-response, straightlining, internal consistency, as well as the availability of information to support coding of education and occupation.

Independent two-sample *t*-tests were used to compare means by mode for continuous variables. For categorical variables, the χ^2 test was used to compare distributions by mode. The *t*-tests and χ^2 tests were based on unweighted samples due to differences in sampling designs between the two surveys. Where weighted estimates were used, the estimates for each survey mode accounted for the sampling design, and design weights were applied to adjust for unequal probabilities of selection.

We compared the overall item non-response between the two surveys, combining the 3 separate codes (*'refusal'*, *'no answer'* and *'don't know'*) used in face-to-face, calculated as the proportion of the questions respondents did not answer, out of all the 138 questions that were asked to all participants. Only the questions which were the same in both surveys, were included in the calculation². We also assessed item non-response in the question about the highest level of education completed by the respondent (*"starting from the top and moving down the list, please select the highest level of education you have completed from these options"*).

To compare the extent of straightlining, we analysed two attitudinal scales available in both questionnaires – the trust in institutions scale (8 questions), and the evaluations of democracy scale (10 questions). Both were measured as 11-point Likert scales. For each respondent in each scale, we calculated a straightlining score, defined as the average deviation from the previous answer (Loosveldt et al., 2018; Maslovskaya et al., 2024). The score for each participant was calculated as follows:

² The face-to-face questionnaire contains 29 additional "ask-all" questions not included in the self-completion version. Consequently, our analysis covers 100% of the self-completion questions and 82.6% of the face-to-face questions.

$$S_i = \frac{\sum_{q=2}^{N_{iq}} \text{abs}(X_q - X_{q-1})}{N_{iq}},$$

where X_q and X_{q-1} are the scores provided by the participants for questions q and $q-1$, respectively, while N_{iq} is the number of questions for which participant i provided a valid score. A lower value in this score is associated with low variability in the responses, and consequently with a higher tendency for *straightlining*. We compare mean values of the straightlining scores across the two surveys. The calculation only includes cases where all the relevant items have been answered.

We assessed the potential effects of survey mode for *occupation coding*. Occupation data was collected with three open text questions (“*what is the name or title of your job*” / “*in your main job, what kind of work do you do most of the time*” / “*what training or qualifications are needed for the job?*”). Coding in each survey was conducted by different organisations with different levels of experience of occupational coding, and thus some differences in the results might be expected. We assessed the quality of occupation coding by defining a dichotomous variable equal to 1 when an occupation code could be assigned for the participant, and 0 when the occupation code is missing.

We also evaluated *internal consistency* using *coefficient alpha* (α , or Cronbach’s α), a commonly used measure of reliability for a set of indicators (Baldwin, 2019; Reise et al., 2013). Values range from zero to one, with higher values indicating greater internal consistency between the indicators (Finch & French, 2015). We assessed the internal consistency of responses in the face-to-face and self-completion surveys across a set of three questions measuring attitudes toward gays and lesbians. Specifically, respondents were asked: “*Gay men and lesbians should be free to live their own life as they wish*”; “*If a close family member was a gay man or a lesbian, I would feel ashamed*”; and “*Gay male and lesbian couples should have the same rights to adopt children as straight couples*”. These items were measured on a 5-point Likert scale ranging from ‘*strongly agree*’ to ‘*strongly disagree*’. As these had ordinal levels of measurement, Cronbach’s α was estimated from their polychoric correlation matrix. We applied the Feldt’s test of equality to test whether the difference in Cronbach’s α between the face-to-face and self-completion surveys was statistically significant (Feldt et al., 1987).

5. Results

RQ1: Response rates

Table 4 presents the response rates for the face-to-face and self-completion surveys.

Table 4: Response rates for GB ESS Round 10 face-to-face and self-completion surveys

	Face-to-face	Self-completion
Gross sample	5720	8000
Complete and usable partial completes	1137 ^a	2908 ^b
Known ineligible cases	381	8
Non-response		
Unknown eligibility	763	5084
Noncontacts and unusable partial completes	3439	0
AAPOR RR4 (eligibility rate $e = 0.93$)	21.51%	38.08%

Notes:

a: There was no fixed rule on inclusion of usable cases but only two cases in the data set answered less than 75% of 'ask all' questions. Includes units where those aged 15-17 were the target respondent (N=12).

b: Usable partial completes include at least 75% of 'ask all' questions.

Using the AAPOR RR4, we applied the same ineligibility rate to both surveys based on the reported face-to-face rates, since the ineligibility in self-completion appeared to be much lower because it relied on calls being made to a helpline to report that the sampled address was a business or vacant. The self-completion survey achieved a significantly higher response rate (AAPOR RR4 = 38%) compared to the face-to-face survey (AAPOR RR4 = 22%), a very large effect size. This was contrary to our expectations and represents a response rate towards the top end of what is currently being achieved in self-completion surveys in GB.

RQ2: Sample representativeness

The results are shown in Table 5. Estimates from the face-to-face mode had slightly larger standard errors than those from the self-completion mode due to the smaller effective sample size (Table 1).

Sex

First, we compared the distribution of respondent sex. In the face-to-face survey this is coded by interviewers, usually without asking the respondent, whilst in the self-completion survey respondents were asked to record this themselves. Although both surveys produced slightly lower proportions of male respondents (45% for face-to-face and 46% for self-completion) compared to the population benchmark of 48%, these differences were not statistically significant and were in line with existing literature. We conclude that the change of survey design here made little difference, and men continue to be slightly underrepresented, in line with our hypothesis H2.1.

Age

The findings were consistent with existing literature and supported our hypothesis H2.2. Both surveys significantly under-represented young adults aged 18–24: 7% for both the face-to-face survey and the self-completion survey, compared to the population benchmark of 10%. Similarly, respondents aged 25–34 were under-represented in both modes (10% for face-to-face and 13% for self-completion) relative to the population benchmark of 17%. However, for this age group, the self-completion survey proportion was closer to the benchmark, which was unexpected based on the literature.

Although previous research suggests that self-completion surveys are more likely to under-represent older respondents than face-to-face surveys, both modes in this study significantly over-represented those aged 65+. The proportion was 33% in the face-to-face survey and 28% in the self-completion survey, compared to a population benchmark of 24%. Again, estimates from the self-completion survey were closer to the benchmark.

Overall, the self-completion survey performed slightly better than the face-to-face survey in reflecting the population distribution by age. It was particularly more successful at reducing the over-representation of older adults and slightly improving representation of younger respondents, a group traditionally harder to reach in general population surveys.

Education

In terms of those with less than lower secondary education, both the face-to-face survey at 15% and self-completion survey at 16% were slightly lower than the population estimates (18%), but these differences were not statistically significant. Relative to the population benchmark of 10%, the self-completion survey slightly over-represented individuals with lower secondary education at 12%, while the face-to-face survey was not significantly different to the population at 10%. Both surveys significantly over-represented respondents with upper tier upper secondary education: 16% in the face-to-face survey and 14% in the self-completion survey, compared to a population benchmark of 12%.

Overall, the face-to-face survey performed slightly better in reflecting the population in terms of education than the self-completion study. There was one estimate that significantly

deviated from the population estimate for the face-to-face survey compared to two estimates for the self-completion survey. Nonetheless, both surveys showed reasonable representation of the less-educated, an encouraging finding given that this group is frequently under-represented in general population surveys.

Region

Considering region, both the face-to-face and self-completion surveys broadly reflected the population benchmarks, except for London, and the self-completion also over-represented those in the South East. Respondents from London were significantly under-represented in both modes: 7% in the face-to-face survey but much less, i.e. 11%, in the self-completion survey, compared to the population benchmark of 14%. This is in line with previous evidence that residents of large cities are harder to reach through face-to-face methods than through self-completion surveys. In contrast, the self-completion survey slightly over-represented respondents from the South East (16% compared to the population benchmark of 14%) whereas the face-to-face survey more closely matched the benchmark at 14%.

Table 5: Demographic characteristics

Variables	Face-to-face					Self-completion					Population estimates
	n	Prop (%)				n	Prop (%)				
		Est.	Std. error	95% CI			Est.	Std. error	95% CI		
				LL	UL				LL	UL	
Sex											
Male	494	45	0.017	41.61	48.24	1298	46	0.010	43.70	47.67	48
Female	631	55	0.017	51.76	58.39	1581	54	0.010	52.33	56.30	52
Age											
18–24	52	7*	0.010	4.98	9.06	142	7*	0.006	5.89	8.44	10
25–34	119	10*	0.010	8.55	12.57	350	13*	0.007	11.51	14.22	17
35–49	222	21	0.015	18.46	24.19	610	22	0.008	20.20	23.46	24
50–64	304	29	0.016	25.95	32.35	809	30*	0.010	28.42	32.19	25
65+	421	33*	0.015	29.73	35.69	898	28*	0.009	26.39	29.82	24
Education ^a											
ISCED 1: Less than lower secondary	176	15	0.011	13.00	17.53	467	16	0.007	14.63	17.53	18
ISCED 2: Lower secondary	98	10	0.011	8.09	12.36	333	12*	0.007	11.17	13.87	10
ISCED 3: Upper secondary (lower tier)	65	7	0.009	5.38	8.85	222	8	0.005	6.83	8.98	8
ISCED 3: Upper secondary (upper tier)	153	16*	0.014	13.75	19.39	353	14*	0.008	12.93	16.00	12
ISCED 4 / 5: Advanced vocational, sub-degree	148	15	0.012	13.05	17.85	431	17	0.008	15.24	18.36	17
ISCED 5 / 6: Bachelor's degree or equivalent, master, or doctoral degree	350	36	0.017	32.99	39.67	832	33	0.010	30.64	34.48	35
Region											
North East	59	5	0.007	3.79	6.43	107	3	0.001	3.16	3.67	4
North West	124	10	0.010	8.36	12.37	321	11	0.002	10.39	11.34	11
Yorkshire and the Humber	123	10	0.007	8.89	11.77	259	9	0.002	8.14	8.95	8
East Midlands	85	8	0.009	6.66	10.23	222	8	0.002	7.29	8.07	8
West Midlands	83	8	0.011	5.58	10.06	258	9	0.003	8.42	9.46	9
East of England	118	11	0.011	8.69	12.86	313	11	0.002	10.87	11.83	10
London	80	7*	0.008	5.71	8.83	312	11*	0.003	10.75	11.97	14
South East	150	14	0.012	11.45	16.25	448	16*	0.003	15.35	16.45	14
South West	123	12	0.011	9.67	13.86	299	10	0.002	9.66	10.57	9
Wales	63	6	0.010	4.65	8.77	150	5	0.002	4.65	5.29	5
Scotland	117	9	0.009	7.76	11.42	219	7	0.002	6.62	7.32	8

Notes: Unweighted *n* and weighted estimates accounting for the sampling design with design weights applied. Population estimates based on GB mid-year estimates for June 2021 (ONS, 2024). ^(a): The sample is based on respondents from England and Wales, and the population estimates are from the 2021 Census for England and Wales for people aged 18+(ONS, 2022). (*): denotes statistically significant differences relative to the population estimates.

Data quality

Table 6 shows the comparison of the data quality indicators between the two unweighted samples.

Table 6: Comparison of data quality indicators

Indicator	Face-to-face					Self-completion					p-value
	n	Mean	Std. err.	95% CI		n	Mean	Std. err.	95% CI		
				LL	UL				LL	UL	
Proportion of item non-response											
Out of 138 items	1125	0.011	0.001	0.008	0.013	2908	0.023	0.001	0.021	0.026	<0.001
Education	1125	0.016	0.004	0.009	0.026	2908	0.014	0.002	0.011	0.019	1.000
Mean straightlining											
Trust in institutions scale	1027	1.879	0.028	1.823	1.935	2820	2.016	0.018	1.980	2.051	<0.001
Democracy scale	1005	1.911	0.033	1.844	1.978	2663	1.939	0.021	1.898	1.980	0.419
Occupation codes	1100	0.012	0.003	0.007	0.019	2853	0.180	0.008	0.165	0.196	<0.001

Note: Mean values, standard errors, and confidence intervals for each survey are weighted using design but not post stratification weights. The differences in the unweighted sample sizes are due to missing or inapplicable values on each item.

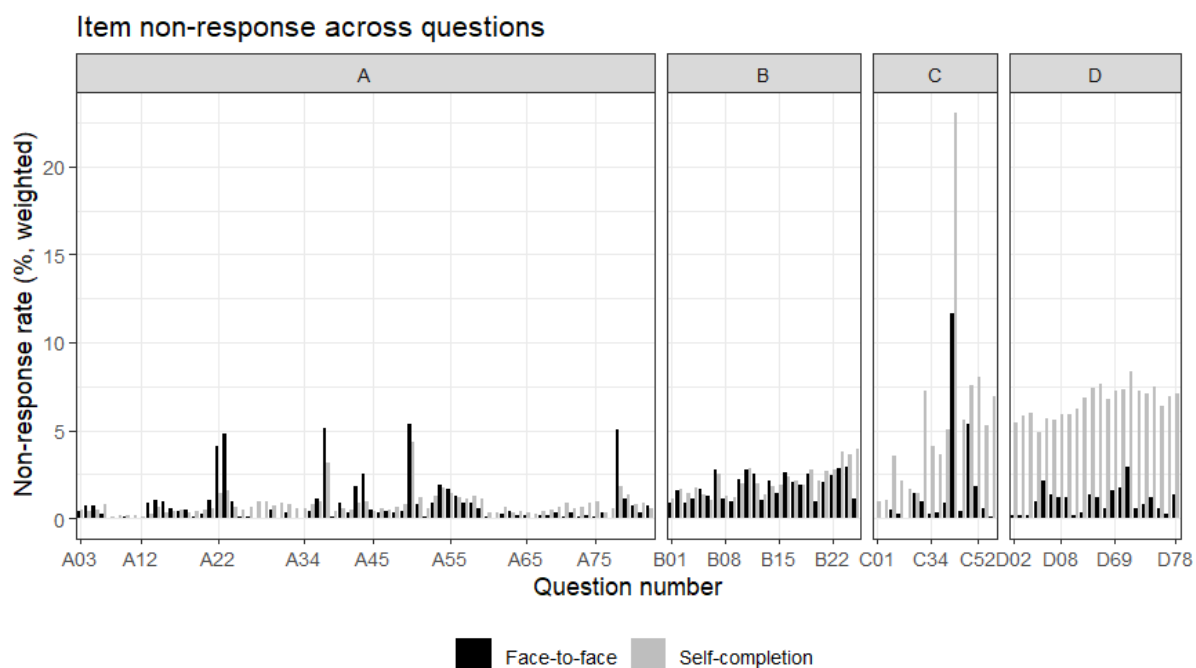
Item non-response

From Table 6, and considering item non-response, the mean was significantly higher in the self-completion survey (2.3%) than in the face-to-face survey (1.1%) ($t = -7.027$, $p < 0.001$), although the effect size was small. This is consistent with previous findings in the literature and our hypothesis H3.1 and can be partially explained by the role of interviewers in encouraging item response in face-to-face data collection. Higher item non-response might be a sign of less engagement, but it might also reflect that there is less pressure to answer when there is no eagerly awaiting interviewer.

Figure 1 compares item non-response rates by questions for both surveys. The main differences are observed in the last group of questions (later modules), mostly belonging to sections C (demographics) and D (rotating module on digital social contacts) of the questionnaire in self-completion survey, whereas in face-to-face survey the item non-response is more evenly spread across the whole questionnaire, with some exceptions. The main spike in item non-response is observed in the question about income in section C, which has a 12.2% non-response rate in the face-to-face survey, and as high as 22.4% in the self-completion survey. This result is expected, as non-response rates for income questions are frequently high due to both task complexity and question sensitivity (Jabkowski & Piekut, 2023).

We found no difference in the rates of item non-response for the *education* question, which is reassuring ($\chi^2 = 0.008$, $p = 1.0$). This suggests that the absence of an interviewer does not negatively affect respondents' ability to report their education level despite the long list they are required to select their answer from.

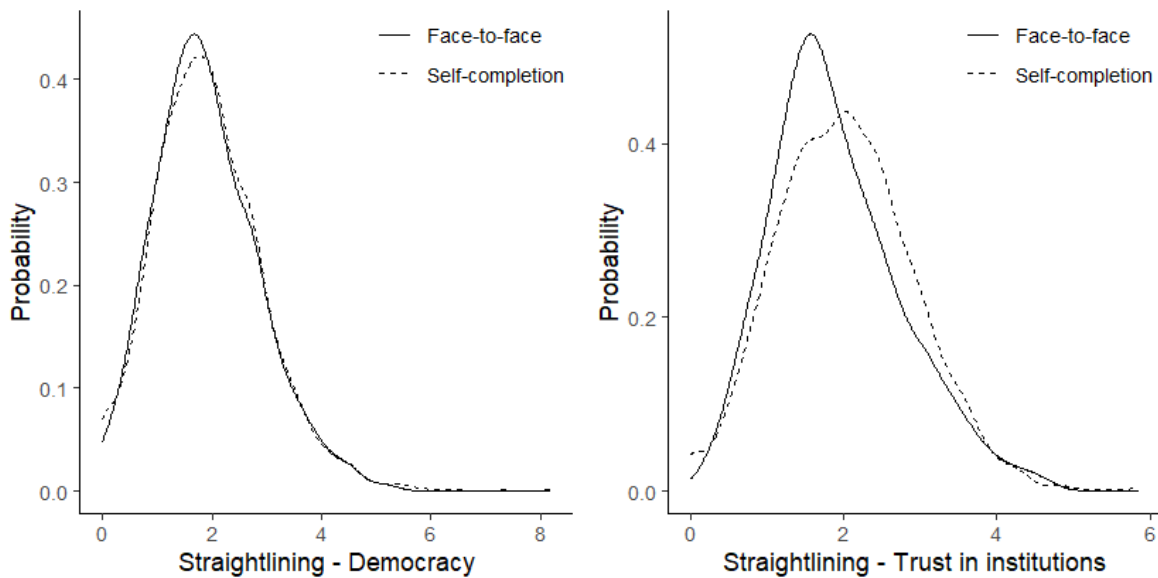
Figure 1: Item non-response across questions common to the face to face and self-completion survey



Straightlining

The results of the analysis of *straightlining*, reported in Table 6 and presented in Figure 2, indicate that there are no significant differences across surveys for the democracy scale (the *t*-test for independent samples with equal variances confirms that there is no difference in the mean value of this indicator across the unweighted samples ($t = -0.808$, $p = 0.419$)). However, there is a significant difference in *straightlining* for the trust in institutions scale, with respondents in the self-completion survey exhibiting a higher score, suggesting a lower tendency for *straightlining*; the *t*-test for independent samples assuming unequal variances confirms that there is a higher tendency of *straightlining* in this scale in the face-to-face survey compared to the self-completion survey ($t = 3.802$, $p < 0.001$). However, the difference is very small as shown in Figure 2. These results for the second block of attitudinal questions are consistent with the findings from the AAPOR report (AAPOR Executive Council Task Force et al., 2010) which, based on experimental results, suggest that individuals have a higher probability of engaging in *straightlining* when being interviewed face-to-face.

Figure 2: Distribution of straightlining scores for two scales in both surveys



Occupational coding

We also examined how productive the *occupation* coding was across the two surveys. The results (Table 6) indicate a significantly higher proportion of uncoded occupations in the self-completion survey compared to the face-to-face survey. The proportion of uncoded occupations in the self-completion survey (unweighted mean of 18.0%) is significantly higher than in the face-to-face survey (1.2%) ($\chi^2 = 213.7, p < 0.001$). This difference might be partially explained by the fact that different organisations conducted the occupation coding for the two surveys but most likely reflect differences in the information available to coders.

Internal consistency

The *internal consistency* of the three-item scale was assessed using Cronbach's α . For the self-completion survey, the scale reliability coefficient was 0.90, while for the face-to-face survey, this was 0.87. A Feldt's test of equality indicated that this difference was statistically significant ($F = 1.301, p < 0.001$). While both values exceeded the threshold of good internal consistency (Baldwin, 2019), results suggest that respondents in the self-completion survey provided more consistent responses across the items on attitudes towards gays and lesbians compared to those in the face-to-face survey. This might be partly due to reduced social desirability bias in self-completion surveys, leading to more uniform responses.

6. Conclusions and Discussion

A direct comparison of face-to-face data collection with a self-completion combination of web and paper on a general social survey in GB suggests that, for this type of study, face-to-face may no longer be the gold standard. The self-completion survey achieved a much higher response rate and better than expected sample composition, which was generally as good as for the face-to-face survey, whilst being much faster and considerably cheaper. The face-to-face survey had slightly less item non-response and better captured occupational and income data. Although both surveys showed high internal consistency, the self-completion survey showed higher consistency than the face-to-face, while there was no difference between the surveys on the measurement of education. Overall, we argue that this self-completion survey in GB outperformed the face-to-face survey. The priorities for the future would be to improve the measurement of occupation and income in self-completion surveys and look for ways to improve data quality.

When comparing the costs for the Round 10 face-to-face and self-completion approaches, for the self-completion survey, the actual cost per completed questionnaire was around one sixth than the cost of a face-to-face interview. Notwithstanding that different providers were used for the two surveys; self-completion offered a substantial cost saving compared to face-to-face which should be taken into consideration when discussing the advantages or disadvantages of using one mode of data collection over the other.

The findings in this paper provide survey commissioners and methodologists planning data collection with reassurance that a self-completion survey can offer a high-quality alternative to a face-to-face survey in contexts similar to that of this study. We believe this holds as long as the questionnaire can be rendered on paper as well as online, and if certain guidelines are followed precisely, such as unconditional and conditional incentives, communication and timing of letters (Dillman et al., 2014). However, more experiments would be needed to isolate the impact of each design feature. If these results are replicated more widely it would suggest that for general social surveys self-completion methods may offer a new gold standard.

Our study is not without limitations. The results are based on one survey and may not be generalisable to other studies. The comparison focused on self-completion surveys combining web and paper in a sequential design and not web only or paper only self-completion surveys. In this article we do not look at differences between the web and paper or in-person and video interviews. For brevity, we focused our analyses on certain sections and items of the questionnaire used. We did not (or could not) explore all quality criteria (such as, within household selection and other characteristics of the achieved samples). The fieldwork took place towards the end of the COVID-19 pandemic which may have influenced response preferences.

Future research should look at how distributions in the survey data differ between modes and the implications for the quality of both data collection techniques. The ESS is undertaking various other projects to gain more insight into self-completion data collection and how it compares to the interviewer-administered methods. For example, as part of the Survey Futures project, an ESS self-completion study has been launched, which to date shows

promising results, with response rates similar to those presented here (Reece & Lynn, 2025). If these results continue to hold true, we may be able to conclude that the pandemic was not the underlying cause of our findings in the present paper, but that a new gold standard might be emerging in the field of social surveys of the general population.

Acknowledgements

The production of this paper was supported by UKRI-ESRC strategic research grant ES/X014150/1 for “Survey data collection methods collaboration: securing the future of social surveys”, known as *Survey Futures*. *Survey Futures* is directed by Professor Peter Lynn, University of Essex, and is a collaboration of twelve organisations, benefitting from additional support from the Office for National Statistics and the ESRC National Centre for Research Methods. Further information can be found at www.surveyfutures.net. The research reported here forms part of Research Strand 4 of *Survey Futures*, led by Professor Olga Maslovskaya, University of Southampton, which focuses on methods for surveys without field interviewers.

The overall design and development of the European Social Survey upon which the two surveys compared in this paper are based, is funded by the Members of the ESS ERIC. ESS Round 10 fieldwork in the UK was funded by the ESRC as part of their obligations as a member of ESS ERIC. The ESS parallel run self-completion survey was funded by UKRI-ESRC ‘Comparing in-person to self-completion interviews: the example of the 2021 European Social Survey in the UK, Professor Rory Fitzgerald: City, University of London ES/W011824/1.

The authors would like to direct thanks Nick Moon and Ivor Knox who helped to design and implement the self-completion study conducted in Great Britain.

7. References

- AAPOR. (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. American Association for Public Opinion Research. <https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf>
- AAPOR Executive Council Task Force, Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D. A., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J. A., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K., & Zahs, D. (2010). Research Synthesis: AAPOR Report on Online Panels. *Public Opinion Quarterly*, 74(4), 711–781.
- Babbie, E. (2016). *The Practice of Social Research* (14th Edition). Cengage Learning.
- Baldwin, S. A. (2019). *Psychological Statistics and Psychometrics Using Stata*. Stata Press.
- Beullens, K., & Loosveldt, G. (2016). Interviewer Effects in the European Social Survey. *Survey Research Methods*, 10(2), 103–118.
- 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*, 1–12.
- Brown, M., & Hancock, M. (2015). *National Child Development Survey 2013 Follow-up: A guide to the datasets*. UCL Institute of Education. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/ncds_2013_follow_up_guide_to_the_datasets.pdf
- Cabinet Office. (2016). *Consultation response: Community Life Survey: Development and implementation of online survey methodology for future survey years* [Report]. Cabinet Office. https://assets.publishing.service.gov.uk/media/5a802969e5274a2e87db836d/community_life_survey_consultation_response_final.pdf
- Čehovin, G., Bošnjak, M., & Lozar Manfreda, K. (2023). Item Nonresponse in Web Versus Other Survey Modes: A Systematic Review and Meta-Analysis. *Social Science Computer Review*, 41(3), 926–945.
- Charman, C., Mesplaine-Cowan, S., & Collins, D. (2025). *The Post-pandemic Role of Face-to-face Fieldworkers: Second report*. <https://surveyfutures.net/wp-content/uploads/2025/10/report-6-role-f2f-interviewers-post-pandemic-second-report.pdf>
- Conrad, F. G., Couper, M. P., & Sakshaug, J. W. (2016). Classifying open-ended reports: Factors affecting the reliability of occupation codes. *Journal of Official Statistics*, 32(1), 75–92.
- de Leeuw, E. D. (1992). *Data Quality in Mail, Telephone and Face-to-Face Surveys*. TT-Publications.

- de Leeuw, E. D. (2008). Choosing the Method of Data Collection. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 113–135). Psychology Press.
- de Leeuw, E. D. (2018). Mixed-Mode: Past, Present, and Future. *Survey Research Methods*, 12(2), 75–89.
- Díaz de Rada, V. (2021). Is the Kish Household Sampling Method Better than the Birthday Method? *Italian Sociological Review*, 11(2), 485–508.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th Edition). John Wiley & Sons, Inc.
- Dixon, J., & Tucker, C. (2010). Survey Nonresponse. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (2nd Edition). Emerald Group Publishing Ltd.
- Domarchi, C., Ndebele, N., Maslovskaya, O., Fitzgerald, R., Lynn, P., & Comanaru, R. (forthcoming). *Recruitment methods for surveys without field interviewers in the UK: Evidence review* [Survey Futures Report].
- Durrant, G. B., & Steele, F. (2009). Multilevel modelling of refusal and non-contact in household surveys: Evidence from six UK Government surveys. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 172(2), 361–381.
- European Social Survey European Research Infrastructure Consortium (ESS ERIC). (2023). *ESS10—Integrated File, Edition 3.2* [Dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10e03_2
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement*, 11(1), 93–103.
- Finch, H. W., & French, B. F. (2015). *Latent Variable Modelling with R*. Routledge.
- Fitzgerald, R. (2024). *A cross-national survey transformation: The move to self-completion interviewing on Europe's flagship cross-national general social survey* [Keynote]. 5th European Social Survey International Conference, 08 July - 10 July 2024, Lisbon, Portugal.
- Fitzgerald, R., & Sibley, E. (2021). *Responding to the Pandemic: A 3 Country Self-Completion Push-to-Web Experiment in the European Social Survey*. Webinar: Survey Methodology Seminar Series, 15 April 2021, London: ESS City, University of London and NatCen Social Research.
- Gaziano, C. (2005). Comparative Analysis of Within-household Respondent Selection Techniques. *Public Opinion Quarterly*, 69(1), 124–157.
- Goyder, J. (1987). *The Silent Minority: Nonrespondents on Sample Surveys*. Westview Press.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd Edition). Wiley.

Haan, M., Ongena, Y., & Aarts, K. (2014). Reaching Hard-to-Survey Populations: Mode Choice and Mode Preference. *Journal of Official Statistics*, 30(2), 355–379.

Hanson, T. (2023). *A self-completion study based on the 2021 European Social Survey in Great Britain* [Methodological Report]. European Social Survey ERIC, City, University of London.

Hanson, T., Fitzgerald, R., & Comanaru, R. (2024). *Self-Completion Study Based on the 2021 European Social Survey in Great Britain* [Data Collection]. UK Data Service.
<http://doi.org/10.5255/UKDA-SN-9038-1>

Heerwegh, D., & Loosveldt, G. (2006). Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly*, 72(5), 836–846.

Ipsos MORI. (2018a). *Labour Market Survey Response Rate Experiments: Report for test 1: Materials experiment* [Report]. Ipsos MORI Social Research Institute.
https://gss.civilservice.gov.uk/wp-content/uploads/2018/04/Test-1_Full-report_FINAL-for-publishing.pdf

Ipsos MORI. (2018b). *Labour Market Survey Response Rate Experiments: Report for test 2, tranche 1: Incentives experiment* [Report]. Ipsos MORI Social Research Institute.
<https://gss.civilservice.gov.uk/wp-content/uploads/2018/04/Test-2-Tranche-1-report-FINAL-for-publishing.pdf>

Jabkowski, P., & Piekut, A. (2023). Not random and not ignorable. An examination of nonresponse to income question in the European Social Survey, 2008-2018. *Field Methods*, 36(3), 213–228.

Jäckle, A., Lynn, P., & Burton, J. (2015). Going online with a face-to-face household panel: Effects of a mixed mode design on item and unit non-response. *Survey Research Methods*, 9(1), 57–70.

Kocar, S., Brown, M., & Calderwood, L. (2023). *Report 3: Occupation coding in self-completion surveys: Evidence Review* (Survey Futures Report 3).
<https://surveyfutures.net/wp-content/uploads/2025/02/report-3-occupation-coding-self-completion-surveys-evidence-review.pdf>

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), 213–236.

Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*, 52(4), 526–538.

Loosveldt, G. (2008). Face-to-Face Interviews. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 201–220). Psychology Press.

Loosveldt, G., Wuyts, C., & Beullens, K. (2018). Interviewer Variance and Its Effects on Estimates. *Quality Assurance in Education*, 26(2), 227–242. <https://doi.org/10.1108/QAE-06-2017-0030>

Lutig, P. (2024). *ESS round 10 mode experiments in Great Britain and Finland. Findings on mode effects* [Working Paper]. European Social Survey ERIC, City, University of London. <https://europeansocialsurvey.org/sites/default/files/2024-10/round-10-experimental-comparison-final.pdf>

Lynn, P. (2020). Evaluating Push-to-Web Methodology for Mixed-Mode Surveys Using Address-Based Samples. *Survey Research Methods*, 14(1), 19–30.

Maslovskaya, O., Chang, G., Howe, S., Aizpurua, E., & Perelli-Harris, B. (2024). *Effectiveness of Using Quick Response (QR) Codes in Survey Invitation Letters: Experiment conducted in Generations and Gender Survey in the UK* [Working Paper]. University of Southampton.

Maslovskaya, O., Lynn, P., Calderwood, L., Durrant, G. B., Fitzgerald, R., Nicolaas, G., & Williams, J. (2025). *Survey Futures Position Statement on Response Rates: Supporting Material*. https://surveyfutures.net/wp-content/uploads/2025/06/Response-Rates-Postition-Statement_SupportingMaterial.pdf

NatCen Social Research. (2020). *British Social Attitudes 37 Technical Report* [Report]. National Centre for Social Research. https://natcen.ac.uk/sites/default/files/2023-08/bsa37_technical-details_0.pdf

NatCen Social Research. (2022). *British Social Attitudes 2022 user guide* [Report]. National Centre for Social Research. https://doc.ukdataservice.ac.uk/doc/9283/mrdoc/pdf/9283_bsa_2022_user_guide.pdf

NatCen Social Research. (2024). *British Social Attitudes 41 Technical Report* [Report]. National Centre for Social Research. <https://natcen.ac.uk/sites/default/files/2024-06/BSA%2041%20Technical%20Details.pdf>

Neuman, W. L. (2012). Designing the Face-to-Face Survey. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (1st Edition, pp. 227–248). Springer.

Olson, K., & Smyth, J. D. (2017). Within-Household Selection in Mail Surveys: Explicit Questions are Better than Cover Letter Instructions. *Public Opinion Quarterly*, 81(3), 688–713.

Olson, K., Smyth, J. D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N. A., McCarthy, J. S., O'Brien, E., Opsomer, J. D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z. T., Turakhia, C., & Wagner, J. (2019). *Report of the AAPOR Task Force on Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys* (pp. 1–327) [Task Force Report]. AAPOR. <https://aapor.org/wp-content/uploads/2022/11/Report-of-the-Task-Force-on-Transitions-from-Telephone-Surveys-FULL-REPORT-FINAL.pdf>

Olson, K., Stange, M., & Smyth, J. (2014). Assessing Within-Household Selection Methods in Household Mail Surveys. *Public Opinion Quarterly*, 78(3), 656–678.

- ONS. (2017). *Labour force survey performance and quality monitoring report, July to September 2017* [Report]. Office for National Statistics.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports/labourforcesurveyperformanceandqualitymonitoringreportjulytoseptember2017>
- ONS. (2019). *Labour force survey performance and quality monitoring report, January to March 2019* [Report]. Office for National Statistics.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports/labourforcesurveyperformanceandqualitymonitoringreportjanuarytomarch2019>
- ONS. (2020a). *How face-to-face interviewer attitudes and beliefs moderate the effect of monetary incentive on UK Labour Force Survey response rates*. Office for National Statistics.
- ONS. (2020b). *Labour Force Survey performance and quality monitoring report: January to March 2020*. Office for National Statistics.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports/labourforcesurveyperformanceandqualitymonitoringreportjanuarytomarch2020>
- ONS. (2020c). *Labour Market Survey: Technical report* [Technical Report]. Office for National Statistics.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourmarketsurveytechnicalreport#the-labour-market-survey-statistical-test>
- ONS. (2022). *Census 2021, England and Wales* [Dataset]. Office for National Statistics.
<https://www.ons.gov.uk/census>
- ONS. (2024). *Estimates of the population for the UK, England, Wales, Scotland, and Northern Ireland* [Dataset]. Office for National Statistics.
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>
- ONS. (2025). *Labour Market Transformation – update on progress and plans: April 2025* [Technical Report]. Office for National Statistics.
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/labourmarkettransformationupdateonprogressandplans/april2025>
- Ormston, R., Martin, C., Rogers, L., Huskinson, T., Irvin, E., Rimmington, E., & Lynn, P. (2024). *Mixed mode research: Report to inform the Scottish Government Long Term Survey Strategy* [Public Services and Government]. Scottish Government.
- Peycheva, D. N., Sakshaug, J. W., & Calderwood, L. (2021). Occupation Coding during the Interview in a Web-First Sequential Mixed-Mode Survey. *Journal of Official Statistics*, 37(4), 981–1007. <https://doi.org/10.2478/jos-2021-0042>

- Reece, N., & Lynn, P. (2025). An experimental comparison of within-household selection methods for self-completion surveys. *11th Conference*. European Survey Research Association, Utrecht, The Netherlands.
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and Modelling Psychological Measures in the Presence of Multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140.
- Reuning, K., & Plutzer, E. (2020). Valid vs. Invalid straightlining: The complex relationship between straightlining and data quality. *Survey Research Methods*, 14(5), 439–459.
- Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and Interviewing. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research* (2nd Edition, pp. 437–470). Emerald Group Publishing Ltd.
- Smith, P. (2020). The Impact of Covid-19 on High Quality Complex General Population Surveys [Social Research Association]. *The SRA Blog*. <https://the-sra.org.uk/SRA/Blog/The%20impact%20of%20Covid19%20on%20high%20quality%20complex%20general%20population%20surveys.aspx>
- Smyth, J. D., Olson, K., & Stange, M. (2019). Within-Household Selection Methods: A Critical Review and Experimental Examination. In P. J. Lavrakas, M. W. Traugott, C. Kennedy, A. L. Holbrook, E. D. de Leeuw, & B. T. West (Eds.), *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment* (1st Edition, pp. 23–45). John Wiley & Sons, Inc.
- Stange, M., Smyth, J. D., & Olson, K. (2016). Using a Calendar and Explanatory Instructions to Aid Within-household Selection in Mail Surveys. *Field Methods*, 28(1), 64–78. <https://doi.org/10.1177/1525822X15604825>
- Stoop, I. A. L. (2005). *The Hunt for the Last Respondent: Nonresponse in Sample Surveys* (SCP Report 2005/8). Social and Cultural Planning Office.
- TNS BMRB. (2013). *Community Life Survey: Summary of web experiment findings*. Cabinet Office. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/325872/Annex_B_-_Summary_of_web_experiment_findings_2012-13.pdf
- TNS BMRB. (2014a). *Community Life Survey Technical Report 2013-14* [Technical Report]. Cabinet Office. <https://www.gov.uk/government/publications/community-life-survey-2013-to-2014-technical-report>
- TNS BMRB. (2014b). *Community Life Web Survey Technical Report 2013-14* [Technical Report]. Cabinet Office. https://doc.ukdataservice.ac.uk/doc/7737/mrdoc/pdf/7737_community_life_2013-14_web_survey_technical_report.pdf
- Tucker, C., & Lepkowski, J. M. (2007). Telephone Survey Methods: Adapting to Change. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. de Leeuw, L. Japiec, P. J. Lavrakas, M. W. Link, & R. L.

Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 1–26). John Wiley & Sons, Inc.

Villar, A., & Fitzgerald, R. (2017). Using mixed modes in survey research: Evidence from six experiments in the ESS. In M. J. Breen (Ed.), *Values and Identities in Europe: Evidence from the European Social Survey* (pp. 259–293). Routledge.

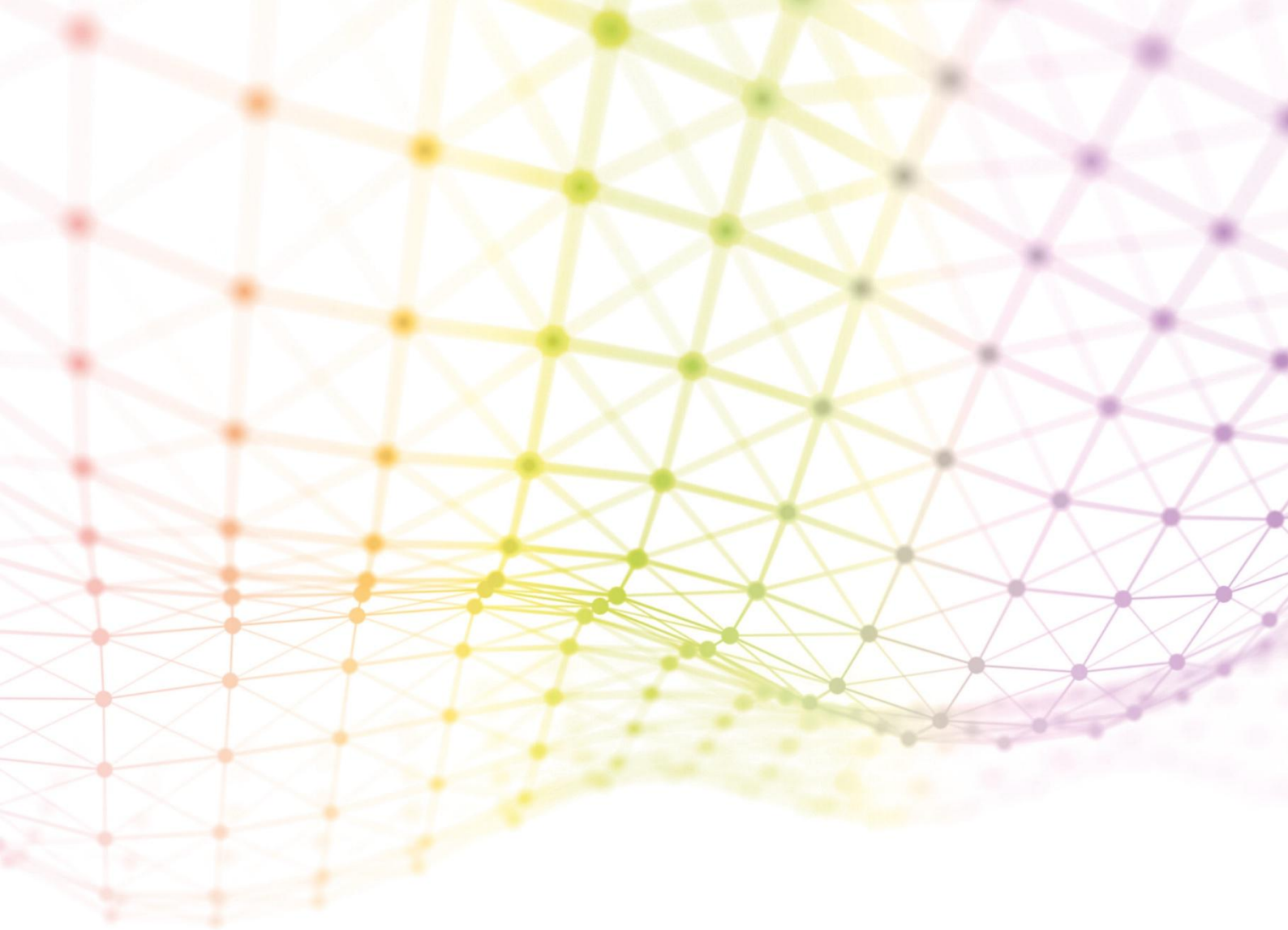
West, B. T., & Blom, A. G. (2017). Explaining Interviewer Effects: A Research Synthesis. *Journal of Survey Statistics and Methodology*, 5(2), 175–211.

Williams, D., & Brick, J. M. (2018). Trends in U.S. Face-to-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology*, 6(2), 186–211.

Williams, J. (2015). *Community Life Survey: Investigating the Feasibility of Sampling All Adults in the Household* [TNS BMRB Report]. Cabinet Office.
https://assets.publishing.service.gov.uk/media/5a806a11ed915d74e622e50e/The_Community_Life_Survey_Investigating_the_Feasibility_of_Sampling_All_Adults_in_the_Household_FINAL.pdf

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.

Zhang, C., & Conrad, F. (2014). Speeding in Web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127–135.



University of Essex



University of
Southampton



Economic
and Social
Research Council

NCRM NATIONAL CENTRE FOR
RESEARCH METHODS

Office for
National Statistics

UCL

WARWICK
THE UNIVERSITY OF WARWICK

MANCHESTER
The University of Manchester

CITY
UNIVERSITY OF LONDON

**National Centre
for Social Research**

LSE LONDON SCHOOL
OF ECONOMICS AND
POLITICAL SCIENCE

Uster University

Unil
UNIL | Université de Lausanne

Ipsos

verian
Formerly Kantar Public

www.surveyfutures.net