The interviewer contribution to variability in response times in face-to-face interview surveys

Abstract

Survey researchers have consistently found that interviewers make a small but systematic contribution to variability in response times. However, we know little about what the characteristics of interviewers are that lead to this effect. In this paper we address this gap in understanding by linking item level response times from wave three of the UK Household Longitudinal Survey (UKHLS) to data from an independently conducted survey of interviewers. The linked data file has a complex, hierarchical structure with response latencies nested within respondents and questions, which are themselves nested within interviewers and areas. We propose the use of a cross-classified mixed-effects location scale model to allow for the decomposition of the joint effects on response times of interviewers, areas, questions, and respondents. We present a new method for producing interviewer specific intra-class correlations (ICCs) of response times and evaluate how interviewer demographic characteristics, personality, and attitudes to surveys and to interviewing affect the length of response latencies and the interviewer-specific ICCs. Hence, the paper makes both methodological and substantive contributions to the investigation of response times.

INTRODUCTION

In recent years survey researchers have become increasingly interested in the length of time it takes respondents to answer questionnaire items, so-called ‘response latencies’. This is, in part, because item-level response time data is now relatively straightforward to capture via automatically generated time-stamps in computer assisted and on-line self-completion questionnaire software (Couper and Kreuter, 2013). While early investigations of response latencies required costly bespoke measurements at the item level via ‘active timers’ controlled by interviewers, it is now straightforward to obtain very precise timings between adjacent key-strokes or mouse clicks using paradata (Yan and Tourangeau, 2007; Couper 1998). In addition to the greater ease and reduced cost of data collection, the focus on response latencies has arisen because latencies have some potentially attractive properties for survey researchers seeking to improve data quality. For example, unusually long response times may be diagnostic of poorly constructed questions that respondents find difficult to answer (Bassili and Fletcher, 1991; Bassili and Scott, 1996; Bassili and Krosnick 2000). Particularly short latencies, on the other hand, may enable detection of respondents who speed through the questionnaire providing poor-quality and error-laden answers (Zhang and Conrad, 2015; Yan et al 2014; Krosnick, 1991). Better understanding of the factors influencing interview length also holds out the possibility of cost efficiencies in interviewer administered surveys, if interview length can be reduced without damaging data quality (Olson and Pankhurst, 2013; Turner et al, 2015).

Researchers have investigated how characteristics of questions, respondents, and interviewers are related to the time it takes to complete a question. With regard to questions, Yan and Tourangeau (2008) found response times in a web survey were longer as question complexity – measured by the number of words and clauses – increased, a finding echoed by Olson and Smyth (2015) in an analysis of a telephone survey. Unsurprisingly, question characteristics that require additional input from interviewers also take longer than questions without such features. Thus, studies have found questions with more response alternatives, that use show-cards, that require open text answers, and which have help screens to be associated with longer response times (Couper and Kreuter, 2013; Olson and Smyth, 2015). Anomalously in this regard, Couper and Kreuter found the presence of interviewer instructions to be associated with *shorter* response times, while Olson and Smyth found no significant association between questions with interviewer instructions and response times. The position of an item in the questionnaire also matters, with latencies decreasing towards the end of a survey (Couper and Kreuter, 2013; Loosveldt and Beullens, 2013). This is presumably because respondents learn the 'rules of the game' in the early parts of the questionnaire and so are able to navigate through it more quickly towards the end. This effect may be exacerbated by fatigue and boredom for both respondent and interviewer as the interview reaches its latter stages, resulting in shorter and quicker answers. Thematic continuity has also been found to be important, with questions that cover the same topic as the previous one eliciting shorter latencies than ones that switch to a different theme (Tourangeau et al, 1991) and item batteries which mix the direction of attitude questions taking longer than a fixed order alternative (Mayerl and Gihel, 2018). Regarding question domain, demographic questions have been found to have shorter response times compared to attitudinal questions, which in turn take less time to complete than behavioural questions (Yan and Tourangeau 2008; Olson and Smyth, 2015; Draisma and Dijkstra 2004). Questions that invoke norms of social desirability have also been shown to produce shorter response times (Andersen & Mayerl, 2017). Collectively, then, the existing evidence supports the uncontroversial idea that, if a question is in some way difficult or complex, it takes more time to answer because greater thought and attention is required to determine a response.

The flipside of this question complexity effect is observed when considering respondent characteristics; indicators of lower cognitive ability are associated with longer response times (Faust et al 1999). In particular, older respondents have consistently been shown to take longer to answer (Yan and Tourangeau 2008; Couper and Kreuter 2013; Loosveldt and Beullens, 2013; Zhang and Conrad, 2014; Olson and Smyth, 2015; Gummer and Rosmann, 2015; Vandenplas et al, 2017), as have respondents with lower levels of education (Couper and Kreuter, 2013; Gummer and Rosmann, 2015). Olson and Smyth, however, found no association between education and response times, while Loosveldt and Beullens (2013) and Vandenplas et al (2017) found a mix of positive, negative, and non-significant associations between education and response times across countries in the European Social Survey.

Interviewers have also been found to influence response times. For example, Olson and Peytchev (2007) found that average interview length decreased over the course of the fieldwork period as interviewers conducted more interviews and became more familiar with the survey. Subsequent studies have established that interviewers exert a small but systematic influence on the duration of question modules (Loosveldt and Beullens, 2013) and on item level response latencies (Couper and Kreuter, 2013; Olson and Smyth, 2015). When it comes to explaining this effect, more experienced interviewers have been found to have, on average, shorter response times compared to their less experienced counterparts (Olson and Peytchev, 2007; Couper and Kreuter, 2013; Kirchner and Olson, 2017). However, Olson and Smyth (2015) found no difference between interviewers with more than a year's experience compared to those with less experience. There is also evidence that older interviewers tend to produce somewhat longer latencies, although this effect was only apparent for male respondents in Couper and Kreuter's analysis (2013).

Thus, although it is now well established that interviewers contribute to variability in response times, we know little about what causes this effect and, where relationships between interviewer characteristics and response times have been found, the direction and magnitude of effects have not been consistent across studies. Part of the reason for this inconsistency is likely to be the rather different designs of the surveys and diversity in the measures used. However, it also relates to the paucity of interviewer characteristics that it has been possible for previous researchers to include in their models. In particular, no study to date has considered more psychological variables, such as attitudes and personality, which scholars have found to be predict a range of survey errors for both interviewers and respondents (West and Blom, 2017; Cheng et al, 2018; Lugtig and Jackel, 2014; Turner et al, 2015; Campanelli et al, 1997; de Leeuw, 1999; Durrant et al, 2010).

It is to this gap in the existing evidence that this paper seeks to contribute. Using data from a large, nationally representative survey in the UK, we use cross-classified mixed-effects location scale models (Brunton-Smith et al, 2017) to decompose variability in response times into question, respondent, interviewer, and area components. We then use a range of measured variables to model the variance at each level. For interviewers, we are able to include measures attitudes toward surveys, to persuading respondents to participate and job satisfaction, as well as a reduced form of the 'Big Five' personality inventory.

The remainder of the paper proceeds as follows. In the next section we set out the structure and sources of our combined dataset, describe how the various measures used in our models were derived, and specify our expectations regarding the likely directions of effects. We then provide a formal account of our analytical strategy and statistical assumptions before setting out the key findings from our analysis. We conclude with a consideration of the limitations of our approach and the implications of our findings for understanding response times in interviewer administered surveys and for survey practice more generally.

DATA AND MEASURES

The data for this study are drawn from three independent sources: response latency and individual questionnaire data come from wave three of the UK Household Longitudinal Survey (UKHLS) and are linked to an independently conducted survey of interviewers who worked on the UKHLS at wave three, and to area characteristics derived from the 2011 census. Data are linked using deterministic matching on unique identifiers for respondents, interviewers, and areas. The UKHLS is a multi-purpose household panel survey covering topics of health, work, education, income, family and social life to help understand the long-term effects of social and economic change (Scott and Jessop 2013). The survey has a stratified, multistage design with a sample of postcode sectors selected with probability proportional to their population size and 18 households selected from each sector. All household members aged 16 and over are invited to provide an individual interview. Data collection for wave three of UKHLS took place between January 2011 and July 2013 with interviews carried out face-to-face in respondents' homes using computer-assisted personal interviewing (CAPI). Hard-to-reach respondents were interviewed over the phone at the end of fieldwork in order to boost response rates. The 2% of respondents who were interviewed by telephone are excluded from our analysis sample because it was not possible to obtain comparable measures of response latency for them. A total of 30,685 individual interviews were conducted representing a response rate of 61% (Knies 2018).

Response Latencies

Response latencies are measured at the item level, using the timings produced by a latent timer embedded within Blaise, the questionnaire software used for UKHLS, which automatically generates time stamps between key strokes in the CAPI audit trail. The response latency for a question is the time between an interviewer keying the respondent's answers to two adjacent questions. This means that the latency includes the time taken for the interviewer to read and for the respondent to answer the question. For this reason, latent timers produce longer latencies than 'active timers' which are controlled manually to include only the respondent's answer, although in practice active and passive timers are very highly correlated (Mulligan et al, 2003).

The length of response latencies in the UKHLS ranged between 0 and 179 seconds. 1.3% of the total sample (39,373 latencies) were zero seconds and 17% were very short latencies of 1-3 seconds. These are implausibly short times to ask and answer a survey question. An investigation into the causes of these very short latencies was undertaken by listening to audio recordings of a sample of such questions. This revealed that the zero second latencies were caused by interviewers scrolling backward in the questionnaire to an earlier question and then scrolling forward again, over-writing the original latencies with zeros. It also showed that the latencies of 1-3 seconds duration were mainly due to interviewers not reading the questions, although it was not clear from the recordings why the questions had been skipped. Because our focus in this paper is on interviewer influences on response latencies, the decision was taken to remove the zero latencies but to keep the latencies of 1-3 seconds.

Some latencies were also very long, likely arising due to extraneous reasons such as the interviewer having technical problems with the laptop, or a phone call interrupting the interview (Turner et al. 2015). We therefore dropped all very long latencies, which we defined as those of more than 100 seconds, amounting to 0.3% of all latencies. The 100 second threshold is a judgment taken by the authors about the maximum length of time it might reasonably take a respondent to answer a question in this survey and, being subjective, this threshold is somewhat arbitrary. However, our results are qualitatively unaffected by setting the threshold at lower and higher levels. Respondents answered an average of 223 questions (minimum=42; maximum=433) resulting in a total of 3,100,288 response latencies. Figure 1 shows that the distribution of response latencies is positively skewed with a mean length of 10.8 seconds and a median of 7 seconds. We therefore log-transformed the raw latencies using the natural logarithm and use linear models in our analysis.

FIGURE 1 HERE

Interviewers

Detailed information on interviewers comes from a survey of interviewers working for the data collection agency, NatCen, that was carried out in 2014 using online and paper self-completion (Burton et al. 2014). Invitations to complete the survey were sent to all 823 NatCen interviewers who worked on wave one of Understanding Society, with 473 completing the questionnaire, a response rate of 58%. We were able to link data from the interviewer survey for 362 of the 668 interviewers who worked on Wave three of UKHLS. Of the interviewers for whom no data linkage was achieved, 156 were non-respondents to the interviewer survey and 150 joined the NatCen interviewer panel after wave one of UKHLS, so were not invited to take part in the interviewer survey. We assessed non-random dropout from the interview panel between the interviewer survey and wave three of the UKHLS by regressing a binary indicator of dropout on the interviewer characteristics measured in the interviewer survey. Only interviewing experience (less experienced interviewers more likely to drop out) and job satisfaction (less satisfied interviewers more likely to drop out) were significant at the 95% level of confidence. Moreover, the distributions of the response latencies were virtually identical between the full wave 3 UKHLS and our analysis sample, so we do not consider that non-random drop out of interviewers from the study represents a threat to the generality of our findings.

From the interviewer questionnaire data we use sex, age, and interviewing experience (measured in number of years working for NatCen) as predictors. We also use measures of interviewer job satisfaction, beliefs about the value and importance of surveys, and attitude toward persuading reluctant respondents which have been used in previous interviewer studies (Lehtonen, 1996; Campanelli et al, 1997). A combined measure for each attitude was derived by taking the first principal component of the items in each scale. Full wordings and response alternatives for the items are included in the Appendix. Our expectation is that lower levels of job satisfaction will be associated with shorter response latencies because interviewers who are less stimulated and rewarded by their work may endeavour to finish interviews more rapidly than those who find interviewing more intrinsically rewarding. A more positive attitude toward the role and importance of surveys in society we also expect to be associated with longer latencies, all things equal, as interviewers who believe that survey evidence is societally consequential may place higher value on obtaining complete and accurate data and avoid speeding through the questionnaire. Previous research has found that interviewers who have more positive attitudes toward persuading reluctant respondents achieve, on average, higher rates of cooperation (Durrant et al, 2010; Jäckle et al, 2013) and that interviewers who achieve higher response rates exert less influence on the variability of responses they produce between respondents (Brunton-Smith et al, 2012). We therefore expect more positive attitudes on this dimension to be associated with longer response latencies.

We include measures of interviewer personality using a version of the Big Five personality inventory, which has become the dominant conceptual and empirical framework for personality measurement (Costa and McCrae, 1992). The Big Five comprises five over-arching dimensions which cover the primary ways in which human traits and behaviours differ systematically between individuals (John and Srivastava, 1999). These dimensions are openness, conscientiousness, extraversion, agreeableness, and neuroticism. People who score highly on openness are keen to try new experiences and are typically artistic, creative, and oriented toward intellectual and creative pursuits. Conscientious people are hard-working, responsible and methodical, with an eye to detail and an achievement orientation. Extraversion relates to an individual's gregariousness, sociability, and enjoyment of exposure to socially stimulating environments. The agreeableness dimension concerns an individual's level of interpersonal warmth and ability to get on with others, people who score high on agreeableness tend to be trustworthy and dependable. Neuroticism is a trait which relates to one's level of emotional stability, people who score high on neuroticism are more likely to be anxious and to experience volatility in their mood. We use a shortened, 15 item version of the full Big Five inventory which was designed for survey administration and which has been used previously on both general population (Benet-Martinez and John, 1998) and interviewer (Jackel et al, 2013) samples. We anticipate that less conscientious and less agreeable interviewers will have shorter latencies. For openness, neuroticism, and extraversion we have no strong prior expectations. Nonetheless, we include all five dimensions in our model in an exploratory rather than a theoretically deductive spirit. Finally, we include a single question which asked interviewers to assess the appropriateness of the length of the wave one UKHLS questionnaire. We anticipate that interviewers who rate the interview as having been too long at wave one will have shorter response latencies at wave three as they may try to mitigate the burden of the questionnaire by speeding up the interview.

Areas

Interviewers working on UKHLS are not randomly assigned to areas so it is important to control for compositional differences between interviewer assignments (Campanelli and O'Muircheartaigh, 1999). To do this we link the interviewer and questionnaire data to the middle layer super-output area (MSOA) geography. MSOAs are similar to US census tracts, comprising an average of 4,000 households grouped together based on similarity of housing tenure, with an average size of 0.6 square miles. We link variables from the 2001 UK census to MSOAs to control for characteristics of local areas. A total of 21 raw census count variables were combined using a factorial ecology model (Rees, 1971), with four neighbourhood indices extracted. These measures cover the extent of *economic disadvantage* (areas with a higher number of single parent families, those on income support and unemployed, fewer people in managerial and professional occupations, and less owner occupiers), *urbanicity* (high population density and domestic properties, and relatively little green space) and *population mobility* (higher levels of in- and out-migration and more single person households). We also account for differences in the *age structure* of the area (with higher scores for areas with a younger population). Each interviewer worked in an average of 13.6 MSOAs (min= 1, max= 46) and each area had an average of 1.2 interviewers working in it (min= 1, max= 4). MSOAs are only available for England and Wales, so respondents in Northern Ireland and Scotland are excluded from the analysis, reducing the sample to 16,401 respondents and 290 interviewers.

Respondent characteristics

We include respondent sex, age, and highest qualification from the individual questionnaire as these have been found to be related to response times in existing studies (Yan and Tourangeau, 2008; Couper and Kreuter, 2013). We also include 'response style' indicators which have been used previously as measures of respondents' tendencies to take cognitive 'short cuts' when completing questionnaires (Krosnick, 1991; Roberts et al, 2019). These are counts of: 'Don't know' and mid-point responses, and non-differentiation (or straight-lining) (Zhang and Conrad, 2014). To measure straight-lining we take the average deviation between the current answer compared to the answer for the preceding question (Loosveldt et al, 2018) in a battery of items using the same scale. We then derive a binary indicator by giving respondents in the top decile a value of 1 and the remaining respondents a value of 0. Because respondents are believed to employ these kinds of 'satisficing' response styles in order to minimize the cognitive costs of survey completion (Turner et al, 2015; Zhang and Conrad, 2015), our expectation is that these indicators should be associated with shorter response times. 2,411 respondents did not answer the questions which were used to produce the response style indicators: 1,391 did not complete the self-completion modules and 672 were ineligible to complete the political attitude questions because they were under the age of 18 and were excluded from the analysis. This reduced the sample to 13,990 respondents nested within 288 interviewers within 3,187 areas.

Question characteristics

A team of three coders categorised the characteristics of all 2,444 questions in the UKHLS individual questionnaire. Following existing practice (Olson and Smyth 2015; Couper and Kreuter 2013; Yan and Tourangeau 2008), questions were coded according to: number of words; position in the questionnaire; presence of interviewer instructions, definitions, and help; transition statements; visual emphasis (e.g. bold text); position in a battery of items; use of show cards; number of response options; and whether an open or a closed response was required. Our expectation, in line with the findings in the extant literature, is that question features which require more input from the interviewer, such as a larger number of words, show-cards, transition statements, and so on, will produce longer latencies. Additionally, questions were coded according to whether they addressed a sensitive issue and whether they were demographic, attitude, perception, behavioural, or test questions (the questionnaire contained numeric, verbal, and memory ability tests). Our expectation here is that latencies will be shortest for demographic questions, longest for test items, with attitudes, perceptions, and behavioural questions falling in between these two extremes. This accords with the ordering that would be expected given the salience and accessibility of each underlying cognitive entity (Yan and Tourangeau, 2008). For example, a typical respondent does not need to take much time to report their age in years, their employment status, housing tenure, and so on because these are personally salient and readily accessible in memory. Attitudes and perceptions will, on average, take somewhat longer to formulate a response to as they are likely to address, at least in part, issues with which the respondent is not familiar and which may be of low salience to them personally. Cognitive test items will take longest because they require some degree of introspection, calculation, or recall and are likely to be cognitively demanding for many respondents in a general population sample. A random sample of 10 (234) % of questions were coded by all three coders to assess the level of agreement between them. The overall level of agreement across the 17 codes was 92% with a range from 72% (type of response scale) to 100% (presence of show-card). Table A1 in the Appendix presents proportions of agreement for each code.

ANALYSIS

The data has a complex structure, with response latencies nested within the two-way cross-classification formed by respondents and questions which is further nested within the two-way-cross-classification formed by areas and interviewers. We therefore use cross-classified mixed-effects models (Goldstein, 2010; Raudenbush and Bryk, 2002) of the following form:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + g_m + f_l + v_k + u_j + e_i, \qquad (1)$$

where y_i denotes the log of response latency i $(i = 1, \ldots, N)$ and $\mathbf{x}_i$ is a vector of area, interviewer, respondent, and item level covariates with coefficients $\boldsymbol{\beta}$. The area, interviewer, respondent, and item random intercepts are, respectively, denoted g_m, f_l, v_k, and u_j and e_i

is the response latency specific residual. The random effects are assumed mutually independent, independent of the covariates, and normally distributed with zero means and constant variances: $g_m \sim N(0, \sigma_g^2)$, $f_l \sim N(0, \sigma_f^2)$, $v_k \sim N(0, \sigma_v^2)$, $u_j \sim N(0, \sigma_u^2)$ and $e_i \sim N(0, \sigma_e^2)$. The random effect variances σ_g^2, σ_f^2, σ_v^2, and σ_u^2 capture the variability in adjusted response latencies across areas, interviewers, respondents, and items respectively, and the residual variance σ_e^2 denotes the variability in response latencies that is unexplained by the fixed and random effects in the model. In addition to main effects, we also included interactions between characteristics of interviewers and respondents (e.g. age of interviewer and age of respondent), between characteristics of respondents and the response style indicators (e.g. age of interviewer and straight-lining), and between characteristics of interviewers and types of questions (e.g. age of interviewer and number of words in the question). This enables us to assess whether the effects of characteristics at one level are moderated by those at another level, for instance, do older respondents provide longer response times when interviewed by an older interviewer, or do questions with more words take less time to administer when the interviewer is younger? Models are fitted using R 3.3.1 and the LME4 library (Bates et al., 2015).

The Intra Class Correlation (ICC) for interviewers is derived as $\rho_f = \sigma_f^2 \left(\sigma_g^2 + \sigma_f^2 + \sigma_v^2 + \sigma_u^2 + \sigma_e^2\right)^{-1}$ and can be interpreted as the expected correlation between the response latencies for the same question from a respondent interviewed by two randomly selected interviewers. Interviewer specific ICCs are generated by extending equation (1) to include an auxiliary log-linear equation for the residual variance. The logistic transformation ensures the residual variance does not take negative values (Brunton-Smith et al, 2017), the equation is written as:

$$\ln\left(\sigma_{e_i}^2\right) = \boldsymbol{\alpha} + f_l^{[2]}, \tag{2}$$

where $\ln\left(\sigma_{e_i}^2\right)$ is the log of the now heterogeneous residual variance, $\boldsymbol{\alpha}$ is the intercept and $f_l^{[2]}$ is an additional interviewer random effect. The '$[2]$' superscript distinguishes this random effect from the interviewer random effect in equation 1 which is now denoted $f_l^{[1]}$. The two sets of interviewer random effects are assumed bivariate normal with zero mean and constant variance-covariance. The ICC for interviewers is derived as $\rho_u = \sigma_u^2\left(\sigma_g^2 + \sigma_f^2 + \sigma_u^2 + \sigma_v^2 + \sigma_e^2\right)^{-1}$ and gives the proportion of the total variance in response latencies that can be attributed to interviewers. It is then straightforward to calculate interviewer-specific ICCs which can be used to identify interviewers who induce more similar responses from the respondents they interview compared to other interviewers:

$$\frac{\sigma_{f[1]}^2}{\sigma_g^2+\sigma_{f_l^{[1]}}^2+\sigma_v^2+\sigma_u^2+\exp\left(\boldsymbol{\alpha}+f_l^{[2]}\right)} \tag{3}$$

Note that for the model used to produce the interviewer-specific ICCs in this paper, we do not include any covariates.

RESULTS

We first fit the null model to obtain the variance components and ICCs across the levels of the hierarchical data structure. These are presented in Table 1 alongside the estimates of these components from Couper and Kreuter (2013) and Olson and Smyth (2015). The estimates are not fully comparable across studies because neither Couper and Kreuter (2013) nor Olson and Smyth (2015) include an area level random effect. Additionally, Couper and Kreuter specify questions as the lowest level in the model, while we treat latencies as a separate level. We prefer this specification because it allows variability deriving from question features that is not captured by the fixed effects to be differentiated from random error. Nonetheless, the models are sufficiently similar to enable approximate comparisons of the variability situated at each level.

Our results show that UKHLS interviewers account for just 3% of the total variability in response latencies, compared to 5% for respondents and 45% for questions. Areas contribute only half a percentage point, which suggests that the omission of area controls is unlikely to have materially affected the results of previous studies. The UKHLS estimates are, then, very similar to those observed in comparable US-based surveys, with approximately half located at the question level, approximately 4-7% at the respondent level and a smaller proportion, around 3% or less, at the interviewer level.

TABLE 1 HERE

While the ICCs for the UKHLS are similar to those obtained for these US surveys, they are substantially lower than those reported by Loosveldt and Bullens (2013) on the fifth (2010/11) round of the European Social Survey (ESS). These authors report interviewer ICCs for twelve countries with an average of 11%, ranging from a low of 3% for Portugal to a high of 25% for the Netherlands (the ICC for Great Britain was 11%). The discrepancy between the estimates in Table 1 and those of Loosveldt and Beullens (2013) is likely due to the fact that response times in the ESS were recorded at the question module rather than the item level, so the ESS models do not include a level for questions. This is likely to inflate the interviewer variance as a result of uneven distribution of item types across interviewers, although we cannot rule out the possibility that interviewers do indeed make a more substantial contribution to response times in some ESS countries.

Figure 2 shows the interviewer specific ICCs with 95% confidence intervals for the UKHLS data. Interviewers are indicated by black diamonds and the red horizontal line indicates the population average ICC for interviewers. There is clear and substantial heterogeneity between interviewers in the extent to which they affect variability in response times across questions and respondents, with some interviewers contributing over 3.5% and others a little over 2%.

FIGURE 2 HERE

To explain the variability at each level, we introduce fixed effects for question, area, respondent, and interviewer characteristics. Estimates are presented in Table 2. Considering first the item-level coefficients, we see that characteristics of questions which would be expected to increase the amount of time required for the interviewer to read out the question have, unsurprisingly, longer latencies. These are questions which: contain more words; require a show-card; have a visual emphasis; provide instructions, or help; and which have more response options, or require an open answer. Questions which are the first in a battery of items using the same response alternatives also take longer than subsequent items. This is likely a result of the interviewer reading the full set of response alternatives for the first item but skipping it for some or all of subsequent ones. In terms of thematic focus, demographic questions have the shortest response latencies, followed by behavioural questions, then attitudes and perceptions, with test items taking the longest time to administer. This is consistent with our expectations and with the findings of previous investigations (Yan and Tourangeau, 2008; Olson and Smyth, 2015). Four question characteristics were unrelated to length of the response time: sensitive questions; questions containing a transition statement; whether the question was self or interviewer administered; and position in the questionnaire. Overall, these question level results are similar to those reported in previous studies though with some exceptions. For instance, Couper and Kreuter (2013) found *shorter* latencies for questions with interviewer instructions, questions which come earlier in the questionnaire, and self-administered questions. Olson and Smyth found sensitive questions to have shorter latencies but questions with visual emphasis, position in a battery, and questions with interviewer instructions to be unrelated to latency length.

Turning to the respondent level coefficients, we see that women, younger people, and people with higher levels of educational attainment have, on average, shorter item-level response latencies. This same pattern was found by Couper and Kreuter (2013) while Olson and Smyth (2015) found a positive association with age but gender and educational attainment to be unrelated to response latencies. The response-style indicators show that non-differentiation is associated with shorter latencies, so respondents employing this 'satisificing' strategy do indeed appear to reduce the cognitive costs of completing the survey (Zhang and Conrad, 2014; Krosnick, 1991). However, a tendency to say 'Don't Know' and to select the mid-point on attitude questions are unrelated to response times.

Of the area-level characteristics, living in more economically deprived and urban areas is associated with longer response times. These associations are likely to be picking up residual compositional differences in the kinds of people living in different areas and should, therefore, tend toward zero with the inclusion of a more comprehensive set of individual level covariates.

With regard to interviewers, only three of the measured characteristics are significantly associated with response times. Consistent with existing research, we find that interviewer experience is negatively related to length of the latency (Couper and Kreuter, 2013; Loosveldt and Beullens, 2013; Olson and Peytchev, 2007). We also find that younger interviewers and interviewers who rated the first wave of UKHLS as being too long had shorter question level response times at wave three. This accords with our expectation that interviewers who believe the questionnaire is too long may try to mitigate negative effects on respondents by speeding up the interview. Counter to our expectations, interviewer sex, personality, job satisfaction, attitude to the role and value of surveys, and attitude to persuading reluctant respondents are all untreated to the length of response times. We also tested for interactions between interviewer and respondent and interviewer and question characteristics. While, several interaction terms were significant, all of the moderating relationships were small in magnitude and lacked clear substantive interpretations, so we have not presented them here.

As a final step in the analysis we fitted an Ordinary Least Squares regression using interviewer characteristics to predict between-interviewer variance in the interviewer-specific ICCs. The results, presented in Table 3 show that, as might be expected, more experienced interviewers have lower ICCs. However, this and the other coefficients in the model do not reach significance at the 95% level of confidence and the predictive power of the model is relatively weak, with an R-squared of under 10%. It would seem then that it is necessary to look beyond these sorts of characteristics to understand why interviewers vary in their individual contributions to response times.

TABLE 3 HERE

DISCUSSION

Academic interest in the time it takes people to answer questions stretches back as far as the 19th century when Dutch ophthalmologist F. C. Donders used them as a way of quantifying the level of mental effort required to complete different cognitive tasks (Yan and Tourangeau, 2008). Later in the twentieth century, psychologists used response times to attitude questions to measure attitude strength, with the rationale that more crystalized and strongly held attitudes will be more accessible in working memory and so will take less time to retrieve and report (Fazio, 1990; Dovidio and Fazio, 1992). With the advent of paradata in the form of automatically generated time stamps in computer assisted interview (CAI) software, survey researchers have now also turned their attention to response latencies, with the hope that understanding the processes that generate their underlying variability might prove fruitful for delivering improvements in data quality and cost efficiency. This new line of research has identified a small but consistent interviewer contribution to response times in both face-to-face and telephone surveys (Couper and Kreuter, 2013; Loosveldt and Beullens, 2013; Olson and Smyth, 2015). However, while it is now well established that interviewers do affect response times, considerably less is known about how or why this effect arises. In this study we have sought to address this gap in our understanding by linking interviewer characteristics from a survey of interviewers to response times captured in a face-to-face interview survey. Our results largely corroborate those of existing studies. We find a similar total interviewer contribution to studies, with 3% of the total variability in response latencies attributable to interviewers, compared to 3% found by Olson and Smyth (2015) and 2% by Couper and Kreuter (2013). That these estimates are considerably smaller than those reported by Loosveldt and Bullens (2013) is likely due to the latter study's use of response times measured at the question block rather than the item level.

We also presented a new method for calculating interviewer-specific ICCs. This revealed substantial variability in the interviewer contribution, albeit that no interviewers contributed more than 4% to the total variability in response times. A potential application of interviewer-specific ICCs would be to identify undesirable interviewer behaviour such as skipping questions or fabricating interviews, which would manifest as interviewers making unusually high or low contributions to variability in response times across the respondents they interview. That we do not observe any such outliers in the UKHLS is not surprising given its adherence to the very highest methodological standards of interviewer training, supervision, and monitoring. However, on surveys where standards are not as rigorously enforced, this indicator could be of value in flagging problematic interviewer behaviour.

Question characteristics make the most substantial contribution to response latencies, with 45% of the total variability located at the question level and respondents accounting for a somewhat larger proportion of the variance compared to interviewers, at 5%. Our study is the first, to our knowledge, to estimate the contribution of areas to response times, even though this in effect serves only as a control for non-random allocation of respondents to interviewers. We find an area variance component equating to just half a percent of the total variance in response latencies, suggesting that it is probably not necessary to account for areas in future studies. We also find a similar pattern of associations between question characteristics and response times, with questions that require greater input or cognitive effort on the part of interviewers and respondents producing longer latencies. There are, to be sure, some discrepancies between our question level findings and those reported in existing studies but this is perhaps not surprising given differences in the designs of the surveys and in the way question level characteristics have been operationalized across studies.

A potential application of the question level estimates would be to use predicted scores from the model to obtain accurate estimates of interview length for new surveys. This would require questions in the new survey to be coded according to the characteristics that significantly influence response times. Model parameters could then be used to produce predicted item-level response times and then summing over the response times for all items in the survey. For example, a 10-word test item placed later in a battery with 2 or more response categories, a help screen but with no show card, definition, visual emphasis or interviewer instructions has a predicted response time of 22.5 seconds. Such an approach could potentially produce more accurate estimates of interview length than the approximate heuristics that are currently the prevalent approach in the industry.

At the respondent level we again corroborate the findings of Couper and Kreuter (2013) in identifying age and educational level as predictors of response latencies, with older and less educated people taking longer to answer. We find a small difference by sex, with women providing very slightly shorter item-level response times than men, a pattern which was also found by Vandenplas et al (2017) in the majority of ESS countries they considered. Respondents who provide undifferentiated responses to adjacent attitude items have significantly shorter latencies, an effect also observed by Zhang and Conrad (2015) and consistent with the theory that some respondents use short-cuts to reduce the cognitive costs of questionnaire completion (Krosnick, 1991). Selecting mid-points and providing 'Don't Know' responses which have also been used as indicators of satisficing in the empirical literature (Roberts et al, 2019) were unrelated to response times, lending further support to the contention that these measures are problematic as empirical indicators of satisficing (Sturgis et al, 2008; Luskin and Bullock, 2011; Turner et al, 2015).

Turning to the primary focus of this paper – interviewers - we find interviewing experience to be negatively related to latency length, again mirroring a finding in the Couper and Kreuter study. It would seem that one of the skills interviewers acquire as they accrue experience in their jobs is an ability to control the pace of an interview and that this enables them to conduct interviews in a more time-efficient manner. Although Olson and Smyth (2015) did not find a significant effect of experience, this may be attributable to the small number of interviewers (22) in their study. None of our expectations relating to the interviewer level psychological variables are supported; job satisfaction, attitude to surveys, attitude to persuading reluctant respondents, and the dimensions of the Big Five personality inventory are all unrelated to response times. The set of interviewer characteristics were also unrelated to the interviewer-specific ICCs, although the direction and magnitude of the coefficients were suggestive of a negative relationship between interviewer experience and the size of the interviewer contribution to variability in response times. One possible explanation for the lack of significant effects is that there is insufficient variability between interviewers on these characteristics for them to be consequential for the speed at which they conduct interviews (Couper and Kreuter, 2013). For example, it might be that there are relatively few interviewers who are low on conscientiousness, relative to the general population, because unconscientious individuals are selected out of the interviewer pool through both recruitment and attrition. Another potential reason for the lack of significant interviewer fixed effects is the lag of two years between the time the interviewer survey was undertaken and the fieldwork for wave three of UKHLS. It is certainly possible that some of the interviewer attitudes may have changed in the intervening period and this would likely depress associations between these variables and the response latencies.

Be that as it may, we do not find support here for the idea that variability between interviewers in response latencies is driven by their attitudes and personality. It is also worth noting that we did not find any substantively meaningful interactions between interviewer and respondent characteristics. Because we are using a latent timer that incorporates both the time the interviewer takes to read questions and to input the answers in addition to the time taken by respondents to provide answers, we cannot separately identify the extent to which, interviewers *moderate* respondents' response times. That we do not find any notable interactions between interviewer and respondent characteristics, however, suggests that the interviewer contribution to latencies is mostly or entirely their own direct contribution to the response time.

REFERENCES

Bassili, J. N., & Fletcher, J. F. (1991). Response time measurement in survey research: A method for CATI and a new look at nonattitudes. *Public Opinion Quarterly,* 55: 331–346.

Bassili, J. N. and Scott, B. S. (1996) Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, 60, 390–399.

Bassili, J.N. and Krosnick, J. A. (2000) Do strength-related attitude properties determine susceptibility to response effects?: new evidence from response latency, attitude extremity, and aggregate indices. *Political Psychology*, 21, 107– 132.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Benet-Martinez, V., & John, O.P. (1998). Los Conco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75(3): 729-750.

Brunton-Smith, I., Sturgis, P., & Leckie, G. (2017). Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location-scale model. *Journal of Royal Statistical Society, Series A*, 180 (2), 551-568.

Brunton-Smith, I., Sturgis, P., & Willams, J. (2012). Is success in obtaining contact and cooperation correlated with the magnitude of interviewer variance? *Public Opinion Quarterly*, 76(2): 265-282.

Burton, J., Knies, G., & Al Baghal, T. (2014). *Understanding Society – UK Household Longitudinal Study: Interviewer Survey 2014, User Manual*. Colchester: University of Essex.

Campanelli, P., & O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity,* 33(1): 59-76.

Campanelli, P., Sturgis, P., & Purdon, S. (1997). Can you hear me knocking? An investigation into the impact of interviewers on survey response rates. NatCen.

Cheng, A., Zamarro, G. & Orriens, B. (2018) Personality as a predictor of unit nonresponse in an Internet Panel, *Sociological Methods and Research*, 1-27.

Costa, P.T., & McCrae, R.R. (1992) Four ways five factors are basic. *Personality and Individual Differences,* 13(6): 653-665.

Couper, M. P. (1998) *Measuring survey quality in a CASIC environment*. Joint Statistical Meeting of the American Statistical Association, Dallas, Aug

Couper, M.P., & Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society Series A*, 176 (1), 271-286.

de Leeuw, E. D. (1999). How Do Successful and Less Successful Interviewers Differ in Tactics for Combating Survey Nonresponse? *Bulletin De Methodologie Sociologique* 62: 29–42.

Dovidio, J.F., & Fazio, R.H. (1992). New technologies for the direct and indirect assessment of attitudes. In Questions about questions: Inquiries into the cognitive bases of survey (eds Tanur J.M.), pp. 204-237. New York: Russel Sage Foundation.

Draisma, S., & Dijkstra, W. (2004) Response latency and (para) linguistic expressions as indicators of response error. In *Methods for Testing and Evaluating Survey Questionnaires* (eds S. Presser, J. Rogthgeb, M. Couper, J. Lessler, E. Martin, J. Martin and E. Singer), pp. 131–147. Hoboken: Wiley.

Durrant, G. B., R. M. Groves, L. Staetsky, & Steele, F. (2010). Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys. *Public Opinion Quarterly*, 74, 1–36.

Faust, M. E., Balota, D. A., Spieler, D. H. and Ferraro, F. R. (1999) Individual differences in information processing rate and amount: implications for group differences in response latency. Psychological Bulletin, 126, 777– 799.

Fazio, R. H. (1990) A practical guide to the use of response latency in social psychological research. In *Review of Personality and Social Psychology*, vol. 11, Research Methods in Personality and Social Research (eds C. Hendrick and M. S. Clark), pp. 74–97. Newbury Park: Sage.

Gummer, T., & Rossmann, J. (2015). Explaining interview duration in web surveys: A multilevel approach. *Social Science Computer Review,* 33(2): 217-234.

Jackle, A., Lynn, P., Sinibaldi, J. & Tipping, S. (2013). The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-operation with Face-to-Face Surveys. *Survey Research Methods*, 7: 1–15.

John, O.P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspective. In *Handbook of personality: Theory and research* (eds Pervin, L.W. and John, O.P.). New York/ London: The Guilford Press.

Knies, G. (2018). *Understanding Society: The UK Household Longitudinal Study Waves 1-7. User Guide*. Institute for Social and Economic Research, Colchester.

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

Lehtonen, R. (1996). Interviewer Attitudes and Unit Nonresponse in Two Different Interviewing Schemes. In *International Perspectives on Nonresponse* (ed. S. Laaksonen), Proceedings of the Sixth International Workshop on Household Survey Nonresponse, Statistics Finland, Helsinki 130–140).

Loosveldt, G., & K. Beullens (2013). The Impact of Respondents and Interviewers on Interview Speed in Face-To-Face Interviews. *Social Science Research*, 42: 1422–1430.

Loosveldt, G, Wuyts, C., & Beullens, K. (2018). Interviewer variance and its effects on estimates. *Quality Assurance in Education*, 26 (2): 227-242.

Lugtig, P., & Jackle, A. (2014). Can I just check..? Effects of edit check questions on measurement error and survey estimates. *Journal of Official Statistics*, 30(1): 45-62.

Luskin, Robert C., and John G. Bullock. (2011) “Don’t Know” Means “Don’t Know”: DK Responses and the Public’s Level of Political Knowledge. *Journal of Politics* 73 (April): 547-57

Mulligan, K., J. T. Grant, S. T. Mockabee, & Monson, J.Q. (2003). Response Latency Methodology for Survey Research: Measurement and Modeling Strategies. *Political Analysis*, 11: 289–301.

Olson, K., & Parkhurst, B. (2013). Collecting Paradata for Measurement Error Evaluations. In *Improving Surveys with Paradata: Analytic Uses of Process Information* (ed. F. Kreuter, pp. 43–72). Hoboken, NJ: John Wiley & Sons.

Olson, K., & Peytchev, A. (2007). Effect of interviewer experience on interview pace and interviewer attitudes. *Public Opinion Quarterly*, 71: 273–286.

Olson, K., & Smyth, J.D. (2015). The effect of CATI questions, respondents and interviewers on response time. *Journal of Survey Statistics and Methodology,* 3: 361-396.

Rees, P. H. (1971) Factorial ecology: An extended definition, survey and critique of the field. *Economic Geography* 47:220–33.

Roberts, C., Gilbert, E., Allum, N. & Eisner, L. (2019) Satisficing in surveys: A systematic review of the literature, paper presented at the Conference of the European Survey Research Association, Lausanne, Switzerland, July.

Scott, A., & Jessop, C. (2013). *UK Household Longitudinal Study (UKHLS) Wave 3*. Technical Report. NatCen, London.

Sturgis, P., Allum, N., & Smith, P. (2008). An experiment on the measurement of political knowledge in surveys. *Public Opinion Quarterly*, *72*(1): 90-102.

Tourangeau, R., Rasinski, K., & D’Andrade, R. (1991). Attitude Structure and Belief Accessibility. *Journal of Experimental Social Psychology*, 27: 48–75.

Turner, G., Sturgis, P., & Martin, D. (2015). Empirical indicators of survey satisfying: do they measure what we think they do? *Journal of Survey Statistics and Methodology*, *3*(1): 89-108.

Vandenplas, C., Loosveldt, G., Beullens, K. & Denies, K. (2017) Are interviewer effects on interview speed related to interviewer effects on straight-lining tendency in the European Social Survey? An interviewer-related analysis, *Journal of Survey Statistics and Methodology,* 0, 1-23.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience, and question complexity on web survey response times. *Applied cognitive psychology*, 22: 51-68.

West, B.T., & Blom, A.G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology,* 5(2): 175-211.

Zhang, C., & Conrad, F. (2015) Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2): 127-135.

TABLES AND FIGURES

Table 1: variance components and Intraclass correlations for the null model across studies

Level of model	Sturgis	Couper (females)	Couper (males)	Olson
Interviewer f_l	0.016	2.2	2.6	0.015
ICC	2.9%	1.5%	1.9%	3.2%
Area g_m	0.003	-	-	-
ICC	0.5%	-	-	-
Respondent v_k	0.027	3.2	6.2	0.035
ICC	4.9%	3.8%	6.3%	7.4%
Question u_j	0.251	-	-	0.255
ICC	44.9%	-	-	53.4%
Residual e_i	0.256	138.5	131.2	0.171
ICC	46.9%	94.7%	91.8%	36%

Table 2 Cross-classified multi-level model estimates including question, respondent, interviewer, and area characteristics (dependent variable=log of response latency)

Independent variables	Coefficient (standard error)
Intercept	1.546 (0.136)*
<u>*Question characteristics*</u>	
Question type (ref=Demographic)	
Attitude and perception	0.152 (0.040)*
Behaviour	0.111 (0.039)*
Test	0.600 (0.080)*
Other	0.026 (0.031)
Response options (ref= Yes/no)	
2 or more categories (single response option)	0.045 (0.032)
2 or more categories (multiple response option)	0.501 (0.060)*
Scale including Likert scale	0.370 (0.036)*
Open answer	0.382 (0.043)*
Number of words	0.012 (0.001)*
Showcard (ref = No)	0.134 (0.037)*
Help screen (ref= No)	0.120 (0.048)*
Transition statement (ref=No)	0.061 (0.062)
Battery (ref=Not a battery question)	
First in a battery	0.126 (0.071)
Later in a battery	-0.297 (0.027)*
Definition available (ref=No)	-0.000007 (0.060)
Sensitive questions (ref=No)	-0.048 (0.036)
Visual emphasis (ref=No)	0.160 (0.049)*
Interviewer instructions (ref=No)	0.175 (0.033)*
interviewer-administered (ref=Self)	-0.023 (0.085)
Section of questionnaire (ref= 1st quarter)	
2^{nd} quarter	0.018 (0.030)
3rd quarter	0.043 (0.038)
4th quarter	-0.070 (0.086)

*= p<0.05; ref = reference category; n = 3,100,288 (latencies); 13,904 (respondents) 1,894 (questions); 288 (interviewers); 3,187 (areas)

Table 2 Contd. Cross-classified multi-level model estimates including question, respondent, interviewer, and area characteristics (dependent variable=log of response latency)

Independent variables	Coefficient (standard error)
Intercept	1.546 (0.136)*
Respondent characteristics	
Gender (Female)	-0.007 (0.002)*
Age (years)	0.004 (0.00009)*
Highest qualification (ref = Degree)	
Other higher degree	0.030 (0.005)*
A-level	0.025 (0.001)*
GCSE	0.028 (0.001)*
Other qualification	0.056 (0.002)*
No qualification	0.035 (0.005)*
Total “Don’t knows”	0.002 (0.001)
Total mid-points	-0.00004 (0.0003)
Straight-lining	-0.027 (0.005)*
Interviewer characteristics	
Gender (female)	0.007 (0.016)
Banded age (ref= 71-79)	
61-70	-0.029 (0.029)
51-60	-0.053 (0.031)
41-50	-0.088 (0.037)*
31-40	-0.145 (0.057)*
Agreeableness	0.001 (0.011)
Conscientiousness	-0.001 (0.011)
Extroversion	-0.005 (0.007)
Neuroticism	0.010 (0.007)
Openness	-0.005 (0.009)
Survey too long (ref= Strongly agree)	
Agree	-0.040 (0.020)*
Disagree	-0.075 (0.023)*
Strongly disagree	-0.090 (0.045)*
Years of experience as interviewer (ref = 0-4 years)	
5-10 years	-0.038 (0.029)
11-42 years	-0.070 (0.030)*
Job satisfaction	0.011 (0.008)
Attitude to surveys	-0.008 (0.008)
Attitude to persuasion	-0.013 (0.008)
Area characteristics	
Economic disadvantage	0.007 (0.002)*
Urbanicity	0.011 (0.003)*
Transitory population	0.003 (0.002)
Age and housing structure	-0.002 (0.002)
Random effects	
Area	0.003
Interviewer	0.014

Question	0.124
Respondent	0.027
Residual	0.256

*p<0.05; ref = reference category; n = 3,100,288 (latencies); 13,904 (respondents) 1,894 (questions); 288 (interviewers); 3,187 (areas); AIC = 4620715.

TABLE 3 OLS Regression predicting interviewer specific ICCs using interviewer characteristics

Independent variables	**B**	**Standard error**	**Beta**
Female	-0.0004	0.0004	-0.0619
Banded age (71-79)			
61-70	-0.0006	0.0006	-0.0987
51-60	-0.0011	0.0007	-0.1985
41-50	-0.0012	0.0008	-0.1381
31-40	-0.0018	0.0013	-0.0961
Years of experience (ref: 0-4)			
5-10 years	-0.0008	0.0007	-0.1340
11-42 years	-0.0012	0.0007	-0.2190
Surveys too long (ref: strongly agree)			
Agree	0.0000	0.0004	0.0086
Disagree	-0.0005	0.0005	-0.0897
Strongly disagree	-0.0016	0.0010	-0.0982
Job satisfaction	0.0001	0.0002	0.0343
Attitude to surveys	0.0001	0.0002	0.0381
Attitude to persuasion	-0.0003	0.0002	-0.1180
Agreeableness	-0.0002	0.0002	-0.0404
Conscientiousness	0.0000	0.0002	0.0057
Extroversion	-0.0002	0.0002	-0.0983
Neuroticism	0.0001	0.0002	0.0477
Openness	-0.0003	0.0002	-0.1087
Constant	0.0346	0.0023	
R-square	0.0964		

n=288 (interviewers)

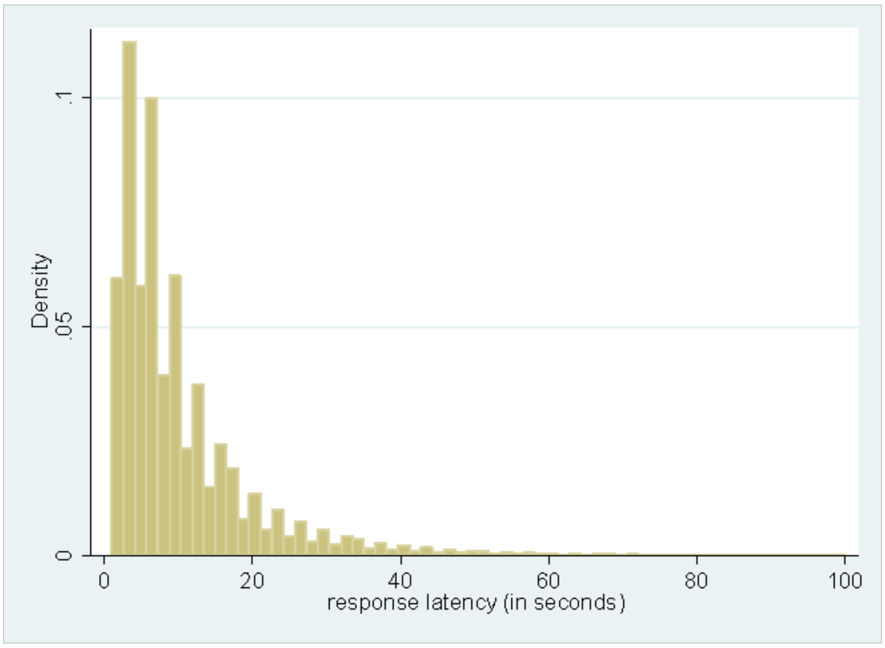

Figure 1 Response latency distribution (in seconds), Wave 3 UKHLS

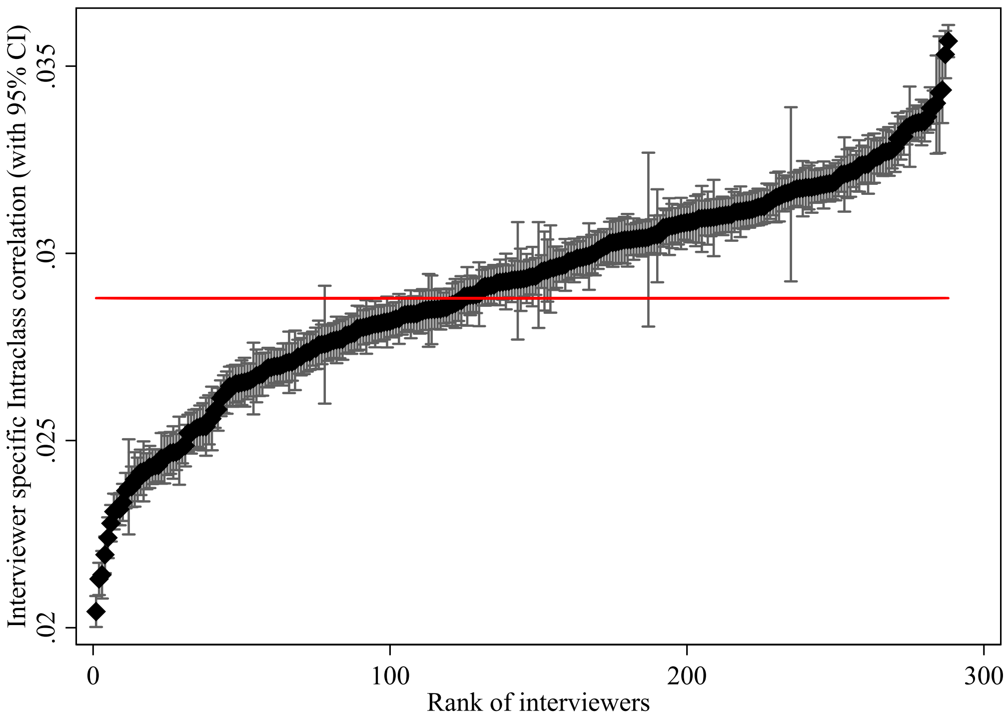

Figure 2 Interviewer specific Intra-Class Correlation coefficients

APPENDIX A

All questions in the wave three UKHLS questionnaire were coded by three research assistants at the University of Southampton. A random 10% sample of questions were coded by all coders in order to assess the proportion of agreement between the coders. The list of characteristics coded for each question is presented in Table A1.

Table A1: Characteristics of questions and agreement between three Research Assistants in coding of 10% of the sample of questions (223 questions).

Question Characteristic	Agreement between 3 coders in 10% of questions
Response category: 1. Yes/no 2. 2 or more categories 3. Scale including Likert scale 4. Open numeric answer 5. Open textual answer	72% (160)
Number of response categories (0:31 (mean 4.4))	92% (206)
Number of words (2:84 (mean 17.3))	Within 3 words difference: 90% (201)
Question type 1. demographic 2. attitude and perception 3. behaviour 4. test 5. other	99% (220)
Response options 1. single response 2. multiple responses 3. numeric answer 4. textual answer	95% (212)
Show card 1. yes 2. no	100% (223)
Help 1. yes 2. no	98% (219)
Sensitive question 1. yes 2. no	73% (163)
Transition statement 1. yes 2. no	73% (162)

First in a battery 1. yes 2. no	98% (219)
Later in a battery 1. yes 2. no	91% (202)
Definition 1. yes 2. no	98% (218)
Visual emphasis 1. yes 2. no	97% (216)
Interviewer instructions 1. yes 2. no	93% (207)
Page of question in questionnaire	100% (223)
Administration of questionnaire (self- vs interviewer-) 1. self-administered 2. interviewer-administered	100% (223)
Module number	100% (223)

Table A2: Characteristics of questions (1,894 questions)

Variables and categories	Frequencies	Percentages
<u>*Question characteristics*</u>		
Question type		
Demographic	312	16.5
Attitude and perception	419	22.1
Behaviour	561	29.6
Test	27	1.4
Other	575	30.3
Response options		
Yes/no	318	16.8
2 or more categories (single response option)	680	35.9
2 or more categories (multiple response option)	79	4.2
Scale including Likert scale	569	30.0
Open answer including both numeric and textual	248	13.1
Number of words	Min: 2.0, max: 84.0, mean: 17.1, SD: 10.4	
Showcard		

Yes	209	11.0
No	1685	89.0
Help screen		
Yes	73	3.8
No	1821	96.2
Transition statement		
Yes	72	3.8
No	1822	96.2
Battery		
Not a battery question		
First in a battery	1123	59.3
Later in a battery	55	2.9
	716	37.8
Definition available		
Yes	46	2.4
No	1848	97.6
Sensitive questions		
Yes	131	6.9
No	1763	93.1
Visual emphasis		
Yes	63	3.3
No	1831	96.7
Interviewer instructions		
Yes	511	27.0
No	1383	73.0
Interviewer-administered		
No (self-administered)	902	47.6
Yes	992	52.4
Section of questionnaire		
1st quarter	430	22.7
2nd quarter	406	21.4
3rd quarter	183	9.7
4th quarter	875	46.2

Table A3: Respondent (13,904), interviewer (288 interviewers) and area (3,187 areas) characteristics

Independent variables	Frequencies	Percentages
Respondent characteristics		
Gender		
Female	7,845	56.4
Male	6,059	43.6
Age (years)	Min: 16.0, max: 99.0, mean: 48.8, SD: 17.3	
Highest qualification		
Degree	3,319	23.9
Other higher degree	1,700	12.2
A-level	2,749	19.8
GCSE	2,950	21.2
Other qualification	1,480	10.6
No qualification	1,706	12.3
Total "Don't knows"	Min: 0.00, max: 32.00, mean: 0.31, SD: 1.26	
Total mid-points	Min: 0.00, max: 36.00, mean: 8.97, SD: 4.66	
Straight-lining		
Yes	1,372	9.9
No	12,532	90.1
Interviewer characteristics		
Gender		
Female	165	57.3
Male	123	42.7
Banded age		
71-79	23	8.0
61-70	117	40.6
51-60	109	37.8
41-50	32	11.1
31-40	7	2.4
Agreeableness	Min: 3.00, max: 7.00, mean: 5.86, SD: 0.75	
Conscientiousness	Min: 3.67, max: 7.00, mean:6.03; SD: 0.75	
Extroversion	Min: 2.00, max: 7.00, mean: 5.17, SD: 1.12	
Neuroticism	Min: 1.00, max: 7.00, mean: 3.01, SD: 1.17	
Openness	Min: 2.67, max: 7.00, mean: 5.19, SD: 0.92	
Survey too long		
Strongly agree	9	3.1
Agree	133	46.2
Disagree	89	30.9
Strongly disagree	57	19.8
Years of experience as interviewer		
0-4 years	23	8.0
5-10 years	136	47.2
11-42 years	129	44.8
Job satisfaction	Min: -1.66, max: 3.14, mean: 0.01, SD: 0.10	
Attitude to surveys	Min: -2.40, max: 3.40, mean: -0.01, SD: 1.00	

Attitude to persuasion	Min: -3.29, max: 2.57, mean: 0.10, 0.10
Area characteristics	
Economic disadvantage	Min: -2.42, max: 3.65, mean: -0.03, SD: 0.98
Urbanicity	Min: -1.67, max: 4.82, mean: -0.01, SD: 0.96
Transitory population	Min: -2.31, max: 11.79, mean: 0.03, SD: 1.04
Age and housing structure	Min: -3.11, max: 9.08, mean: 0.01, SD: 0.97

Interviewer Attitude Scales

Job Satisfaction

Now thinking about your interviewing job at NatCen, how satisfied were you with the following aspects of the job? (Very satisfied, Satisfied, Somewhat satisfied, not satisfied)

1. The amount of pay you receive
2. The interesting nature of the job
3. The ability to work independently
4. The interaction you have with people
5. The level of flexibility to choose your times of work
6. The level of flexibility to choose how much work to take on
7. The need to work during the 'evening and weekend'

Attitude to surveys

Below follows a series of statements in general (Strongly agree, agree, disagree, strongly disagree)

1. Surveys are important for science, politics, and the economy
2. Surveys help make society more democratic
3. Most surveys are carried out in a responsible way
4. In most cases survey results are correct
5. Far too many surveys are carried out in the UK

Attitude to persuading reluctant respondents

8. Below follows a series of statements on persuading respondents (Strongly agree, agree, disagree, strongly disagree)

1. Reluctant respondents should always be persuaded to participate
2. With enough effort, even the most reluctant respondent can be persuaded to participate
3. If a respondent is reluctant, a refusal should be accepted
4. It does not make sense to contact reluctant target persons repeatedly
5. If you catch them at the right time, most people will agree to participate
6. Respondents persuaded after great effort do not provide reliable answers