*Understanding Society*

**Working Paper Series**

**No. 2017 – 04**

**April 2017**

# Assessing the risk of mode effects

## Review of proposed survey questions for waves 7-10 of *Understanding Society*

**Jo d'Ardenne, Debbie Collins, Michelle Gray, Curtis Jessop, Sophie Pilley**

**NatCen Social Research**

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

## Non-technical summary

Mixed mode surveys involve the same set of survey questions being presented to participants using different means. In *Understanding Society* the main modes are face to face and web. However, differences in how questions are presented to survey participants – whether the question is read out by an interviewer or presented on a screen - can affect the answers people give. This is known as a mode effect and it can have implications for how survey results are interpreted.

Learning more about mode effects is important, particularly being able to identify and take account of these effects when analysing survey data. This project contributes to this effort. Researchers at NatCen Social Research developed a set of criteria to assess the risks of mode effects for each question to be included in waves 7 to 10 of *Understanding Society*, excluding those that would be asked in the self-completion module. All questions were assessed and assigned a rating that identified whether the question was at no, low, medium or high risk of mode effects. This report describes the development of the risk assessment approach.

# Assessing the risk of mode effects

# Review of proposed survey questions for waves 7-10 of *Understanding Society*

Jo d'Ardenne, Debbie Collins, Michelle Gray, Curtis Jessop, Sophie Pilley

**Abstract**

Learning more about mode effects is important, particularly being able to identify and take account of these effects when analysing survey data. This project contributes to this effort. Researchers at NatCen Social Research developed a set of criteria to assess the risks of mode effects for questions to be included in *Understanding Society* waves 7 - 10, excluding those included in the self-completion module. All questions were assessed and assigned a rating that identified whether the question was at no, low, medium or high risk of mode effects. This report describes the development of the risk assessment approach.

**Keywords:** mixed mode surveys; mode effects; coding risks

**JEL Classification:** C81, C83

**Correspondence:** Joanna.D'Ardenne@natcen.ac.uk

## Contents



## Tables

# 1 Background

The *Understanding Society* scientific leadership team at the Institute for Social and Economic Research (ISER), University of Essex have a programme of methodological research designed to assess the potential impact of mode effects and develop strategies to mitigate their impact on survey findings. As part of this programme of work NatCen Social Research undertook a review of proposed questions to be asked in waves 7-10 of *Understanding Society* to identify questions that might be 'at risk' of mode effects and to categorise what features of a question make it at risk. The focus of the review was to identify questions that could have mode effects if they were asked in both face-to-face and web modes. From 2016 *Understanding Society* has included a mixed mode design in which a proportion of the sample is invited to participate in the survey via the web first, with a face to face interviewer follow up of web non-respondents.

The review was conducted by researchers within NatCen's Questionnaire Development and Testing (QDT) Hub. The final output of this project is a spreadsheet documenting each question, whether each question has any features that are associated with mode effects, a rating on the severity of risk (low, medium or high), and a description on the potential impact on the data collected.[1]

This report documents how this review was undertaken and describes the outputs from it.

# 2 Development of the risk assessment code frame

The first stage of this work was to develop a code-frame to identify questions that may be at risk. In developing our code-frame we drew on the proposed classification of question characteristics relevant to measurement error (Campanelli *et al.*, 2013), 2013), which the authors proposed are important in the design of questionnaires to be used in mixed mode surveys. However we also drew on elements of the Questionnaire Appraisal System –QAS (Willis and Lessler, 1999) and the wider methodological literature. We identified three overlapping sets of factors that the literature suggests can increase the risk of mode effects.

- Risk of interviewer effects
- Risk of satisficing
- Question and answer presentation issues.

We describe these risk factors further in the remainder of this section.

[1] The spreadsheet can be accessed at https://www.understandingsociety.ac.uk/files/working-papers/2017-04-appendix.xlsx.

## 2.1 Risk of interviewer effects

The content (or topic) of a question can impact on whether or not interviewer effects occur. For example, when answering sensitive questions people may respond differently in interviewer administered (IA) modes compared to self-completion (SC) modes. Socially desirable reporting is more likely to occur in interviewer administered modes. Some authors argue that interviewer presence is related to positivity bias in rating scales (where people are more likely to give 'positive' or 'agree' answers if a scale is interviewer-administered).

Our final code-frame included three types of risk based on question content. The codes and coder instructions are shown in Table 2-1.

**Table 2-1 Risk factors for interviewer effects**

| | **Rationale** | **Coder instructions** |
|---|---|---|
| ***1. Fear of disclosure*** | Embarrassing, illicit or illegal behaviours are more likely to be reported in self-completion modes compared to interviewer administered modes (e.g. Tourangeau, Rips and Rasinski, 2000). | Could there be negative consequences for the participant if the information was made public? This could include illegal behaviours, illicit behaviours or other information the participant would prefer to remain private. |
| ***2. Socially desirable reporting*** | It may be that a question is not about an illegal/ illicit behaviour but it could still have a socially desirable response (e.g. drinking, voting, housework questions).<br><br>Again socially desirable reporting is more prevalent in interviewer administered modes compared to self-completion modes (e.g. Tourangeau, Rips and Rasinski, 2000). | Could participants edit their answers to 'look good' in front of an interviewer?<br><br>Please try and think about how different groups of people could react to these questions. If any groups may edit their answers code '1. Yes.'<br><br>If people may adjust their answer based on interviewer characteristics (e.g. sex, ethnicity etc.) code '1.Yes' |
| ***3.Positivity bias /***<br><br>***Rating scales*** | Face-to-face responses are more likely to have extreme 'positive' ratings compared to the same questions asked in a self-completion mode (e.g. Ye, Fulton and Tourangeau, 2011). | Is the question asking the participant to rate something on a scale e.g. their satisfaction with something or their levels of agreement with something?<br><br>Include agree/disagree scales, fully labelled verbal scales (e.g. excellent-poor) or end point labelled scales (e.g. 0-10). Include three point scales |

## 2.2 Risk of satisficing

It is generally assumed that the risk of respondent satisficing is greater for difficult questions than it is for easy questions (Krosnick, 1991). In interviewer administered questionnaires, interviewers can:

- explain complex tasks;
- provide additional information if required; and
- motivate participants to perform more complex tasks in a way that maximises data quality (e.g. encourage consultation of documents).

Therefore, interviewer presence could decrease satisficing behaviours by decreasing question difficulty, increasing respondent motivation and increasing optimal answering strategies. Theoretically this means that inherently difficult questions could be answered in optimal ways in interviewer-administered modes and non-optimal ways in self-completion modes. Our code-frame includes some items adapted from the QAS (Willis and Lessler, 1999) where we think interviewer presence could have a positive impact on participants' answering strategies. Our final code-frame includes five codes on question difficulty, see Table 2-2.

**Table 2-2 Risk of satisficing**

| | **Rationale** | **Coder instructions** |
|---|---|---|
| ***1. Complex question stem*** | Interviewers are trained to read out the entire question, including all instructions. In self-completion modes participants may not read the entire question, especially if it is long-winded or complex. | Does the question stem include lengthy instructions, introductions, or explanations?<br><br>Do not include optional interview read outs or help screens as part of the stem.<br><br>Please note that not all text-fills variations will be read to respondents. |
| ***2. Extra information*** | In self-completion modes participants may be less likely to seek out information displayed on help screens or read text that appears after the question. | Does the questionnaire include additional explanatory notes on top of those included in the question's stem e.g. definitions of key words or other forms of help? |

| | | |
|---|---|---|
| | | Include all optional interviewer read-outs, clarifications and checks.<br><br>Include instructions that appear after the question mark. |
| ***3.Computation*** | In self-completion modes participants may be less motivated to carry out calculations and could adopt 'shot-cutting' strategies. In interviewer administered modes assistance can be provided with calculations. | Does the question involve any mental calculations e.g. adding, subtracting or averaging?<br><br>Only include questions where you suspect that respondents may have trouble adding up an answer or that might be susceptible to 'rounding.' E.g. age of child / number of people you look after are unlikely to require computation.<br><br>Include dates and numeric questions if appropriate but exclude verbal frequency scales (Always-Never). |
| ***4. Document consultation*** | In self-completion modes participants may be less motivated to get documentation to improve the accuracy of their answer. | Does question request that the participant should consult documents to enhance accuracy? |
| ***5. Open questions*** | Less information is given in open questions in self-completions compared to interviewer administered questions. This can lead to differences in how open responses are coded. | Is a completely open textual answer required?<br><br>Exclude open numeric questions.<br><br>Short textual answers (e.g. participant name and address field were also excluded during the checking phase). |

## 2.3 Question and answer presentation

Some types of mode effects are associated with how questions and answer options are presented: visually or aurally. For example, respondents may not be able to recall longer lists of answer options in aural modes (e.g. telephone). The number of scale points can also influence whether or not mode effects occur (e.g. mid-points are more likely to be selected in visually administered modes compared to aurally administered modes).

The main focus of this review was to consider mode effects that might arise in a face-to-face/ web mixed mode survey context. For example, when viewing questions on the web, participants may not be able to see all answer options on a screen without scrolling. This could increase the likelihood of primacy effects occurring in web (for long list questions) compared to a face-to-face interview where the answer options are listed on a card that is provided to the respondent. In addition, some more 'unusual' formats of question (like ranking tasks) may benefit from an interviewer providing assistance with data entry.

Our final code-frame includes five codes on question and answer presentation that could result in mode effects in a CAPI/web survey. These are shown in Table 2-3.

**Table 2-3 Question and answer presentation issues**

| | **Rationale** | **Coder instructions** |
|---|---|---|
| ***1. Ranking tasks*** | Ranking tasks are not recommended for self-completion modes (Campanelli *et al.*, 2013). | Is the question a ranking task (example shown below)?<br><br>Please rank the following 3 items from 1 meaning most important to 3 meaning least important.<br>• Less traffic [2]<br>• Less crime [1]<br>• More / better shops [3] |
| ***2. Battery of scalar questions*** | Non-differentiation (flat-lining) may be more apparent in web modes than face-to-face modes.<br><br>Note: The 'four' items per cut-off is an arbitrary cut-off for the | Is the question part of a battery of questions that all use the same answer scale? Only include as a 'battery' item if **4 or more questions use the same scale in an unbroken sequence**. |

| | | |
|---|---|---|
| | purposes of coding. | Include agree/disagree scales, other verbal scales, 0-10 scales etc. |
| ***3. Number of answer options*** | In self-completion modes participants may be less likely to read all options compared to in a CAPI mode with a Showcard (Kaminska and Foulsham, 2013). | Are five or more answer options shown (or read) to the respondent?<br><br>Include both scales and categorical responses. Do NOT include non-substantive or hidden options (such as Don't Know or Not Applicable). |
| ***4. Scales with mid-points*** | Mid-points are more likely to be selected in visual modes than aural modes (Campanelli *et al.*, 2013).<br><br>On satisfaction and agree/disagree<br><br>questions, there is more selection of mid-points in CAWI than IA (Campanelli *et al.*, 2013). | Does the question have an answer scale with a mid-point? (E.g. is it a scale with 3, 5, 7, 9 or 11 points)?<br><br>Include agree/disagree scales, fully labelled verbal scales (e.g. excellent-poor) or end point labelled scales (e.g. 0-10). |
| ***5. Hidden codes and interviewer coded items*** | There is limited evidence on how best to translate interviewer coded items into questions for participants and what impact this has on the data collected.<br><br>Hidden codes may not be used in the same way in self-completion modes. Hidden modes can either be shown to participants (in which case they might be more likely to be selected more in web) or remain hidden (in which case they may be less likely to be selected in web). | Is the item an interviewer observation (that is not read out to respondents)?<br><br>Are there any spontaneous or 'hidden' answer codes which are not shown to participants?<br><br>Exclude hidden don't know/ no opinion answer options. |

### 2.4 Overlap in conceptual framework

As mentioned in the introduction to section 2, there is a conceptual overlap in how the three sets of factors described above interact to produce mode effects. For example, positivity bias (described in Table 2-1) has been attributed by some authors to acquiescent reporting in the presence of an interviewer (see for example, Ye, Fulton and Tourangeau, 2011). However, other authors ascribe the same effect as resulting from aural versus visual pathways of question processing (for example see Dillman, Smyth and Christian, 2009). In developing our code frame the goal was to ensure we captured all relevant factors. The allocation of a factor code to a particular group heading is subjective: we were not attempting to definitively attribute causal mechanisms.

## 3 Coding questions in the W7-W10 questionnaire

Five researchers were briefed on how to use the code-frame. Each question was reviewed by two coders independently (i.e. coders were not able to check what code had been assigned by the other coder). Coding was binary (e.g. coders could only code 'Yes' or 'No' for each risk factor for each question).

Initially a block of 100 questions was selected to test the code-frame on. The 100 questions chosen varied in terms of whether they were factual or attitudinal, what types of answer category they used and so on. After this coding was complete the inter-coder reliability scores were compared. Reasons for disagreement between coders were explored in a debriefing session and coding instructions were adapted accordingly. After this all questions (N=1,486) were double coded using the adapted code-frame and coder instructions. In the case of further coder disagreement a 'final code' was allocated by a third coder. Table 3-1 presents the final inter-coder agreement levels, the number and percentage of questions that were coded 'yes' and the predicted impact on data for each risk factor.

**Table 3-1 Inter-coder reliability and implications for interpretation of risk**

| Code | Coder agreement (%) | Final code | | | Predicted impact on data |
|---|---|---|---|---|---|
| | | Column | Yes (N) | Yes (%) | |
| **Risk factors for interviewer effects** | | | | | |
| Fear of disclosure | 95.5 | CY | 33 | 2.3 | Illicit or 'sensitive' activities less likely to be reported in IA modes |
| Socially desirable reporting | 59.6 | DC | 364 | 25.7 | Socially desirable answers over-reported in IA mode. |
| Positivity bias / Rating scales | 94.1 | DG | 149 | 10.5 | Greater reporting of positive answers in IA mode. |
| **Risk factors for satisficing** | | | | | |
| Complex question stem | 90.6 | DK | 69 | 4.9 | Full question less likely to be read in SC mode. Impact on data quality unclear. |
| Extra information | 88.9 | DO | 391 | 27.6 | Extra information: Interviewer information not viewed in SC mode. Impact on data quality unclear. |
| Computation | 91.0 | DS | 110 | 7.8 | SC participants may be less motivated to carry out calculations and more likely to use 'short-cutting' strategies. Impact on data quality unclear. |

| | | | | | |
|---|---|---|---|---|---|
| Document consultation | 99.9 | DW | 1[2] | 0.1 | SC participants may be less motivated to check documentation. Impact on data quality unclear. |
| Open questions | 98.9 | EA | 46 | 3.2 | Less information given in SC compared to IA. Impacts on ease and accuracy of coding. |
| **Risk factors for question and answer presentation effects** | | | | | |
| Ranking tasks | 99.9 | EI | 2 | 0.1 | Not recommended for SC. Item non-response may be higher in SC. Impact on data quality unclear |
| Battery of scalar questions | 97.8 | EM | 364 | 25.7 | Non-differentiation (flat-lining) may be more apparent in SC than IA. |
| Number of answer options | 93.7 | EQ | 377 | 26.6 | In SC participants may be less likely to read all options compared to IA. Increased risk of primacy effects in SC. |
| Scales with mid-points | 92.9 | EU | 246 | 17.3 | Mid-points more likely to be used in SC as a form of flat-lining. |
| Hidden codes and interviewer coded | 91.7 | EY | 93 | 6.6 | Interviewer fields need reframing for SC. Hidden codes will be used less (or not used at all) in SC. |

[2] Note: Only one variable in the questionnaire (an interviewer observation) made an explicit reference to documentation checking. However, in practice documentation consultation occurs at many items. Further coding of this requires that items that benefit from documentation be marked in the questionnaire specification.

# 4 Assigning risk ratings

After the coding was complete risk ratings (low, medium and high) were produced for each question. In our view certain codes represented more serious mode effect risks than others. Some codes had clearer implications for directional biases in the data collected. Codes on 'interviewer effects' made predictions about specific answers being differentially reported in different modes whereas for other codes the implications on data outputs were less clear. We therefore gave codes that predicted a directional or predictable bias greater weight.

The inter-coder reliability for risk of 'socially desirable reporting' was low compared to other codes. For this reason a greater weight was attached to social desirability bias in cases where coders were in agreement.

Finally, questions with multiple risk factors were assigned a higher risk rating than questions with single risk factors. The relative weights given to each risk factor and how final scores were generated are shown in Table 4-1 below.

**Table 4-1 Weighting of codes into final risk scores**

<table>
<tr><td colspan="2">Risk factors for interviewer effects</td><td colspan="2" rowspan="2">Risk scoring for interviewer effects<br>(Column FT)</td></tr>
<tr><td>Code</td><td>Code score</td></tr>
<tr><td>Fear of disclosure</td><td>3</td><td>Total Score</td><td>Risk rating</td></tr>
<tr><td rowspan="3">Socially desirable reporting</td><td rowspan="3">If both coders agree= 2<br><br>If coder disagreement and final code is yes=1</td><td>0</td><td>No risk</td></tr>
<tr><td>1-2</td><td>Low risk</td></tr>
<tr><td>3-4</td><td>Medium risk</td></tr>
<tr><td>Positivity bias / Rating scales</td><td>2</td><td>5+</td><td>High risk</td></tr>
<tr><td colspan="2">Risk factors for satisficing</td><td colspan="2">Risk scoring for satisficing<br>(Column FV)</td></tr>
</table>

| | | | |
|---|---|---|---|
| Complex question | 1 | **Total Score** | **Risk rating** |
| Extra information | 1 | **0** | **No risk** |
| Computation | 1 | **1-2** | **Low risk** |
| Document consultation | 1 | **3-4** | **Medium risk** |
| Open questions | 3 | **5+** | **High risk** |
| **Risk factors for question and answer presentation effects** | | **Risk scoring for question and answer presentation effects** (Column FX) | |
| Ranking tasks | 1 | **Total Score** | **Risk rating** |
| Battery of scalar questions | 1 | **0** | **No risk** |
| Number of answer options | 2 | **1-2** | **Low risk** |
| Scales with mid-points | 1 | **3-4** | **Medium risk** |
| Hidden codes and interviewer coded | 2 | **5+** | **High risk** |

A summary risk rating (on risk of any mode effect) was then generated for each question, using the rules set out in Table 4-2.

**Table 4-2 Final Risk Rating**

| **Final risk rating of any mode effect**<br><br>(Column FZ) | |
|---|---|
| **If ALL risk ratings = no risk** | **No risk** |
| **If ANY risk ratings = low risk but no Medium nor High risk** | **Low risk** |
| **If ANY risk ratings = Medium risk but no High risk** | **Medium risk** |
| **If ANY high risk ratings** | **High risk** |

These rules were used to generate summary the summary ratings for risk of interviewer effects, satisficing and question and answer presentation effects, as well as an overall final risk rating. Table 4-3 shows the total number of questions rated as having no, low, medium or high risk following coding for each type of risk factor and overall. Questions judged to have a high overall risk of mode effects were those that were rated as having a high risk of interviewer effects and or high risk of question and answer presentation effects.

**Table 4-3 Total number of questions rated as being at risk of mode effects**

| | **Risk of interviewer effects** | **Risk factors for satisficing** | **Risk factors for question and answer presentation effects** | **Final risk rating of any mode effect** |
|---|---|---|---|---|
| **No risk** | **1044** | **980** | **799** | **405** |
| **Low risk** | **336** | **450** | **496** | **758** |
| **Medium risk** | **88** | **56** | **184** | **299** |
| **High risk** | **18** | **0** | **6** | **24** |

# 5 References

Campanelli, P., Nicolaas, G., Jäckle, A., Lynn, P., Hope, S., Blake, M. and Gray, M. (2013) 'A Classification of Question Characteristics Relevant to Measurement Error and Consequently Important for Mixed Mode Questionnaire Design'. First presented at the *Royal Statistical Society*. London, 11 October 2011.

Dillman, D. A., Smyth, J. . and Christian, L. . (2009) *Internet, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ: Wiley.

Kaminska, O. and Foulsham, T. (2013) *Real-world eye-tracking in face-to-face, web and SAQ modes*. ISER Working Paper Series. 2013–7.

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

Tourangeau, R., Rips, L. . and Rasinski, K. (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Willis, G. and Lessler, J. (1999) 'Question appraisal system QAS-99', *Rockville, MD: Research* …. Available at: http://appliedresearch.cancer.gov/areas/cognitive/qas99.pdf (Accessed: 30 September 2016).

Ye, C., Fulton, J. and Tourangeau, R. (2011) 'More Positive or More Extreme? A Meta-Analysis of Mode Differences in Response Choice', *Public Opinion Quarterly*, 75(2), pp. 349–365.