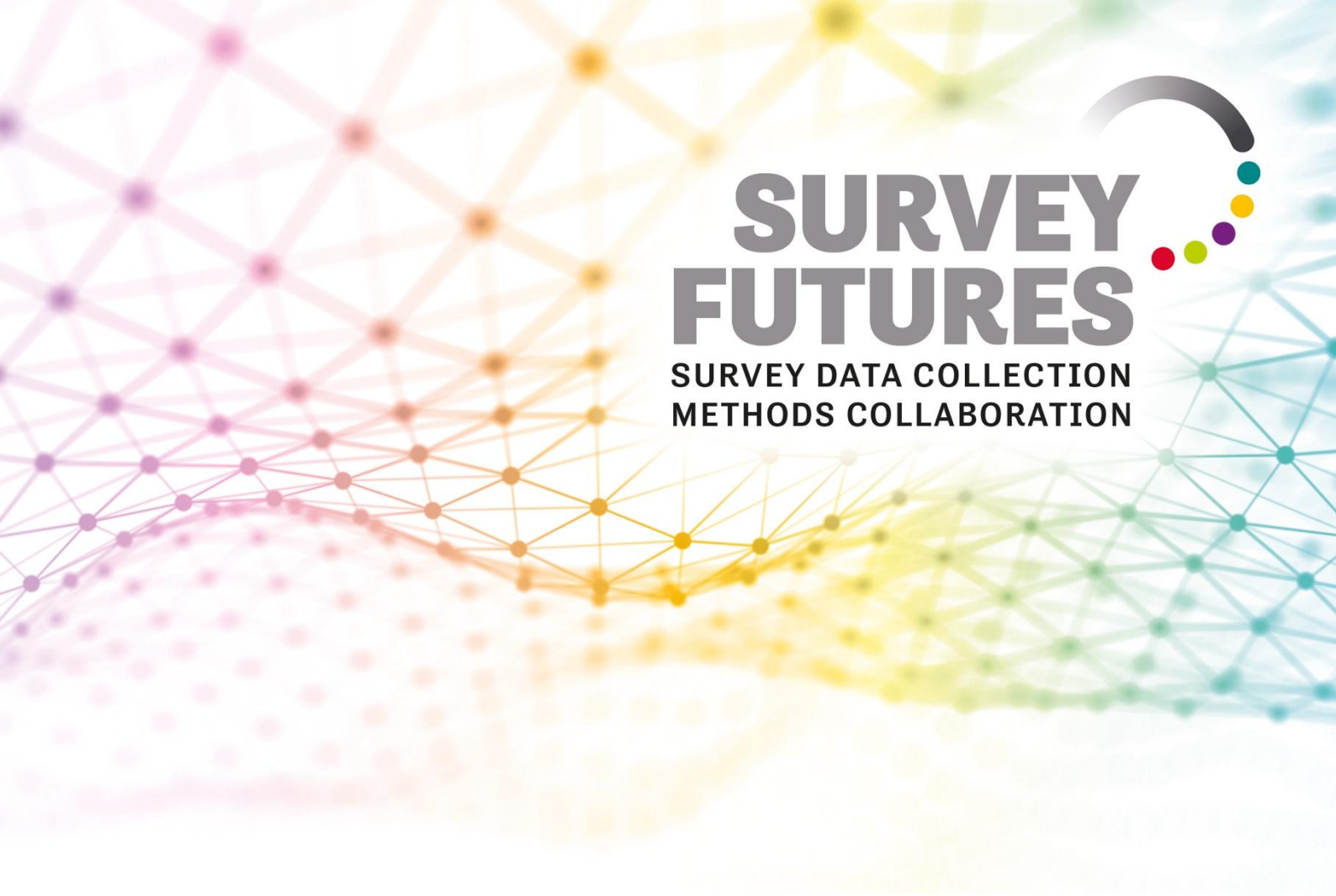


Survey Practice Guide 8: When to anticipate mode effects in mixed-mode surveys

Georgia D. Tomova, Richard J. Silverwood, Liam Wright

Centre for Longitudinal Studies, University College London

May 2026

www.surveyfutures.net

Survey Futures is an Economic and Social Research Council (ESRC)-funded initiative (grant ES/X014150/1) aimed at bringing about a step change in survey research to ensure that high quality social survey research can continue in the UK. The initiative brings together social survey researchers, methodologists, commissioners and other stakeholders from across academia, government, private and not-for-profit sectors. Activities include an extensive programme of research, a training and capacity-building (TCB) stream, and dissemination and promotion of good practice. The research programme aims to assess the quality implications of the most important design choices relevant to future UK surveys, with a focus on inclusivity and representativeness, while the TCB stream aims to provide understanding of capacity and skills needs in the survey sector (both interviewers and research professionals), to identify promising ways to improve both, and to take steps towards making those improvements. 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.

Research Strand 6 of Survey Futures ("Mode effects") has three related parts, on reduction of mode effects by design (led by Jo d'Ardenne, formerly of NatCen), on evaluation of mode effects (led by Professor Annette Jäckle, University of Essex), and on handling mode effects in analysis (led by Liam Wright, UCL). This Survey Practice Guide stems from the third of these parts.

The suggested form of citation for this guide is:
Tomova GD, Silverwood RJ & Wright L. (2026) 'When to anticipate mode effects in mixed-mode surveys', *Survey Futures Survey Practice Guide* no. 8 Colchester, UK: University of Essex. Available at https://surveyfutures.net/practice-guides/.

1. Introduction

Many surveys have been transitioning to mixed-mode designs where data are collected using more than one mode of data collection, e.g. by employing both a web survey and a face-to-face interview. This shift has been driven by several interconnected pressures, but concerns have been raised that the use of mixed-mode designs can lead to differences in the responses provided by each mode, so called 'mode effects' (de Leeuw et al., 2008) or 'mode measurement effects' (Klausch et al., 2013), which may introduce bias in the analyses of mixed-mode data. It is important to distinguish these from differences that reflect that different types of people who may respond via each mode (known as 'mode selection' or 'selection effects'). In short, mode (measurement) effects refer to *how* people respond, whereas mode selection effects refer to *who* responds by each mode (Burton & Jäckle, 2020). **Table 1** provides short definitions of each key concept. Throughout this guidance, we focus exclusively on mode measurement effects.

In some cases, mode effects can be prevented or reduced during the survey design stage, for example by presenting questions as consistently as possible across modes, or by using self-administered questions for sensitive items even within interviewer-administered surveys. Targeted guidance also exists for specific transitions, most notably for surveys moving to the web, such as the Measurement Effect Risk Framework (MERF) in *Survey Practice Guide 2* (d'Ardenne et al., 2025). That being said, the prevailing view is that mode effects cannot always be fully prevented, given there exist fundamental differences between modes that are likely to drive mode effects, such as whether an interviewer is physically present when it comes to sensitive questions (Tourangeau & Yan, 2007).

This practice guide sits within a wider set of complementary resources for both survey designers and analysts of mixed-mode survey data. It draws on the findings of a recent systematic review of mode effects (Tomova et al., 2026), providing a general overview of the main findings with the aim of helping users to identify scenarios (e.g. combinations of modes or mode characteristics) that increase the risk of mode effects. This guide is intended to support survey designers in making informed decisions about whether a certain mixed-mode combination warrants particular attention, when mitigation strategies to reduce mode effects may need consideration, or whether mode effects are simply unlikely to be of concern. For those interested in appropriately designing mixed-mode surveys in situations where mode effects are likely to occur, the complementary *Survey Practice Guide 2* (d'Ardenne et al., 2025) contains a framework which can help guide the process in such scenarios. Although specifically focused on the transition from interviewer-administered to online surveys, it offers practical guidance that can help in the development and testing of questionnaires more widely. Beyond the design stage, the present guide is also relevant to those analysing existing mixed-mode data. By understanding the scenarios in which mode effects are most likely to occur, and to what extent, it can help inform decisions on whether mode effects may need to be accounted for in statistical analyses. A separate Survey Practice Guide (Wright et al., 2026 [forthcoming]) is dedicated to approaches for handling mode effects in the

analyses of mixed-mode data, including detailed practical guidance and recommendations for analysts.

1.1. Definitions

Table 1. Definitions of key concepts.

Concept	Definition
Mode	The method of survey data collection, e.g. face-to-face interview, web survey
Mixed-mode survey	A survey in which more than one mode is used to collect the same data
Mode (measurement) effect	A systematic difference in the measurement of survey responses attributable to the mode of data collection, i.e. a difference in *how* people respond by each mode
Mode selection effect	A systematic difference in survey responses attributable to different respondent composition in each mode of data collection, i.e. a difference in *who* responds by each mode
Mode experiment	A study in which the mode of data collection is experimentally (i.e. randomly) assigned with the aim of obtaining estimates of the mode measurement effect that are free of mode selection effects

1.2. Evidence informing this practice guide

The occurrence and magnitude of mode effects can be most robustly studied experimentally, by conducting experiments in which respondents are randomly assigned to a particular mode. In contrast, attempting to do this in observational data is more likely to result in differences that are a combination of mode effects and selection effects.

The evidence and recommendations in this guide are drawn from our recent systematic review, which is the largest and most comprehensive conducted to date (Tomova et al., 2026). It examined 90 mode effect experiments conducted in general populations (or age-, sex-, or region-specific strata thereof). The experiments were conducted in more than 25 countries, with nearly half conducted in the USA, and six in the UK. They were published between 1967 and 2024 (though with the majority from 1998 onwards). Across all included studies, we reviewed mode effects on 3,545 items, resulting in 4,113 mode effect estimates in total (since some items were examined in more than one way), corresponding to 128 mode comparisons (since some of the 90 experiments compared more than one pair of modes).

For the purposes of the review, we applied a threshold of 0.2 SD to determine whether a mode effect is 'meaningful', which is a commonly used rule of thumb for a 'small' standardised effect (Sullivan & Feinn, 2012). While such thresholds are arbitrary, it is necessary to select one since items will vary considerably, making any synthesis on such a large scale extremely challenging. It is also important to note that whether a 0.2 SD difference is meaningful in practice will vary depending on the specific item. For example, consider life satisfaction measured by the ONS4 "Overall, how satisfied are you with your life nowadays?" item on a 0-10 scale (0 = "not at all", 10 = "completely") (ONS, 2025). Suppose the standard deviation is 2 (close to the observed 1.8 [Tenaglia, 2024]), then a difference between modes of 0.2 SD would represent a difference of 0.4 points on the scale. Although still small at the individual level, a difference of this size may still lead to systematic observable differences between modes, or result in apparent changes over time when switching modes. As another example, a 0.2 SD is approximately the difference in average height between 15- and 16-year-old girls (Cohen, 2009).

1.3. Scope and limitations

This guidance is focussed on general differences in the measurement of items between modes (e.g. whether there is evidence that responses differ between paper and web or paper and face-to-face surveys), but does not extend to item-level features such as differences in scales, ordering of response options, or specific wording of questions (however, many of these features are discussed in *Survey Practice Guide 2* [d'Ardenne et al., 2025]). The guidance therefore relates to general mode characteristics, such as whether an interviewer is present or whether the questions are delivered visually or aurally, rather than to the actual design of individual survey items. We focus on the difference in average responses to items between modes, rather than specific behaviours such as 'straightlining'. This reflects both the scope of the systematic review on which this guidance is based, as well as the nature of the available evidence (i.e. the actual questionnaire is not always available as part of articles on mode experiments, and alternative types of metrics are commonly not reported).

We are only able to provide recommendations for mode comparisons for which sufficient evidence exists. For example, no relevant experiments involving video interviewing had been conducted at the time of reviewing the literature (though some evidence has been published since, e.g. see *Survey Futures Working Paper 11* [Kocar et al., 2025]). We are also not able to discuss the direction of mode effects - which depends on the coding of the individual question and responses – and we therefore focus on the size of the mode difference. For some item categories, evidence may be available from only one or two studies, which may affect the extent to which findings will generalise to other settings. Experiments were conducted in a variety of countries around the world with different populations, and some of them were conducted a long time ago, both of which may further reduce comparability and relevance. When comparing mode characteristics, it is not always possible to separate them completely (e.g. questions delivered aurally are often also delivered in the presence of an interviewer), which may lead to misattribution of the findings. Finally, it is worth noting

that, even when mode is randomly assigned, mode experiments are not necessarily immune to mode selection, which may be introduced in such settings by differential non-response or non-compliance (to the assigned mode) between the groups. Nevertheless, many experiments have been designed and analysed thoughtfully. The database introduced in section 3 contains detailed information on all studies and can be used to gauge the quality of individual papers.

2. Mode effects: evidence and recommendations

In this section, we briefly summarise the evidence for each type of mode comparison or mode characteristic and provide a brief recommendation on whether it is important to consider strategies to mitigate against mode effects, or whether this is unlikely to be of substantial concern in the context of mixed-mode surveys.

It is important to keep in mind that these recommendations are general and reflect the evidence as a whole, rather than specific individual scenarios. For certain mode combinations, although mode effects may be unlikely in general, they may still occur for particular items. Where possible, we discuss the categories of items where mode effects are more or less likely to arise (see **Table 2** for which items were assigned to which categories). Ultimately, just as the decision to adopt a mixed-mode design for a particular survey will be informed by the overall risk of mode effects rather than the risk of mode effects in a few specific items, our recommendations are also intentionally broad, as we believe this is likely to be most generalisable and useful in practice.

We present the evidence and recommendations organised by comparing different mode characteristics (e.g. presence or absence of an interviewer) to make it easier to convey the key messages most relevant to survey design decisions (**Table 3**). We further present the results in an alternative way, organised by direct mode comparisons (e.g. paper vs web) (**Table A1**). Further details and a more comprehensive discussion of all results are available in the review itself (*Survey Futures Working Paper 12* [Tomova et al., 2026]).

Table 2. Item categories used when examining the underlying evidence informing this practice guide and the types of items that were assigned to each category.

Item category	Types of items
Adverse experiences	Victimisation, Discrimination, Sexual abuse, Harassment, Physical abuse, Adverse childhood experiences, Bereavement, Emotional abuse, Financial abuse
Attitudes & expectations	Government & politics, Attitudes towards others, Attitudes towards behaviours, Attitudes towards surveys, General attitudes, Attitudes towards environment, Attitudes towards risk, Self-esteem, Personality, Expectations, Religion
Behaviour	Substance use, Risk behaviour, Criminal history, Time use, Violence, Conflict, Truancy
Demographics	Respondent demographics

Education	Educational achievements
Employment, income & assets	Assets, General finances, Employment, Income, Financial planning
Health and wellbeing	Alcohol, Mental health, Smoking, General health, Health conditions, Cognition, Exercise, Diet, Weight, Subjective wellbeing, Oral health, Sleep, Disability, Healthcare usage, Life satisfaction
Leisure	Leisure, Technology use, Hobbies
Sexual activity, pregnancy & fertility	Sexual activity, Contraception, Sexually transmitted infections, Fertility, Pregnancy
Social	Marriage & cohabitation, Social media, Family, Friends, Other relationships, Social life

Table 3. Evidence and recommendations for mixed-mode design combinations, organised by type of mode characteristics.

MODE CHARACTERISTICS	OVERALL SUMMARY OF EVIDENCE	DIFFERENCES IN ITEM DOMAINS	CAVEATS	RECOMMENDATION
Interviewer presence (physical) e.g. Face-to-face vs Web; Face-to-face vs Telephone; Paper[a] administered in presence of interviewer vs in absence	Overall, the evidence was mixed and was highly dependent on the specific item domain.	The most sensitive items such as adverse experiences, sexual activity, and social variables, were more likely to lead to mode effects when modes differed based on whether the interviewer was physically present or not. For other domains, such as attitudes, behaviours, employment, and health, mode effects were less likely.	For some items there was more available evidence to make substantive conclusions than others. For example, there was evidence on health variables from 26 experiments, but only 6 for adverse experiences, sexual activity, and social items, which could also explain the differences in findings.	The physical presence of an interviewer is most likely to introduce mode effects for the most sensitive types of items. If interested in collecting such data, where possible, consider combining modes that do not differ based on interviewer presence. Where relevant, ensure technical documentation explains the risks of mode effects and refers to methods for handling mode effects in the analyses.
Written or aural question delivery e.g. Telephone vs Paper[a]; ACASI[b] vs CASI[c]; Face-to-face vs Paper[a]	Overall, mode effects were consistently likely to occur where modes differed based on written or aural presentation.	The only domain for which the evidence did not suggest substantial mode effects are likely to occur was behavioural items. For all other domains, the	The evidence for the domains exhibiting the largest and most common mode effects (sexual activity and social items) came from relatively fewer number of	When combining (or transitioning between) modes that differ in whether questions are presented visually as text or aurally, anticipate that mode effects are likely to

		evidence was generally consistent.	studies than other domains, which may mean the evidence is not as reliable.	occur, and ensure that any technical documentation explains the risks of mode effects and refers to methods for handling mode effects in the analyses.
Computer-assisted or computer-unassisted modes e.g. PAPI[d] vs CAPI[e]; Telephone vs T-ACASI[f]; Paper[a] vs Web	Overall, mode effects can occur, but are unlikely to be substantial, and their magnitude and occurrence differs across domains.	More sensitive items like those referring to social and sexual activity behaviours were more likely to be subject to mode effects than others such as behaviour and adverse experiences. Mode effects were also more likely for education items. Even within domains, the evidence was mixed and there was large variability across items.	Some studies that reported use of CAPI[e] may have additionally used CASI[c] for more sensitive items – where this was not appropriately reported in the studies, it may explain the larger mode effects observed for more sensitive topics. The evidence for mode effects in education may have been skewed by mode effects reported in large-scale educational assessments.	Combining modes with and without computer-assisted elements may give rise to mode effects, but the evidence is not consistent enough to provide a clear recommendation. Consider consulting the database of mode effects (Section 3) for specific mode contrasts and specific items of interest for more targeted and informed decision-making.
Answers reported directly to an interviewer or not	Overall, this mode characteristic led to the largest and most common mode effects observed in the literature.	The evidence was consistent across the domains. Mode effects were particularly large and common for attitudinal, sexual activity,	No significant concerns were identified.	Combining or transitioning between modes which differ in terms of whether answers are reported directly to an interviewer or are self-

e.g. CAPI[e] vs CASI[c]; Telephone vs Paper[a]; Face-to-face vs Web		and social items, but all domains had evidence for mode effects. The only domain in which mode effects were relatively less likely to occur were behavioural items.		administered poses the highest risk of mode effects out of all mode characteristics that were examined. Where relevant, consider following elements of the MERF framework(d'Ardenne et al., 2025). Where possible, consider administering questions most likely to be subject to mode effects in a self-administered way even in an interviewer-led mode. Ensure that the appropriate technical documentation explains the risks for mode effects and refers users to methods for handling potential mode effects during analysis.
Answers self-reported on paper or electronically e.g. Paper[a] vs Web	Overall, the mode effects are unlikely to occur based on this characteristic.	Overall, although mode effects were more likely to be observed in some domains (e.g. sexual activity and health), mode	There was evidence for large mode effects in only one domain (education), however the evidence exclusively came from	Where both modes involve self-administered questionnaires without interviewer involvement, the risk of mode effects is

		effects were, on the whole, consistently smaller and less common across all domains.	large-scale educational assessments conducted on paper or electronically, which may not be relevant to surveys, and therefore excluded from consideration here.	low. Therefore, transitioning from paper to web surveys, or employing mixed-mode paper and web designs can proceed without special considerations regarding mode effects.
Interviewer involvement type: only present or directly collecting responses e.g. CASI[c] vs CAPI[e]; ACASI[b] vs T-ACASI[f]; Paper[a] (in presence of interviewer) vs Face-to-face	Overall, mode effects were likely to occur and were substantial in some domains.	Some domains were particularly likely to give rise to mode effects in this setting, e.g. sexual activity, adverse experiences, employment and income. Other domains (e.g. attitudes, behaviours, health and wellbeing) were less likely in comparison, but mode effects were still common.	The evidence for some domains only came from single studies (e.g. attitudinal and employment items), which means they may not generalise to many settings. Some other types of items (e.g. adverse experiences, sexual activity) had evidence that only came from around 25 mode effect estimates in total for each, which is also quite small. However, this is all somewhat mitigated by the fact that, on the whole, the evidence that mode effects are larger and more likely to occur	When combining modes that differ based on whether the interviewer is simply present or directly collecting the responses, consider implementing design choices that might minimise mode effects, e.g. for sensitive items such as those related to sexual activity, consider collecting these in a self-administered way, even when the interviewer is otherwise present. When transitioning between such modes, or where mode effects cannot be prevented, ensure that the appropriate technical documentation explains

		based on this mode characteristic, was consistent across the available evidence.	the risks for mode effects and refers users to methods for handling potential mode effects during analysis.

[a]Self-completed paper questionnaire (distinct from PAPI, which is interviewer-administered) [b]Audio computer-assisted self-interviewing, [c]Computer-assisted self-interviewing, [d]Paper-and-pencil interviewing, [e]Computer-assisted personal interviewing, [f]Telephone audio computer-assisted self-interviewing

3. Mode effects database

3.1. Introduction

The recommendations provided in this practice guide focus on providing evidence synthesised across all findings. Although useful for making general statements about whether mode effects are likely to occur for a given mode combination, this may not be sufficient when there is interest in whether mode effects are likely to exist for specific items, for which the evidence may differ from the overall pattern. Another challenge is that findings may also vary across populations, which is an additional consideration when interpreting the guidance.

To make it easier to find more targeted and relevant evidence on mode effects, in addition to the general guidance, we also collated all mode effect estimates (n=4,113) we extracted from the literature into a freely accessible online database. This database can be used to locate evidence on mode effects for specific items, populations, or surveys in which experiments were conducted. This way, practitioners can locate evidence that is most relevant and most closely aligned with their own needs.

3.2. Overview of contents

The database is accessible at https://cls-data.github.io/mode-effects-database/.

It is organised in three sections: a homepage, a 'Mode effects' section, and a 'Details of each study' section. The homepage contains general information about the database, the underlying systematic review, and offers a set of general instructions on how to navigate the contents, including how to perform searches.

The 'Mode effects' section provides information on the study, item, mode comparison, and the mode effect, with a separate entry (row) for each mode effect estimate. General information is available to help contextualise the setting in which the specific mode effect was obtained, including year of publication, country of origin, source population category, and a brief description of the survey population. This is followed by the name of the survey item, and the category in which it belongs to (e.g. health, behaviour, etc), the modes being compared, and the measures obtained in each mode. The mode effect estimate column is highlighted to make it easier to identify and is reported alongside measures of uncertainty (standard error and 95% confidence intervals). Where possible, a standardised effect size is also provided.

The 'Details of each study' section provides additional study-level information. The section contains a brief overview of the sampling and randomisation process, as well as the response rate (*Note*: the latter was frequently not available). It also provides a brief quality appraisal, in particular discussing potential sources of bias and whether sufficient information was provided in the study. This information was collected so that those wishing to use the database can make an informed decision on whether certain

mode effect estimates are useful for their own purposes, given the available evidence and its quality.

Figure 1 provides an example entry of the database for a mode effect reported between paper and tablet responses to the “Did not always wear a seat belt (%)” item from the Youth Risk Behaviour Survey.

Figure 1. A) An example entry in the 'Mode effects' section of the database, showing an estimated mode difference between paper and tablet responses to the "Did not always wear a seat belt (%)" item from the Youth Risk Behaviour Survey. The database contains general information on the study, survey item, mode differences, and associated uncertainty. **B)** An example entry in the 'Details of each study' section of the database, corresponding to the mode effect entry shown in panel A. This section contains further information on the sampling and randomisation design, response rate (where reported), and any potential challenges such as sources of bias or insufficient reporting, which may affect the validity of the results.

A)

Study								Variable				Mode		Measure						
Study	Year of publication	DOI	Source population category	Survey name	Survey population profile	Survey sweep	Country	Category	Sub-category	Variable type	Variable	Reference mode	Alternate mode	Reference mode measure	Alternate mode measure	Effect measure	Mode effect estimate	Mode effect standard error	Mode effect confidence interval (lower)	Mode effect confidence interval (upper)
All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All
Li et al. 2024	2024	10.1016/j.jadohealth.2023.10.032	Cross-sectional survey	Youth Risk Behavior Survey (YRBS)	US public and private school students in grades 9-12 in the 50 states and the District of Columbia	Sweep in 2021	USA	Behaviour	Risk behaviour	Binary	Did not always wear a seat belt (%)	Paper	Tablet	41.70	40.30	Mean difference	*-1.40*	*1.10*	*-3.55*	*0.75*

B)

	Design								Quality and appraisal								
DOI	Source population category	Survey name	Survey population profile	Survey sweep	Country	Sampling	Study design	Study modes	Study response rate	Study response rate by mode	Other exclusions (post-response)	Final sample size	Compliance	Quality of reporting	Selection	Differential item non-response	Other comments
All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All
10.1016/j.jadohealth.2023.10.032	Cross-sectional survey	Youth Risk Behavior Survey (YRBS)	US public and private school students in grades 9-12 in the 50 states and the District of Columbia	Sweep in 2021	USA	Three-stage cluster sampling: counties, schools, and classes. Within each sampled county, three large and one small schools were selected systematically with probability proportionate to enrollment size. In schools agreeing to participate both by paper and tablet questionnaires, classes were randomly sampled.	In the third-stage sampling, classes were randomly assigned to respond by paper or tablet.	Paper, Tablet				8,329 (4,684 Paper, 3,645 Tablet)	Not reported, assumed full	Unadjusted mode effect estimates were not presented. No outcome SDs were reported. Overall response rate was not reported.	Despite the YRBS aiming to be nationally representative, the sample for this experiment was described as not intended to be nationally representative as it only included schools who agreed to participate by both modes. Analyses were unweighted.	30 out of 89 items were reported to exhibit significantly different missingness by mode. 20 had higher missingness in tablet, and 10 - in paper. Item non-response was reported for all variables and differential missingness was discussed.	Odds ratios were adjusted for sex, race/ethnicity and school grade. Although a specific reason was not provided. Randomisation was applied at class rather than student level.

4. References

Burton, J., & Jäckle, A. (2020). *Mode effects*. Understanding Society Working Paper Series. https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2020-05.pdf

Cohen, J. (2009). *Statistical power analysis for the behavioral sciences* (2. ed., reprint). Psychology Press.

d'Ardenne, J., Bull, R., Das, A., Perera, Z., & Sexton, O. (2025). Survey Practice Guide 2: How to mitigate against measurement effects when surveys move online. *Survey Futures Survey Practice Guide.* Colchester, UK: University of Essex. Available at: https://surveyfutures.net/practice-guides/.

de Leeuw, E. D. de, Hox, J. J., Dillman, D. A., & European Association of Methodology (Eds). (2008). *International handbook of survey methodology*. Lawrence Erlbaum Associates.

Klausch, T., Hox, J. J., & Schouten, B. (2013). Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions. *Sociological Methods and Research*, *42*(3), 227–263. https://doi.org/10.1177/0049124113500480

Kocar, S., Brown, M., Sanchez, C., Ploubidis, G., & Durrant, G. (2025). Measurement Mode Effects in Live Video Interviewing: Evidence from a Mixed-Mode Quasi-Experimental Survey. *Survey Futures Working Paper No. 11*. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/.

ONS. (2025). *Personal well-being user guidance*. ONS Website, Methodology,. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide

Sullivan, G. M., & Feinn, R. (2012). Using Effect Size—Or Why the P Value Is Not Enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Tenaglia, S. (2024). Housing and wellbeing: Analysing English Housing Survey data 2013-2020. *What Works Wellbeing*. https://whatworkswellbeing.org/wp-content/uploads/2024/01/Housing-wellbeing-analysis-FINAL-report.docx.pdf

Tomova, G. D., Silverwood, R. J., & Wright, L. (2026). *Mode effects on survey item measurement: A systematic review of the experimental evidence*. *Survey Futures Working Paper No. 12*. Colchester, UK: University of Essex. Available at https://surveyfutures.net/working-papers/.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Wright, L., Tomova, G. D., Pina-Sanchez, J. & Silverwood, R.J. (2026). A Tutorial on Sensitivity Analysis for Handling Mode Effects. *Survey Futures Survey Practice Guide (forthcoming)*. Colchester, UK: University of Essex. Available at: https://surveyfutures.net/practice-guides/.

5. Appendix

Table A1. Evidence and recommendations for mixed-mode design combinations, organised by type of mode comparison.

MODE COMBINATIONS	OVERALL SUMMARY OF EVIDENCE	DIFFERENCES IN ITEM DOMAINS	CAVEATS	RECOMMENDATION
Paper and web	Overall, substantial mode effects were rare.	The evidence was generally uniform across categories. Mode effects were reported for health and wellbeing items slightly more than for social or behavioural items.	There was evidence for large mode effects in large-scale educational assessments (when comparing electronic vs paper completion), however these may not necessarily be relevant to most standard survey practice.	Paper and web modes are broadly comparable, even for sensitive questions. Using a mix of paper and web poses very low risk of substantial mode effects. Transitioning paper surveys to web versions therefore also poses minimal risks of meaningful mode effects.
Face-to-face and paper	Mode effects were generally likely to be observed across item categories, with some domains particularly likely to be subject to mode effects.	Health and wellbeing as well as behavioural items were less likely to exhibit mode effects, compared to attitudinal, adverse experiences, and financial items.	For some categories (e.g. employment and income) evidence was available from only a single study or a few studies.	Using a combination of face-to-face and paper modes should be approached carefully as the risk of mode effects is high. Where possible, consider administering questions most likely to be subject to mode effects in a self-administered way even in an interviewer-led mode.

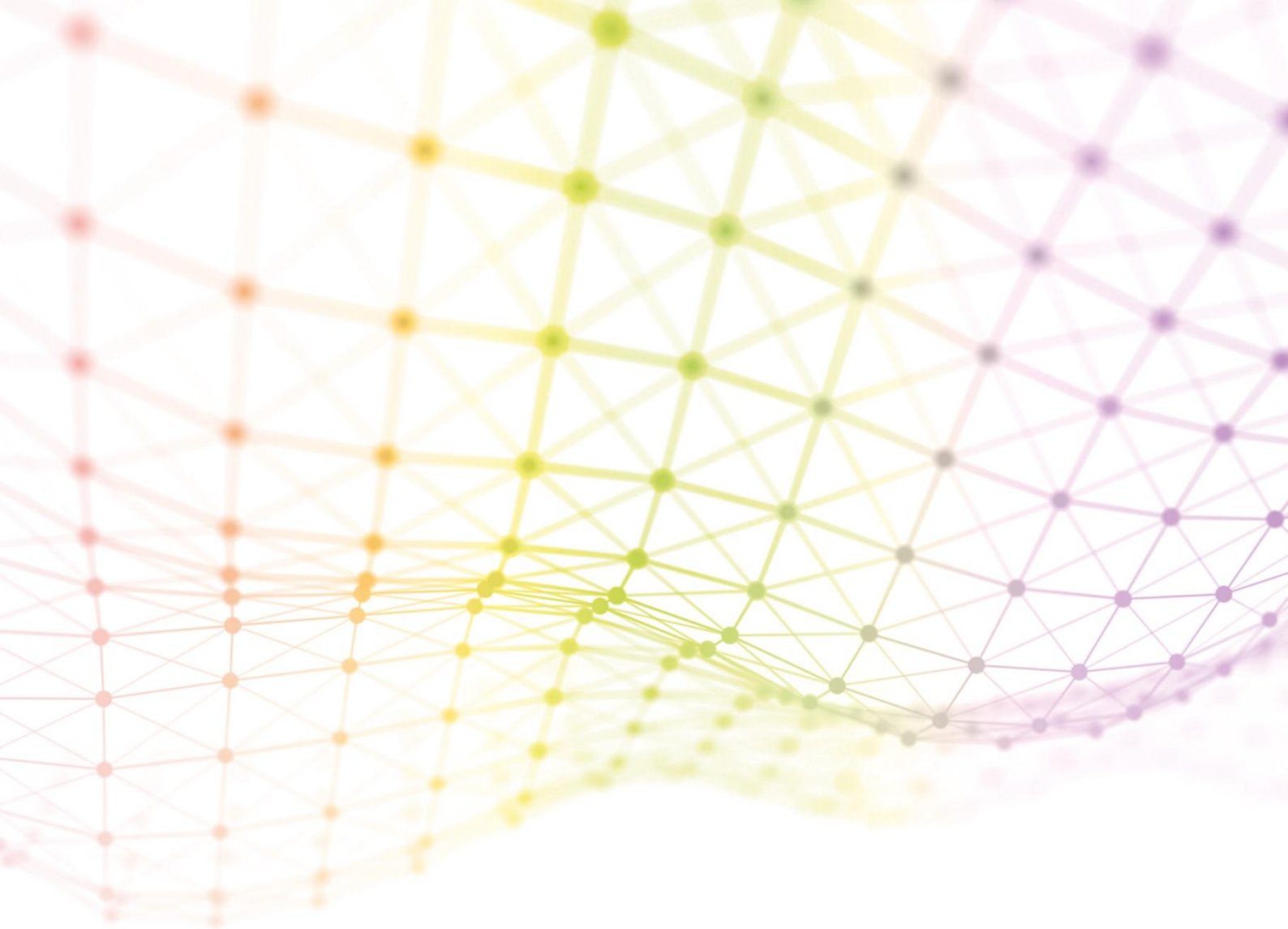
				Since it may not be possible to minimise mode effects at survey design stage, ensure that the appropriate technical documentation explains the risks for mode effects and refers users to methods for handling potential mode effects during analysis.
Face-to-face and telephone	Overall, substantial mode effects were rare.	Behavioural items were more likely to exhibit mode effects, whereas these were rare for attitudinal or health items.	The overall evidence comes from only four domains (adverse experiences, attitudes, behaviour, health and wellbeing) and may not generalise to other settings.	Face-to-face and telephone modes were broadly comparable and pose low risk of substantial mode effects.
Telephone and paper	The combination of these modes exhibited the largest and most commonly occurring mode effects out of all mode combinations.	Around half of attitudinal, health, and social items exhibited substantial mode effects, which is much more common than usual. Mode effects were less commonly reported for behavioural items.	The overall evidence comes from only four domains (attitudes, behaviour, health and wellbeing, social) and may not generalise to other settings. Within the social domain, the	Using a combination of telephone and paper modes should be approached carefully as the risk of mode effects is very high. Since it may not be possible to minimise mode effects at survey design stage, ensure that the appropriate technical

			evidence only came from 10 items.	documentation explains the risks for mode effects and refers users to methods for handling potential mode effects during analysis.
Face-to-face and web	Overall, the combination of face-to-face and web modes commonly resulted in large mode effects.	Substantial mode effects were evident across all available domains, but were particularly severe for social items, where 2/3 of items were subject to large mode effects.	The evidence for mixing face-to-face and web modes was only available for attitudes, health and wellbeing, and social items, and may not generalise to other settings. In addition, within each domain, mode effects were reported for no more than 30 items in total across studies, which is limited.	A face-to-face and web mixed-mode design should be approached carefully due to the high risk of mode effects. When considering the transition of a face-to-face survey to web, follow MERF (d'Ardenne et al., 2025). In addition, ensure that the appropriate technical documentation explains the risks for mode effects and refers users to methods for handling potential mode effects during analysis.
Telephone and web	Overall, mode effects were likely to occur, but the frequency was variable across domains.	While employment and income as well as health and wellbeing items were less likely on average to commonly exhibit large mode effects, this was	The evidence for mixing telephone and web modes was only available for attitudes, employment and income, and health and wellbeing items, and	A telephone and web mixed-mode design should be approached carefully due to the high risk of mode effects. When considering a

		much more common for attitudinal items.	may not generalise to other settings.	transition to web, follow MERF (d'Ardenne et al., 2025). In addition, ensure that the appropriate technical documentation explains the risks for mode effects and refers users to methods for handling potential mode effects during analysis.
Face-to-face (interviewer-led) and CASI[b] or ACASI[a]	Overall, mode effects between a standard face-to-face mode and CASI[†] (or ACASI[‡]) tended to be larger and exhibited frequently across all domains.	Although they were reported for all examined domains, mode effects were particularly likely to be larger for attitudes, behaviours, health, and sexual activity items.	For many of the domains, evidence is only available from a single study (e.g. for health and wellbeing items and employment and income items), and may not be sufficient or generalisable to other settings.	Transitioning from a standard interviewer-led face-to-face interview to CASI[b] (or ACASI[a]) should be approached carefully due to the high risk of mode effects due to the different level of interviewer involvement. Ensure that the appropriate technical documentation explains the risks for mode effects and refers users to methods for handling potential mode effects during analysis.

Paper and CASI[b] or ACASI[a]	Overall, mode effects were least likely to occur for this combination of modes.	Items relating to behaviour, sexual activity, or adverse experiences were virtually free of substantial mode effects according to the evidence. However, health and wellbeing items were a little more likely to be subject to mode effects.	The evidence for some domains came from single studies or from a small number of reported items (e.g. adverse experiences, employment and income) and may not generalise. However, this is also mitigated by the fact that for this combination of modes, mode effects were consistently unlikely to occur.	Paper and CASI[b] (or ACASI[a]) modes were generally comparable and therefore pose low risk of substantial mode effects.

[a]Audio computer-assisted self-interviewing, [b]Computer-assisted self-interviewing



University of Essex

University of Southampton

UKRI Economic and Social Research Council

NCRM National Centre for Research Methods

Office for National Statistics

UCL

Warwick The University of Warwick

Manchester 1824 The University of Manchester

City University of London Est 1894

National Centre for Social Research

LSE The London School of Economics and Political Science

Ulster University

Unil UNIL | Université de Lausanne

Ipsos

verian Formerly Kantar Public

www.surveyfutures.net