**Adapting ESS survey questionnaires to mixed-mode data collection**

Dag F. Gravem, Statistics Norway, dfg@ssb.no

**Abstract**

*The purpose of WP4 in the MIMOD project is to recommend approaches for designing web questionnaires for mixed-mode surveys. Inherent differences between survey modes can lead to measurement differences affecting the quality of national and cross-national estimates. By analysing questions looking for characteristics that are associated with such “pure” mode effects, questions can be pretested and redesigned with the aim of reducing such effects. WP4 is analysing model and national questionnaires for key ESS surveys, and has found discrepancies between Eurostat recommendations and national questionnaires. On a European level, all key ESS surveys are conducted using all main modes. WP4 has also found discrepancies between Eurostat recommendations and national implementations on the one hand, and mixed-mode research on the other. The paper presents two examples of questions that need adaptation to function well in mixed-mode data collection, and that will be pretested at a later stage. The paper concludes with the recommendation that all future Eurostat model questionnaires should be developed as mixed-mode model questionnaires rather than today’s practice of designing them with a preferred mode in mind.*

**Keywords:** Mixed-mode, model questionnaires, mode effects, pretesting

**1. The purpose of MIMOD WP4**

Within the MIMOD project, the purpose of WP4 is to offer recommendations on approaches for developing web questionnaires for mixed-mode surveys in the European Statistical System (ESS). These recommendations will be on a survey level, as well as for key questions and question types.

The recommendations are intended to reduce the risk of measurement errors, or measurement differences, due to *mode effects*. By mode, we mean the method of data collection. The different data collection modes in social surveys include

* Computer-assisted personal interview (CAPI)
* Computer assisted telephone interview (CATI)
* Computer-assisted web interview (CAWI)
* Paper and pencil interview (PAPI)

Mode effects can be divided into *selection effects*, due to different demographic groups having a higher response propensity in one mode than another, and *measurement effects*, due to inherent differences between data collection modes themselves. The latter has also been termed *“pure” mode effects*. (Körner et al., 2014)

Disentangling and correcting for both types of mode effects is the task of WP2. (Luzi and Buelens 2018) In WP4, the aim is to identify questions that are likely to suffer from *measurement effects* and to find possible ways to prevent errors or measurement differences from occurring in the first place.

**2. Theoretical approaches**

There are several theoretical approaches for studying “pure” mode effects. Some take the different characteristics of the *modes* as a starting point. Körner et al. (2014) differentiate between

* Visual and aural modes
* Interviewer-administered and self-completion modes
* Computer-assisted and non-computer assisted modes
* Degrees of social contact, face-to-face (CAPI), via telephone (CATI), and none (CAWI)

E.g. in an aural mode like CATI, there is a limit to the amount of information your working memory can process, requiring shorter questions and fewer answer categories than visual modes.

Another approach is to take different *question* characteristics as the starting point, and evaluate how fit individual questions will be for the different modes. Pamela Campanelli et al. have provided us with a 2013 unpublished manuscript *A Classification of Question Characteristics Relevant to Measurement Error and Consequently Important for Mixed Mode Questionnaire Design* (figure 1, appendix A), which singles out 29 different characteristics and provide evidence-based mode recommendations on which modes to use.

The main characteristics dimensions are

* Question content: sensitive, e.g. factual, subjective, degree of difficulty
* Type of task: e.g. open and closed questions, scalar, yes/no vs. mark all that apply
* Characteristics of task: e.g. number of categories, middle categories, end-labelling, branching
* Implementation of task: e.g. use of instructions, edit checks, use of don’t know options and formatting

**Figure 1. Overview of the classification of question characteristics**



In MIMOD WP4, we adapt this typology to determine the fitness of question types and questionnaires for mixed-mode data collection involving CAWI. Most of the questions we analyse will have more than one of the characteristics in Campanelli et al.’s typology, and the mode recommendations can contradict one another. For each question or question type, these contradictions need to be discussed, and if possibe avoided when the question is redesigned.

*2.1. Experimental theory and survey organisation practice*

De Leeuw and Hox (2015) point out that empirical mode comparisons show relatively small mode effects, except for questions on sensitive topics. Here, self-completion modes produce better results because the presence of an interviewer will cause some respondents to answer less candidly.

However, they also point out that this could be because experimental survey researchers take great care to ensure equivalence across tested modes. In non-experimental practice, such as in the ESS NSIs, this is not always the case. Questions may have been developed and adapted for one specific mode, and function poorly when asked identically in a different mode. WP4 is looking for possible NSI experiences of these kinds.

*2.2. Question design philosophies*

A typology of four question design philosophies for mixed-mode surveys has been suggested (de Leeuw (2008); Dillman, Smyth and Christian (2009)):

* Unimode: the questions are as similar as possible, regardless of mode. May also mean that the questions are suboptimal in each mode. The rationale is that this will minimize mode effects.
* Mode-specific: questions are adapted and adjusted to fit the format and specific needs of each mode
* Mode-enhanced design: the strength of one specific mode is used. This may lead to lesser quality data from other modes
* Generalised mode: questions are designed differently in different modes but intended to elicit the same cognitive reaction. Mode-enhanced for each mode.

In the literature, these design philosophies may be described somewhat differently, and there are gradual transitions between the categories. With reference to survey organisation practice referenced above, a clear design philosophy is not always discernible in NSI questions and questionnaires. The grant call nevertheless states that the MIMOD project should offer recommendations on whether to use a unimode or a mode-specific design philosophy.

**2. Eurostat requirements and recommendations versus NSI practices**

In WP4, we examine Eurostat’s mode and question requirements and recommendations for the ESS social surveys. We also look at the ways different European NSIs have solved this in practice, given their historical, technological, financial, demographic, etc. conditions, in order to identify good and bad practices. This paragraph sums up some preliminary results from this work.

For the surveys European Health Interview Survey (EHIS), Adult Education Survey (AES) and the Community Information and Communication Technology (ICT) survey, model or reference questionnaires are provided by Eurostat. The Labour Force Survey (LFS) and the EU Statistics on Income and Living Conditions (EU-SILC) survey have extensive methodological guidelines and minimum requirements, but no model questionnaires. So far, EHIS, the ICT survey and the EU-SILC have been studied in some detail.

*2.1. EHIS*

The EHIS is model questionnaire explicitly designed for face-to-face modes (“preferred modes”), including suggestions for show cards. Adaptation to other modes is left for each ESS country to decide, although a separate report contains recommendations for CATI modes on some questions. The rationale for preferring interviewer-administered modes seems to be are needed for the clarification of difficult medical terms, and for assisting in calculating e.g. alcohol consumption. From a survey methodology perspective, self-administered modes would be more recommendable for such questions.

The MIMOD survey shows that interviewer-administered CAPI is indeed the most widely used mode for the EHIS survey, by 15 of 28 countries. Self-administered CAWI comes in at second place with 11. Six countries use a combination of these two modes, and CATI and PAPI is also in use. This tells us that Eurostat’s preferences are not necessarily the preferred modes in practice. It also tells us that there is a risk of mode effects on sensitive questions in many countries.

*2.2. ICT survey*

There are no specific mode guidelines or recommendations for the ICT survey, but the layout of its model questionnaire presupposes a visual mode, e.g. by referrals to the “tick all that apply” response format, which must be substituted with a battery of yes/no questions or an open question with interviewer coding. In CAPI interviews, show cards can be used as a visual aid.

The amount of text in the questions and response categories is also better suited for visual modes, as it will be a heavy load on respondents’ working memory in CATI.

Nevertheless, CATI is the most widely used mode for the ICT survey with 16 countries, followed by CAWI which is used by 15 countries. Five countries report using a CATI/CAWI mixed-mode combination – in other words one strictly aural and one strictly visual mode.

*2.3. EU-SILC*

The EU-SILC methodological guidelines states that five modes of data collection are possible for the survey: PAPI, CAPI, CATI, self-administered (paper) by respondent and CAWI, although the latter is not covered in the EU-SILC’s legal basis. (p. 79) Further, priority is to be given to personal interviews (PAPI, CAPI) over the other modes of data collection. CATI has however been allowed on a “gentleman’s agreement” basis for countries with person samples. The motivation given for this is that the interview *length* will be shorter as the whole household is not interviewed.

The document goes on to describe the various survey variables, and many of them have recommendations for implementation. These recommendations lack any discussions of how to implement the questions for different modes, let alone for a mixed-mode survey. Nevertheless, they presuppose interviewer-administered modes, as there are several references to response categories having to be read out loud.

Although CAPI is the most used mode for the EU-SILC (20 countries), the gentlemanly option is widely used (11 countries in 1st wave and 15 in consecutive waves), and the non-legally based CAWI is also gaining in importance (7 countries in consecutive waves).

*2.4. Summary*

When comparing the documentation of the ESS social surveys with how these surveys are conducted in practice, and with recommendations found in literature for e.g. questions on sensitive topics, some clear discrepancies are observable. The ESS countries are not necessarily following Eurostat’s recommendations, and Eurostat’s recommendations are not always in line with research on mode effects.

**3. Questions singled out for further testing**

In the model questionnaires and national questionnaire implementations we have reviewed so far, some question types have been singled out based on Campanelli et al.’s typology because there is a risk of measurement error in a mixed-mode context. This autumn, we will work on adapting the questions for mixed-mode involving web, and pretest different versions in the cognitive labs of WP4’s participating countries, using cognitive interviewing and usability testing techniques.

In this paragraph, we present two examples of the questions that may be included in these tests.

*3.1. An example from the ICT model questionnaire*

**Figure 2. Example from the ICT model questionnaire**



Figure 2 shows Q3 from the ICT survey’s model questionnaire. It presupposes a visual survey mode, and the “check all that apply” format is not possible in interviewer-administered modes. This corresponds with characteristic 14 in the Campanelli typology. This format has been shown to yield lower endorsement rates than the yes/no for all format. There is also some evidence of primacy effects, i.e. the top categories are picked but not the bottom ones, when respondents spend less than average time on the item. A yes/no format would on the other hand force respondents to actively assess each question.

A second issue with Q3 is the possible inherent difficulty of the concepts used, which is included in characteristic 5 in the Campanelli typology. According to literature, one can expect more satisficing in self-completion than in interviewer modes, and more satisficing in CATI than in CAPI modes due to the social interaction, motivation and aid offered by the interviewer.

In this question, however, it is also worth noting that the questions look more like lists of theoretical variables rather than fully operationalized questions. E.g. “create or edit electronic documents” may be easier to answer when supplemented with “Word” and “Excel” as examples. This is done e.g. in the Dutch CATI/CAWI National questionnaire, but not in the Norwegian CATI single-mode questionnaire.

A third issue here, which is not covered by the Campanelli typology, is question and response category *length*. In aural modes, extensive question length places high demands on the respondents’ working memory and should be avoided.

*3.2. An example from the EHIS model questionnaire*

**Figure *3*. *Example from the EHIS model questionnaire with show card***



Figure 3 shows a question on alcohol consumption from the EHIS model questionnaire. Alcohol consumption is generally considered a sensitive question, which is characteristic 1 in Campanelli’s typology. The recommendation is clearly to use a self-completion mode. In a CAPI interview, the authors recommend a protocol of handing over the computer to the respondent for self-completion on such items. Further, they recommend pre-testing across modes if CATI is to be used to assess possible mode effects.

A second issue with this question is its use of numeric bands, characteristic 13 in the typology. The use of show cards is not feasible for CATI mode, and Campanelli et al. recommend a branching implementation instead, by splitting it into several questions. Such an implementation is described in a grant report which is briefly referred to in the EHIS technical documentation. (Finger et al. 2011, Figure 4) In NSI practice, Statistics Norway made an adaptation quite similar to this for their CATI only EHIS questionnaire.

**Figure 4. Branching recommendation for EHIS question on alcohol consumption**



Additional issues with the question include its length, and difficulties of recall and calculation, covered by Campanelli characteristic 5. For such questions, the authors recommend CAPI mode for motivation and support to avoid satisficing.

**4. Conclusions and future work**

MIMOD WP4 will continue to analyse the model questionnaires, specifications, requirements and other documentation of the key ESS surveys, as well as national implementations. A selection of questions and question types likely to have mode effects will then be pretested and adapted this fall, based on up-to-date literature and NSI documentations.

The work done within WP4 so far has shown that discrepancies between Eurostat recommendations and national implementations on the one hand, and mixed-mode research on the other. Likely, some conclusions from WP4 will contradict recommendations inn ESS survey documentation and specifications, where one mode is often favoured over others. Practice shows that modes *will* be mixed in many different combinations, and it is the view of this author that all future ESS model questionnaires should be developed as mixed-mode model questionnaires. In the words of Dillman, Smyth and Christian:

*If there is even a small chance of mixing modes in the project, design the questionnaire for the possibility of mixed-mode data collection.[[1]](#footnote-1)*

**5. References**

Campanelli, P. et al. (2013), A Classification of Question Characteristics Relevant to Measurement Error and Consequently Important for Mixed Mode Questionnaire Design, Presented 2011 at the Royal Statistics Society, London, UK

Community Survey on ICT Usage in Households and by Individuals – 2018 Model Questionnaire version 1.4 (2018)

De Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008), International handbook

of survey methodology. NY: Lawrence Erlbaum Associates.

De Leeuw, E. D., Hox, J. J., De Leeuw and Hox (2015), Survey Mode and Mode Effects, in Engel, U. (2015) pp. 22-34.

Dillman, D. A., J. D. Smyth and L. M. Christian (2009):Internet, Phone, Mail and Mixed-Mode Surveys – The Tailored Design Method, John Wiley & Sons, New York.

European Health Interview Survey (EHIS wave 3) Methodological manual (2018 edition), Eurostat manuals and guidelines, Luxembourg.

Engel, U. et al. (eds.) (2015), Improving Survey Methods – Lessons from Recent Research, Taylor & Francis, London.

Finger, J. et al. (2011), Improvement of the European Health Interview Survey (EHIS) modules on alcohol consumption, physical activity and mental health: Final report, European Commission, Berlin.

Körner, T. (2014), Report on the definition, identification and analysis of mode effects. Deliverable for Work Package III of the ESSnet on Data Collection for Social Surveys Using Multiple Modes, Federal Statistical Office Germany, Wiesbaden.

Luzi, O. and B. Buelens (2018), Dealing with mode effects, paper presented at the Q2018 conference, Krakow, Poland.

Methodological guidelines and descriptions of EU-SILC target variables – 2018 operation (Version August 2017), European Commission.

**6. Appendix A – list of question characteristics relevant to measurement error according to Campanelli et al.**

*Question content*

1 Sensitive questions

2 Factual, non-sensitive questions

3 Subjective, non-sensitive

4 Subjective non-sensitive scalar questions

5 Inherent difficulty due to concept and comprehension, and recall issues

*Type of task – open questions*

6 Unconstrained textual/verbal questions

7 Open questions requiring a number

8 Open questions requiring a date

9 Open questions requiring a short textual/verbal response

10 Open questions with interviewer coding

*Type of task – closed questions*

 11 Agree-disagree scales

 12 Unipolar and bipolar rating scales

 13 Numeric bands

 14 Mark all that apply

 15 ‘Yes/No’ for each

 16 Ranking

 17 Battery of ranking questions

 18 Visual analogue

*Characteristics of task*

 19 Use of middle categories

 20 Number of response categories

 21 End-labelling

 22 Branching

*Implementation of task*

 23 Use of instructions, probes, clarification etc.

 24 Edit checks

 25 Spontaneous ‘Don’t know’ in interviewer modes

26 Size of answer box/text field and delineation of answer space for fully open questions

27 Formatting response boxes for numbers, dates and short textual/verbal responses

 28 Formatting of response lists for closed questions

29 Showcards in Face-to-face interviews for long lists of nominal/ordinal response options

1. Guideline 11.7 in *Internet, Phone, Mail and Mixed-Mode Surveys – The Tailored Design Method* by Don A. Dillman, Jolene D. Smyth and Leah Melani Christian [↑](#footnote-ref-1)