

ESS round 10 mode experiments in Great Britain and Finland. Findings on mode effects

Peter Lugtig

Why estimating mode measurement effects is difficult

One of the major problems in estimating the mode measurement effect in (mixed-mode) surveys is that isolation of the causal effect of mode on measurement is difficult. Using face-to-face or postal recruitment for the survey will result in different types of respondents that participate in the survey, and differences in selection effects that are hard to separate from the fact that respondents also react differently to the mode of the interview. Some studies have tried to eliminate selection effects by for example re-interviewing face-to-face respondents in a self-interviewing mode shortly after the original interview (Klausch et al., 2014). Or they have randomized respondents into a survey mode only after successfully recruiting a respondent into the survey (Heerwegh, 2009). Both these designs are a bit artificial, and are in practice always complex to explain and administer to respondents. The ESS did not use such a complex design in their mixed-mode experiments in round 10.

Mode experiments in Round 10 of the ESS and ways to estimate mode measurement effects

ESS round 10 was conducted with a self-completion instrument (web and paper) in 9 countries, with the 22 other countries using face-to-face interviewing. In Finland and Great Britain, a within country experiment was conducted, where the samples were randomly assigned to one of two modes: one part of the sample was recruited and interviewed face-to-face, while the second part of the sample was assigned to postal recruitment and self-interviewing (push-to-web). These differences in interview modes both within countries and between countries, along with comparisons to data collected in earlier rounds provide the ingredients for estimating the mode measurement effects.

Several kinds of comparisons can be made:

1. For the 9 countries using self-completion, we can estimate the change in data associated with the switch from face-to-face interviewing (before round 10) to self-interviewing at round 10 for variables that have been collected at every wave. Apart from the problem that the change in mode between rounds 9 and 10 will affect both measurement and nonresponse, a third problem is that aggregate change between rounds 9 (collected before Covid) and round 10 (collected during Covid) is likely to occur as well for several variables. For the countries that used face-to-face interviewing in both rounds 9 and 10, we can estimate the same type of change however, and comparing both sets of change (countries switching towards self-interviewing vs staying in face-to-face) we can still learn something about how statistics change by moving to a self-interviewing mode. These comparisons are made in a companion report to this one, and can be found in Lugtig (2024a). Because findings in this report partly build on the longitudinal analyses, the reader is referred to this companion report for more information about the size of mode effects across the entire European Social Survey.

2. For Great Britain and Finland, we can compare respondents from the within country experiments on all variables collected in ESS round 10 common to both modes. In this comparison, we still have the problem that answers may differ due to nonresponse and measurement, but there is no “time effect” or “Covid-effect”. This document will compare the self-completion experiments carried out in Great Britain and Finland to their round 10 face-to-face fieldwork, and study in more detail whether changes in means that were found for about 25 variables in the companion report are more likely to be caused by mode selection effects (different respondents in different modes) or mode measurement effects (different answers in different modes). In Finland, we are basing our analyses on all respondents which for both face-to-face and self-completion are aged 15 or older. In the United Kingdom, we have restricted our analysis to Great Britain (i.e., excluding Northern Ireland) and to respondents aged 18 or older to match the geographic scope and age eligibility of the self-completion data collection.

Table 1: Effects that can lead to a difference in interview modes in ESS

Comparisons	Time and Covid-effect	mode Selection effect	Mode measurement effect
8 countries: round 10 vs round 9 - face-to-face vs self-interviewing	xxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxx
20 countries: round 10 vs round 9 - face-to-face vs. face-to-face	xxxxxxxxxxxxxxxx		
2 countries: round 10 within-country experiment - face-to-face vs self-interviewing		xxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxx

The size of raw mode effects in Finland and Great Britain

The companion report (Lugtig, 2024a) compared the sizes of the overall differences between the mode used in rounds 9 and 10. One of the core findings was that mode effects are mainly found in means, but not so much for variances and covariances. The cause of this shift in means when the mode of the survey changed from face-to-face to self-completion is unclear, and could be different for different variables: for example, social desirability, primacy, or recency effects can cause shifts in the overall distribution. However, it is also possible that at least a part of the mode effect can be explained by different selection effects in the survey modes (both nonresponse and coverage differences). Because the different survey modes attract different types of respondents, we find differences on our dependent variables. As a first step in our analyses, we repeat an analysis of the size of mode effects across all variables that were used in ESS rounds 9 and 10 for the experimental dataset in Finland and Great Britain.

The first step in the analyses is to investigate whether mode effects in Finland and Great Britain are similar to those found on aggregate in the 9 countries switching to self-interviewing in round 10 of the ESS. For this, we concentrate on reporting Standardized Mean Differences in the form of Hedges g . Table 1 below shows that the median mode effect is about 0.02, but in absolute terms it is 0.12, which amounts to a small effect on average. Mode effects appear to be on average about the same size as in the longitudinal analysis conducted on round 9 and 10 data (median absolute effect size: 0.11). It is important however to keep in mind that the sample sizes in the current experimental data analysis are quite small, and for that reason, effects may be a bit more variable in this report than in the companion report studying the effects across 28 countries.

Table 2: Distribution of Hedges g for mode effect on means in experimental ESS data

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Hedges g mode effect FI and GB	-0.36	-0.11	-0.02	0.04	0.18	0.51
Absolute Hedges g mode effect FI and GB	0.01	0.06	0.12	0.16	0.24	0.51

Covariate adjustment: procedure and covariates

Because we have experimental data in both Great Britain and Finland, we can dig deeper into the possible causes of the mode effect than was possible when we compared all countries across round 9 and 10. In particular, we can investigate whether the differences in means we observe between the face-to-face and self-interviewing experimental conditions can be explained by mode selection effects.

One advantage of using the within-country experiments conducted in Finland and Great Britain is that there is no effect of time that can lead to change or difference between modes. Vannieuwenhuyze, Loosveldt & Molenberghs (2014) explain different ways in which mode selection and mode measurement effects can be disentangled for the experimental mode setups as encountered in Finland and Great Britain; covariate adjustment models (called ‘back door’ methods) are in such situations recommended to 1) estimate the mode selection effect difference between the two experimental samples and 2) then adjust for mode selection differences. The remaining difference between the face-to-face and self-interviewing samples can then be interpreted as 3) the mode-measurement effect.

The analyses in this paper to implement the covariate adjustment model are as follows:

1. In a first step of any analyses into mode effects, it is important to identify which covariates are related to mode selection and outcome variables in the analyses. Only variables where no mode measurement effect should be present were included as potential covariates. We tried to account for this in our analyses, but of course are never entirely sure whether mode measurement effects are absent or not. This led to about 20 variables that we considered to use to correct for selection effects. The 20 covariates were regressed on the variable ‘mode’, and several outcome variables. Interactions with the variable ‘cntry’ were included to take account of possible country specific effects. We finally selected variables that were both strongly related to mode selection, and at least some of the outcome variables.
2. After identifying these variables, adjustments for selection effects can be made. Given the relatively large size of the samples, we will use statistical matching (as used by Lugtig et al, 2011) to condition the self-interviewing and face-to-face respondents separately for Finland and Great Britain on the variables related to mode-selection, in order to create a subset of respondents that will be identical on the covariates used to explain nonresponse differences.
3. After this, the matched sample will include “identical respondents” that will only differ by the interview mode that was used. That is, respondents unique to the face-to-face sample or self-interviewing sample that are not found in the other sample, are discarded from the dataset so that we can estimate the measurement effect for respondents we find in both the self-interviewing and face-to-face survey.
4. A series of analyses will then compare matched respondents from both interview modes. In our analyses, we will first use all numeric variables that were used in both ESS rounds 9 and 10, as well as the self-completion experiments in round 10, and did not contain a lot of missing values. 105 variables were used in the analyses

After these steps we selected 7 variables to be used in matching:

- gender of the respondent (variable name: ‘gndr’)
- educational attainment in years (variable name: ‘edyrs’)
- age (variable name: ‘agea’)
- country (Great Britain or Finland)
- reporting to have no internet access (variable name: ‘accnone’)
- Whether the respondent already had covid at the time of the survey (variable name: ‘respc19’)
- Whether someone in the household already had covid at time of the survey (variable name: ‘resphh19’)

For the last two variables, it is a possibility that mode measurement effects exist. However, both predictors were strongly related to mode selection, and were therefore included in the matching model.

Matching procedure

The next step in our analyses is to match respondents in both Finland and Great Britain on the set of covariates. For this, we used Coarsened Exact Matching as implemented in the package `{MatchIt}` in R. Age was split into 10 categories of equal size (deciles) to facilitate the matching. Matches could be multiple-to-one in case there were multiple matching candidates with the exact same respondent profile in a mode.

Not all respondents could be matched, and not all matches could be made exact. Covariate balance was improved with 98%, implying that almost all matches were exact, except for a few respondents. From the 2677 respondents who were interviewed face-to-face in Finland and Great Britain combined, 2328 could be matched to a respondent who was interviewed via self-completion (paper and web). From the 3928 self-interviewing respondents, fewer respondents could be matched: a total of 2862 matches were made. We end up with four groups of respondents: respondents who could be matched or not be matched and were interviewed either in the face-to-face or self-interviewing mode.

Analyses on matched dataset

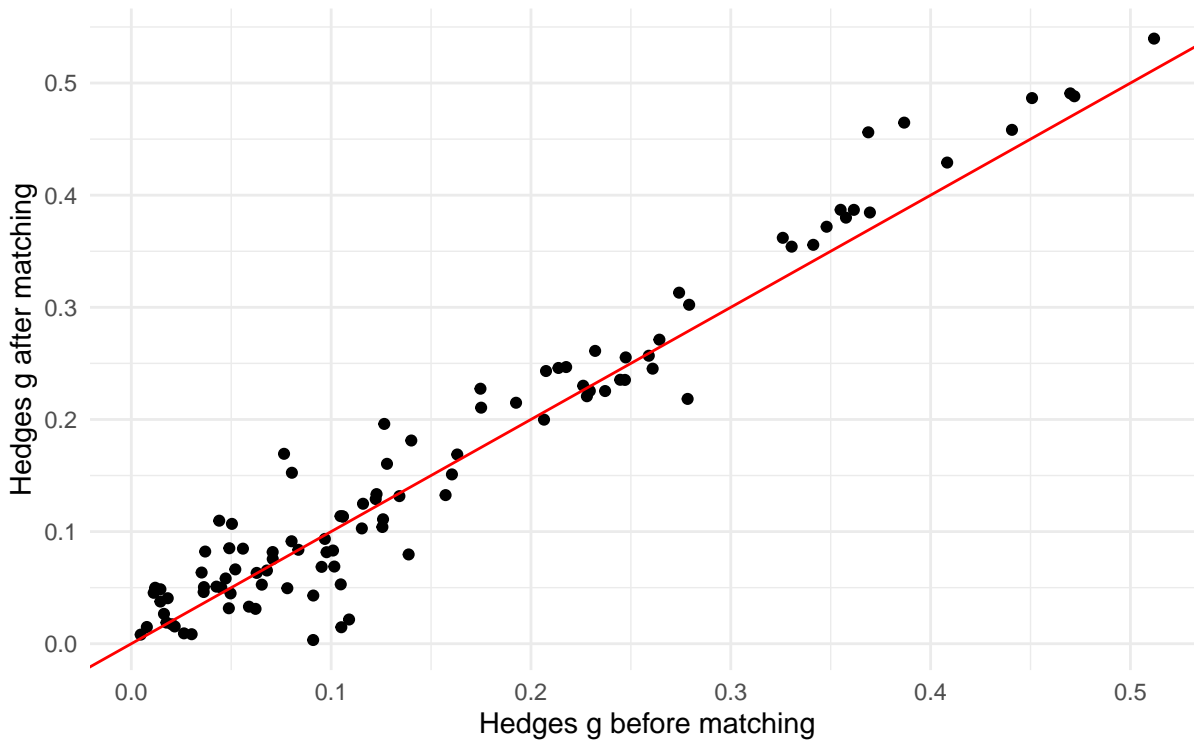
The next step in our analyses is to compare mode effects again in Finland and Great Britain, but now focusing on respondents from self-interviewing and face-to-face interviewing who could be matched and who could not be matched. Because we already showed that differences between Finland and Great Britain are relatively small, we will pool the analyses here across the two countries.

Figure 1 shows the mode effects between face-to-face and self-completion before and after respondents were matched. We hope to see that after matching, selection effects that were present before matching disappear. The mode effects for matched respondents should ideally approach 0. Selection differences between unmatched respondents across the face-to-face and self-interviewing groups should become larger compared to before matching.

What figure 1 shows however is that selection effects can only explain a small part of the mode effects present. On average, selection effects account for only 7% of the observed mode effect on average. This means either that our covariates do not explain selection effects adequately and/or that part of the mode effect can be explained by measurement effects. We do see that for some variables selection effects can fully be explained away after matching the samples on the covariates, but we also find that for some variables, mode effects even increase after matching. Sensitivity analyses on the covariates used in the matching procedure and the matching results did not change these results. Given these findings it seems unlikely that selection effects alone can explain the mode effects. Measurement effects do seem to play a role. We refer to the companion report (Lugtig, 2024a) for illustrations of the mode effects, and some ideas on why the size of mode selection effects and mode measurement effects may be different across variables. Because the sample sizes for the entire ESS are much larger than in the analysis conducted in Finland and Great Britain, these analyses are more suitable for zooming in on the effects of mode on specific points in the distribution of variables. Matching does reduce the average size of mode effects in the experimental data by about 7%, but the absolute median effect size remains more or less the same at about 0.11 in both groups.

Figure 1: Mode effects on means before and after matching

Hedges g for self-completion & face-to-face before matching (x-axis) and after (y-axis)



Conclusion

This report shows that the mode effects that we found when comparing data from round 9 and round 10 of the ESS largely hold for the experimental data collected in Finland and Great Britain. We find, on the whole, similar sizes of the mode effects. The big advantage of comparing the experimental data in this report over the longitudinal results is that we can exclude time as a factor in the occurrence of the mode effect between face-to-face and self-interviewing. We can also exclude the possibility that country-specific reasons for making a switch to self-interviewing in round 10 of the ESS are a cause of changes in means. The data in Great Britain and Finland collected using either the face-to-face mode or self-interviewing were collected at the same time, and assignment to mode was truly experimental. This means that there are only two potential causes of mode effects: selection effects because of the fact that the different survey modes attract different types of respondents, and mode measurement effects that results from differences in the question-answer process of participating in different surveys.

The size of mode effects in Finland and Great Britain appear to be a little more variable than based on the longitudinal analyses conducted in Lugtig (2024a), but on aggregate, the size of the mode effects in absolute terms are quite similar (median hedges g-value in this report 0.12, vs. 0.09 in companion report). Our conclusion is therefore much the same: for most variables, a transition towards self-interviewing will not impact the distributions of variables much. The companion report showed that there are no large effects on variances and covariances for any variables. This does not imply there are no problems. We see that for a minority of variables, about 20% of all tested, effect sizes are about 0.20 or more in terms of Hedges g-values, implying a meaningful change. The causes for these changes are likely to be caused by a combination of causes. First, it is quite likely that mode selection effects cause some of the differences we find.

In this report, we used statistical matching to condition respondents from the face-to-face and self-interviewing modes on a set of 7 covariates that were found to be predictive of both mode selection, and some of the outcome variables where we found strong mode effects. After matching, we compared

matched respondents from both modes again, and find that selection effects are reduced by about 7% after conditioning on the covariates. Although it is quite likely that there are selection effects between both modes that we failed to account for using the available covariates in the ESS, it is highly unlikely that all mode effects can be explained by selection effects. The presence of an interviewer (or not) is another likely cause of mode differences, especially with socially sensitive questions. Both selection effects, and the change of the presence of an interviewer are effects that are inherent to a mode change, and are hard to eliminate. Questionnaire design choices may be a final cause of mode effects.

The transition towards self-interviewing that the ESS is taking on in the next 5 years will benefit from more detailed assessment of these mode effects for individual questions. From a theoretical perspective, we could expect some types of questions (e.g. with long answer scales, or those sensitive to socially desirable answer behavior) to exhibit stronger mode measurement effects. It would be worthwhile to systematically code question characteristics based on theories on mode measurement differences (presentation, communication, answering process) and then relate these question characteristics to the size of the observed mode effect. This would pinpoint the cause of mode effects in more detail, and also provide clues about what to potentially do to reduce mode measurement effects.

Related to this, this report did not answer the question of how mode effects can be resolved. In some cases, a review of question content and cognitive testing could provide in-depth information on whether measurement effects are present, and whether these are perhaps amenable by changing the ESS questionnaire. It is also worthwhile to spend more energy on understanding selection effects and nonresponse bias in countries where the ESS is making the switch from face-to-face to self-interviewing. Experiments where the assignment of mode at the level of the sample is useful for further understanding selection and nonresponse bias, especially when information is available on all sample members. Experiments where the interview mode is assigned to a respondent after he/she has agreed to participate would be very useful to investigate measurement effects in more detail.

References

- Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21(1), 111-121.
- Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological methods*, 17(2), 137.
- Klausch, T., Schouten, B., Buelens, B., & Van Den Brakel, J. (2017). Adjusting measurement bias in sequential mixed-mode surveys using re-interview data. *Journal of Survey Statistics and Methodology*, 5(4), 409-432.
- Lugtig, P. (2024a). Comparison ESS rounds 9 and 10: mode differences. ESS internal report, July 2024
- Vannieuwenhuyze, J., Loosveldt, G., & Molenberghs, G. (2014). Evaluating mode effects in mixed-mode data through the back-door and front-door. *Journal Of Official Statistics*, 30(1), 1-21.