'How long will it take?' An analysis of interview length in the fifth round of the European Social Survey

Geert Loosveldt and Koen Beullens

Centre for Sociological Research, K.U. Leuven

The question 'How long will the interview take?' is frequently asked by interviewers during training and by respondents during the initial doorstep interaction. In this paper, we investigate the impact of different interviewers, countries and some respondent characteristics on interview length in the fifth round of the European Social Survey.

The results show substantial differences between countries with regard to interview length and reinforce that differences between countries are based on much more than just the differences between languages. The results support the obvious suggestion that fewer applicable questions reduce the interview length. Further, interviewing older respondents takes more time, and the duration also increases if a respondent more frequently asks for clarification. The huge impact of interviewers on interview length is the most remarkable result. In all countries, the difference between interviewers accounts for a significant and substantial part of the variance in interview length. More detailed fieldwork monitoring in each country is necessary in order to understand these differences. The results also clearly illustrate the necessity for investment in training, monitoring and follow-up of interviewers in each country participating in a cross national survey.

Keywords: Interview length; interviewer effects; European Social Survey

1 Introduction

'How long will the interview take?' can be considered as one of the standard questions asked by a respondent during their initial doorstep interaction with an interviewer, as the time taken by the interview is an important element of the respondent's evaluation of the cost of participation. The duration of an interview is also important for the interviewer with regard to planning the fieldwork activities and is a factor which influences his or her cost and benefit analysis. Short and well-paid interviews are financially and organizationally more attractive. So the answer to the question 'How long will the interview take?' is relevant to both interviewers and respondents.

During training, interviewers receive information about the expected interview length and instructions on how to reply to a respondent's question concerning this. In this reply, a reference to the modal interview length is appropriate. Based on the total number of questions, the number that are applicable per respondent, degree of complexity, and the format of the questions in the questionnaire, together with information from pilot interviews, the mean interview length can be estimated. It is generally assumed that these characteristics are the dominant determinants of interview length and that any deviation from the anticipated duration will be small. However, we know from experience that interview length can vary within one survey using the same questionnaire (Olson and

Peytchev 2007). In this paper, we will explore the impact of interviewers and respondent characteristics on interview length in the European Social Survey (ESS). First, we describe the measurement and distribution of interview length in the dataset of the fifth round of the ESS. Subsequently, the basic determinants of interview length and some hypotheses are discussed.

2 Data and measurement of interview length

The ESS is a biannual, cross-national survey carried out in several European countries. Face-to-face interviews (either Paper and Pencil Interviews, PAPI, or Computer-Assisted Personal Interviews, CAPI) are used for data collection. In this paper, data from the fifth round organized in autumn 2010 and spring 2011 is used. By December 2011, datasets from 20 countries were available on the ESS website (http://www.europeansocialsurvey.org).

The ESS questionnaire contains several core modules that remain relatively consistent from round to round (e.g. socio-demographic characteristics, national, ethnic and religious identity, socio-political values, etc.) and two rotating modules. In the fifth round, the rotating modules were concerned with 'Work, Family and Well-being' and 'Trust in Criminal Justice'. For this round, the estimated length of an interview was one hour.

For the fifth round of the ESS, the start time and date was recorded at the beginning of each interview (just before the first question, A1) and the end time after the last question in the main questionnaire (immediately after question G88). Although the time recording appears relatively simple, some

---

Contact information: Geert Loosveldt, Centre for Sociological Research, K.U. Leuven Parkstraat 45, 3000 Leuven, e-mail: Geert.Loosveldt@soc.kuleuven.be

problems with impossible or extreme values were observed. In the data from Estonia, the clock for the start time was incorrect (12 instead of 24). Therefore, Estonia was omitted from the dataset and only 19 countries were used in the analysis. In addition, some extreme values were excluded from the analysis: less than 30 minutes (1.81% of respondents) and more than 180 minutes (0.53% of respondents). The selection of these extreme values is based on the idea that on the one hand an interviewer needs a minimum amount of time to read the minimum number of applicable questions on the one hand, whilst on the other hand the interview can last longer due to unexpected circumstances. For the first round of the ESS, the estimate of the lower limit for an extremely short interview, including only the questions that all respondents had to answer (approximately 180), was 45 minutes (Japec 2005). Based on the questionnaire and the empirical distribution of interview length in the fifth round, we opted for a more extreme minimum value of 30 minutes. The selection of the maximum value is more arbitrary. Because of possible unexpected circumstances, we opted to use data from as many respondents as possible and not to reduce the variability in interview length. So, an extreme maximum value is used. Figure 1 shows small frequencies of high values for interview length: only 0.43% of the respondents are in the interval 150–180. Therefore, the impact on the results caused by the selection of this high maximum value should not be overestimated. In combination with a few missing values for the basic variables, we arrive at a valid measurement of interview length in the range of 30 to 180 minutes for 97.4% of the respondents.

Some basic descriptive statistics show that the distribution of interview length for all respondents (n = 36,221) is skewed to the right (Figure 1): mean interview length is 68 minutes and the median equals 65. The variability is considerable: the standard deviation is 23 minutes; the interquartile range is 24 (the first quartile = 54 minutes and the third quartile = 78 minutes). Figure 1 also clearly shows higher occurrences of multiples of five minutes. This pattern is typical for rounding errors. No less than 38.4% of the interviews have a recorded length that is a multiple of five, whereas the interval from 30 to 180 minutes contains 31 multiples of five and the expected percentage of interviews with a length that is a multiple of five is 20%.

The two different data collection methods used in the ESS (Paper and Pencil Interviews, PAPI, and Computer-Assisted Personal Interviews, CAPI) can have an effect on the accuracy of the measurement of interview length. It can be assumed that by using timers in the CAPI system, it is possible to measure the interview length more precisely without rounding errors. We consider the occurrence of multiples of five to be an indication of rounding errors. Of the interviews carried out using the PAPI method, 64.73% show a multiple of five for duration, compared with the expected 20% for those using CAPI. This is a clear indication that the manual recording of times in PAPI is less precise and characterized by more rounding errors than the automatic recording in CAPI.

In CAPI, interviewers do not need to pay attention to the routing in the questionnaire. Therefore, it can also be assumed that using CAPI, interviews progress smoothly and the mean interview length is shorter than with PAPI. The results support this assumption. The difference between CAPI and PAPI with regard to mean interview length is small (2 minutes), but statistically significant (PAPI = 69.4 minutes versus CAPI = 67.4; t = 8.98, p<0.0001). It must be recognised that the distinction between the two data collection modes (CAPI and PAPI) is also a distinction between two groups of countries. In the next section, the differences between countries are examined in greater detail.

Although it is not the aim of this paper to replicate the whole analysis using ESS round 4 data, Figure 2 presents the distribution of interview length in ESS round 4 for the same countries as in round five (mode = 60 minutes, median = 67 minutes, mean = 72 minutes and standard deviation = 22 minutes). The similarity to the distribution in round 5 is striking. The same shape and pattern of frequencies for the multiple of five minutes is observed. Therefore, the typical characteristics of the distribution of interview length observed in round 5 are neither unique nor exceptional.

3 Basic determinants of interview length

In the previous section, considerable variance in interview duration was noted. How can this be explained? The starting point for an analysis is that both respondents and interviewers are responsible for the duration of an interview. Therefore, a distinction must be made between a respondent-oriented and an interviewer-oriented analysis of interview length. This distinction is related to the cognitive and communicative processes underlying the asking and answering of questions in survey interviews.

In a cross-national comparative survey, in addition to the impact of respondent characteristics and interviewers on interview length, it is also necessary to evaluate the differences between countries. Successively, we discuss three determinants of interview length: respondent, interviewer and country

A respondent-oriented approach   A respondent-oriented analysis of interview length is focused on the respondent's cognitive effort and response time, and response latency is the key concept. Response latency is the time it takes a respondent to answer a survey question and is considered to be an indirect indicator of 'the amount of information processing necessary to answer a question' (Bassili and Scott 1996). The characteristics of respondents and questions that influence this (e.g. age, education, complexity of the question and response categories, memory retrieval, etc.) have an effect on information processing and the resulting response time (Bassili 1996). Mostly, the analysis of response latency is related to the survey questions and is considered an efficient way to detect problems with questions (Bassili and Scott 1996; Presser et al. 2004). In general, response time can be considered an important part of a respondents contribution to the interview length, which is informative regard-
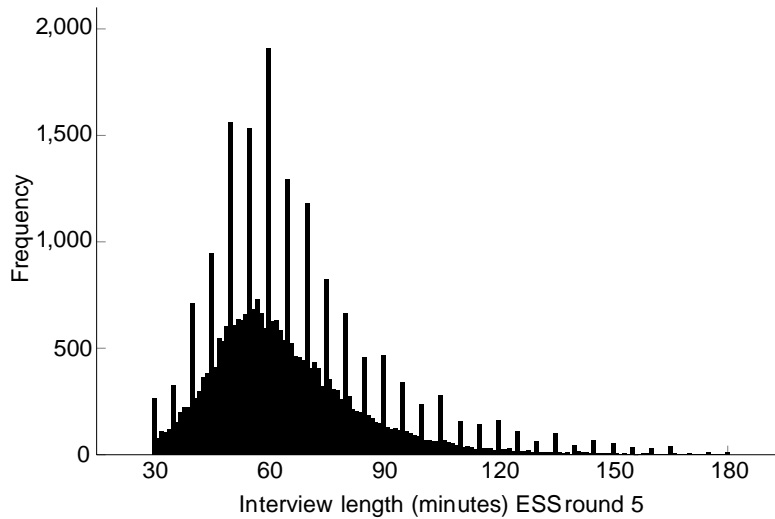
Figure 1. Distribution of interview length for all respondents in ESS round 5 (n = 36221)
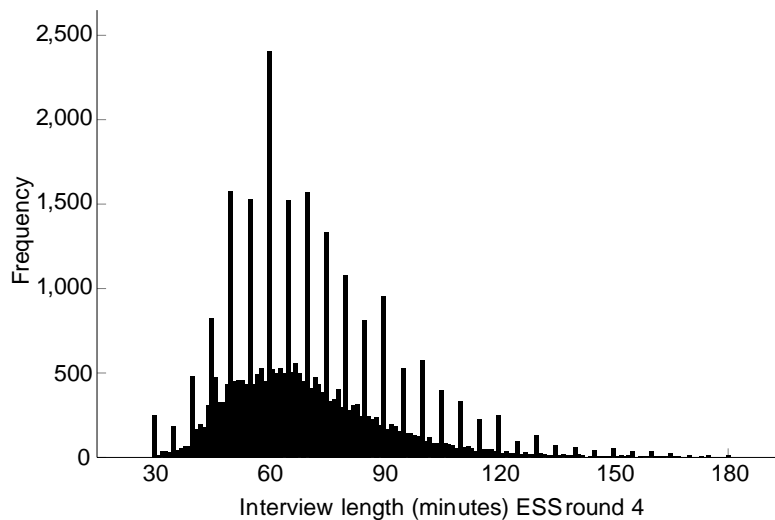

Figure 2. Distribution of interview length for all respondents in ESS round 4 (n = 37662)

ing the quality of the response process and is influenced by the respondents cognitive capacity and motivation. The respondents educational level and age may both be relevant in this context. Both variables are considered as proxy variables for the respondent's cognitive abilities (Olson and Peytchev 2007). It could be assumed that older people need more time to answer questions and that respondents with a higher educational level are more familiar with answering questions, and as a consequence will need less time. With regard to interview length, we therefore expect a positive effect from age and a negative effect from the respondent's educational level. Results published in previous research by Olson and Peytchev (2007) support these expectations.

The ESS questionnaire also contains a separate section in which interviewers record their assessment of some respondent and interview characteristics. Variables from this additional data are appropriate to use as an indirect assessment of a respondent's motivation and capacity to answer items in the main questionnaire. The frequency of asking for clarification and of understanding the questions (questions J1 & J4 with a five-point scale: 1 = never, to 5 = very often) can be considered as an indirect evaluation of the respondent's capacity. It can be assumed that interviews with respondents having less capacity (more requests for clarification and less understanding of the questions) go less smoothly and take more time. A question regarding the frequency with which the respondent tries to answer questions to the best of their ability (question J3) is an indirect assessment by the interviewer of the respondent's motivation. It can be assumed that an interview with highly motivated respondents (making greater efforts to answer questions to the best of their ability) will take more time.

Although the same questionnaire is used for each respondent, every respondent does not necessarily answer the same number of questions. Depending on some respondent characteristics (e.g. the number of household members, partner, type of job, etc.) the number of questions answered varies between respondents. Following on from this, the number of applicable questions can be considered as a summary of the respondent characteristics that can affect the interview length. It is clear that an increase in the number of applicable questions (or a decrease in the number of not applicable questions) must have a positive effect on interview duration.

To sum up, as detailed in this section, we expect an effect on interview length resulting from a respondents age and level of education, the variables assessed by the interviewer (clarification, understanding and ability) of the respondent's motivation and capacity, and the number of (not) applicable questions.

An interviewer-oriented approach   Interview length can also be assumed to be influenced by the way an interviewer performs their task during the interaction and communication process with a respondent. In an interviewer-oriented approach to analysing interview length, the impact of the interviewer on the duration is the central focus. In this approach, interview length is considered as a proxy indicator of several aspects of interviewer behaviour that can each influence the duration of an interview (e.g. the pace of reading questions and instructions, the way the interviewer clarifies questions and reacts to inadequate respondent behaviour, etc.) (Olson and Peytchev 2007). In surveys carried out by interviewers trained according to the key principle of standardized interviewing, it can be assumed that the interviewer has no, or only limited, impact on interview length. In this case, interviewers apply the same basic task rules during the interaction with respondents, and they expend the same efforts to obtain adequate responses. Therefore, the interviewer's effect on interview length should be almost the same for every interview with a similar respondent.

Given the characteristics of standardized interviewing, one can estimate a 'normal' duration for an interview with an ordinary respondent and expect only random differences in interview length between interviewers when they interview similar groups of respondents. However, in 1981 Cannell, Miller and Oksenberg noted that "...interviewers proceed through the interview too quickly; they ask questions at a rapid rate, permit no pause between response and the next question, and jump into the interaction with fast feedback. The atmosphere created is hasty and casual. The interviewers major goal seems to be to finish the interview" (Cannell, Miller and Oksenberg 1981:415).

Research into interview length and interview speed is currently lacking. The few reported results regarding interviewer effects on interview length show that expectations based on the principles of standardized interviewing are not supported. In an analysis of the number of questions completed per minute, Hox reported an intra-interviewer correlation of 0.14 (Hox 1994). This means that 14% of the variance

of the speed of interviews is explained by differences between interviewers. Higher intra-interviewer correlations of interview length are reported by Olson and Peytchev (2007). In their analysis of data from the pre-election 1984, 1988 and 2000 National Election Studies conducted by the Survey Research Center (SRC) at the University of Michigan, they show that 22, 27 and 24 per cent of the variance of interview length in 1984, 1988 and 2000 respectively, is explained by differences between interviewers. These intra-class correlations are quite high and make it clear that some interviewers systematically interview significantly faster or slower than others. The results also show that the level of experience of an interviewer and the order in which interviews are held (experience over a survey's field period) have a significant, negative effect on interview length.

In line with these results, and in contrast with the basic assumptions of standardized interviewing, we must assume that interviewers do their jobs differently to an extent, and accordingly affect interview length. In this context, it is relevant to evaluate the effects of the order of interviews and the interviewer's workload. For both variables, a negative effect on interview length is expected.

A cross-national approach   Although the main focus of this paper is to evaluate the respondent and interviewer oriented approaches to analysing interview length, in the next section the differences between countries is also examined.

In a cross-national comparative survey such as the ESS, it is extremely important that the data collection in each country is organized in the same way. In each country participating in the ESS, a National Co-ordinator (NC), in collaboration with a fieldwork organization, is responsible for the implementation of the survey according to the ESS fieldwork specifications. One important task is the translation of the English source questionnaire into each country's language(s). To avoid resulting differences between countries in the wording and order of questions, the specifications for the translation procedure are very strict. These aim to ensure that identical questionnaires (the same question wording, order, response categories, skipping patterns, etc.) are used for all countries. Specifications for other parts of the data collection process (sampling, interviewer selection and training) are also intended to standardize the implementation of the survey. Because the same main questionnaire and a representative sample were used in each country, there are very few arguments to support differences between countries as being relevant to interview duration. The data collection mode (CAPI or PAPI) could be considered as a relevant country characteristic. However, in the previous section we have already seen that the difference in interview length between the two methods is small. Language is another country-specific variable which could be responsible for differences in interview length between countries. It is possible that in some languages, more words and/or more or longer sentences are needed in order to express the same idea. In the next section the impact of language will be evaluated.

4 Results

4.1 Evaluation of the differences in measured interview length between countries

Given the previously described great importance attached to standardized interviewing and the comparability of fieldwork procedures, no significant differences between countries should be expected regarding interview length. However, the results shown in Table 1 do not support this expectation. With country as a fixed factor in a one-way analysis of variance, the differences between countries is significant (F = 789.9; df = 18; 36202; p<0.0001) and 28% of the variance is explained. The highest mean interview length is observed in the Czech Republic (101.9 minutes) and the lowest mean in Slovenia (50.6 minutes). Both countries used PAPI and also have a high percentage of 'multiple of five' interview lengths. When we remove the Czech Republic from the analysis, the explained variance is halved. Other countries with a remarkably low mean for interview length (<60 minutes) are Switzerland, Israel and Portugal. Countries with a high mean value (>75 minutes) are Germany and Poland. It is clear that this clustering of countries based on mean interview length differs from the grouping based on mode. In Hungary and Russia all the measurements of interview length are in the range between 30 and 180 minutes (100% of valid length). The observed differences support the idea that there are real differences between countries in fieldwork activities and the survey culture.

As previously mentioned, some languages may need more words and longer sentences to express the same idea. As a consequence, language differences might be responsible for systematic differences between countries. However, countries do not perfectly overlap in terms of language. Using the variable 'interviewer questionnaire language' we can test the impact of language more directly. The highest values for mean interview duration are realized for Czech (101.2 minutes) and Catalan (86.7 minutes) languages. In Hebrew, the mean is merely 48.2 minutes and in Slovenian 50.6 minutes.

The difference between languages explains 26.5% of the variance in interview length. This is comparable with the differences between countries. When we use language and country simultaneously as an independent variable in the analysis of variance, the proportion of explained variance (28.9%) barely increases compared to when both variables are used separately. Although the partial effects of both variables are significant, the partial sum of squares for country (SS = 409,286) is much larger than the partial sum of squares for language (SS = 117,800). This means that the difference between countries is more important than the difference between languages. The difference between the French and the German speaking parts of Switzerland (French = 59.6 minutes; German = 56.2 minutes; difference = 3.4) is much smaller than the difference between Germany (75.7 minutes) and the German speaking part of Switzerland (difference = 19.5), and the difference between France (68.3 minutes) and the French speaking part of Switzerland (difference =

8.7). The difference between the two languages in Belgium (French = 61.1 minutes and Dutch = 59.3 minutes; difference = 1.8) is also smaller than the difference between France (68.3 minutes) and the French speaking part of Belgium (difference = 7.2) and slightly smaller than the difference between the Netherlands (61.9 minutes) and the Dutch speaking part of Belgium (difference = 2.6).

The previous analysis assumes that the language of the questionnaire is also the language of the interview. However, this does not necessarily mean that the language of the interview and questionnaire is the same as the language most often spoken by the respondent at home. The language used in the interview differs from the language spoken at home for 6.4% of the respondents. It could be expected that this language difference would hinder the smooth proceeding of an interview, but this expectation is not supported by the data. There is a small but statistically significant increase of about two minutes when the primary language at home and the language of the questionnaire are the same (same language = 68.2 minutes; other language = 66.5 minutes; t = -3.99, p<0.0001). This means the interview takes slightly longer when the language spoken at home is the same as the interview language. It is possible that the same language facilitates a conversation about some of the topics in the questionnaire.

4.2 Effects of respondent characteristics on interview length

In the previous section, we argued that age, education, the interviewer's assessment of the respondents motivation and capacity, and the number of (not) applicable questions are relevant respondent characteristics in order to explain interview length. To evaluate the impact of these variables, a multiple regression analysis for all units together and for each country separately was carried out. Table 2 illustrates the proportion of explained variance in interview length and the standardized regression coefficients of the aforementioned variables. The results presented in Table 2 once again illustrate differences between countries. There are differences in the explanatory power of the regression model, and the pattern and magnitude of the standardized regression coefficients effects is not the same for each country.

Although in most countries the respondent characteristics can explain a significant and substantive proportion of variance in the interview length, in some countries this proportion is remarkably small (less than 5%). This is true for Bulgaria, the Czech Republic, Israel, Portugal and Russia. With the exception of Portugal, the data collection mode in these countries was PAPI. This leads to the suspicion that time recording using PAPI was not sufficiently precise. In all countries, the number of not applicable questions has the strongest and most significant effect on interview length. It is clear that more not applicable questions decrease the interview length.

The second most important variable in the regression model is the interviewer's evaluation of the frequency of the respondent asking for clarification of the questions (1 =

Table 1: Measurement of interview length by country

	n	Mode	Mean	std. dev.	% multiple of five	Valid length
Belgium	1670	CAPI	60.0	16.5	28.8	98.0
Bulgaria	2427	PAPI	61.6	12.9	85.7	99.8
Switzerland	1435	CAPI	57.3	17.5	19.9	95.3
Czech Republic	2343	PAPI	101.9	27.4	85.7	98.1
Germany	2967	CAPI	76.7	23.7	19.8	97.8
Denmark	1555	CAPI	64.8	18.3	32.7	98.7
Spain	1825	CAPI	69.3	20.8	19.2	96.9
Finland	1841	CAPI	61.6	19.2	19.6	98.0
France	1710	CAPI	68.3	19.1	19.8	99.1
United Kingdom	2336	CAPI	59.5	16.6	20.1	96.3
Hungary	1561	PAPI	68.2	14.2	74.8	100
Israel	1947	PAPI	51.2	15.2	70.0	83.6
Netherlands	1825	CAPI	61.9	18.9	18.7	99.8
Norway	1514	CAPI	67.1	21.0	19.6	97.6
Poland	1731	PAPI	77.2	21.0	60.1	98.8
Portugal	2107	CAPI	55.6	8.9	18.6	98.2
Russia	2595	PAPI	66.4	17.8	62.4	100
Sweden	1484	CAPI	64.4	17.9	21.7	98.9
Slovenia	1348	PAPI	50.6	14.8	75.9	96.1

Table 2: Multiple regression with respondent characteristics: R-square and standardized regression coefficients

	R-square	No. of not app.	age	Education	Clarification	Ability	Understanding
All countries	0.07	-0.19	0.06	0.06	0.18	0.05	0.05
Belgium	0.23	-0.36	0.07	0.04°	0.30	0.05	0.03°
Bulgaria	0.02	-0.09	-0.04°	0.07	0.08	0.00°	-0.06
Switzerland	0.17	-0.27	0.12	0.08	0.26	-0.01°	-0.03°
Czech Republic	0.03	-0.08	0.04°	-0.01°	0.10	0.09	0.01°
Germany	0.11	-0.23	0.05	0.10	0.17	0.02°	0.01°
Denmark	0.17	-0.28	0.18	0.02°	0.28	0.00°	-0.01°
Spain	0.16	-0.36	-0.00°	0.07	0.16	0.03°	0.07
Finland	0.18	-0.25	0.22	-0.04°	0.26	-0.04°	0.03°
France	0.20	-0.29	0.11	0.11	0.31	0.07	-0.01°
United Kingdom	0.12	-0.28	0.13	0.06	0.16	0.05	-0.03°
Hungary	0.05	-0.11	0.12	0.09	0.06	-0.06	-0.08
Israel	0.01	-0.06	0.02°	0.05	-0.04°	-0.06°	0.03°
Netherlands	0.14	-0.31	0.12	0.01°	0.27	-0.01°	0.01°
Norway	0.19	-0.32	0.19	-0.03°	0.24	0.06	-0.07
Poland	0.08	-0.21	0.13	0.01°	0.11	-0.07	0.05°
Portugal	0.00	-0.05	-0.05°	0.03°	0.03°	-0.00°	-0.01°
Russia	0.03	-0.12	0.05	0.00°	0.15	-0.02°	0.04°
Sweden	0.21	-0.30	0.22	0.08	0.30	-0.01°	0.01°
Slovenia	0.16	-0.27	0.15	-0.02°	0.27	0.04°	0.04°

Note: Coefficients with ° are not significant at a 5% level. A value for a standardized regression coefficient equal to 0.00 also includes values less than 0.00.

never, to 5 = very often). All the significant effects are positive. This means that the interview duration increases when respondents more frequently ask for clarification of questions. The impact of the two other variables from interviewer evaluations is less convincing. Controlling for the other variables in the regression model, the interviewer's evaluation of the respondent's understanding of the questions has a significant effect in only four countries and the evaluation of answering questions to the best of the respondents ability in seven countries.

As anticipated, interviewing older people takes more time. In 14 out of 19 countries, the effect from the age of respondents is significant, and when significant it is positive. The impact of level of education is less apparent, and is significant for only nine countries. Contrary to expectations based on previous research (Olson and Peytchev 2007), the effect in these nine countries is positive. This means that the interview length increases when the educational level of the respondent is higher. It should be noted that the effect of level of education is a partial effect (controlling for clarification, ability and understanding). It is also possible that this partial effect of education indicates a respondent's tendency to discuss the subject matter of questions and to elaborate on his or her answers.

4.3 Evaluation of interviewer effects on interview length

In the ESS, interviewers are instructed to perform their tasks according to the key principle of standardized interviewing. Standardized interviewing aims to ensure that all respondents answer exactly the same questions under conditions that are as consistent as possible (Groves et al. 2004). The main argument in favour of this interview style is the reduction in differences between interviewers, which could otherwise lead to identifiable interviewer-related error (Fowler and Mangione 1990). Standardized interviewing implies that interviewer effects on interview length will be limited.

The starting point for the evaluation of interviewer effects in a cross-national survey is a three level hierarchical data structure. Respondents (the first level) are nested within interviewers (the second level) and interviewers are nested within countries (the third level). Both interviewers and respondents are considered as random samples. However, the participating countries in the ESS are not a random sample of European countries and the number of countries used in the analysis is rather small (19). Because of this, we do not consider the country as a random factor. This means that there is a three-level hierarchical data structure, but we use the country only to classify the interviewers. This type of data structure results in a two-level conditional hierarchical linear model (Bryk and Raudenbush 1992), in which the respondents are nested within each interviewer and the interviewers are classified into countries. Resultantly, countries are considered as levels of a fixed effect and the resulting model is a mixed model with random and fixed effects (Littell, Milliken, Stroup and Wolfinger 1996).

In a model for the whole sample of 36,221 respondents and 2,979 interviewers, and with the interviewers nested within the countries as a random factor and countries as a fixed factor, one third of the variance in interview length is explained by the differences between interviewers. So the (partial) intra-interviewer correlation for the whole sample after controlling for country is 33%. This percentage is considerable and does not suggest the general application in all countries of the key principle of standardized interviewing. It must be noted that this is an approximate evaluation of the differences between interviewers. Because interviewers are allocated to a particular area, the respondents are not randomly assigned to the interviewers. Therefore, it is possible that interviewers did not interview similar groups of respondents and that the differences between interviewers are partially area effects.

However, we can use the results of this initial model as reference point from which to evaluate the impact of other variables, and the difference between countries in terms of interviewer effects. In the next step we will elaborate on the initial model using the relevant respondent characteristics noted in the previous section (age, education, number of not applicable questions and asking for clarification) and we will execute a multilevel model for each country separately. The introduction of these variables into the model results in new estimates of residual variance and interviewer-related variance ($\sigma^2_{e|c}$ and $\sigma^2_{u|c}$). It can be assumed that the additional variables explain part of the variance in the dependent variable. As a consequence, the variance components in the random part will decrease. The decrease in the residual variance at the respondent level means that part of the variance in the dependent variable is explained by the additional variables. The decrease in the interviewer-related variance means that the distribution of independent variables is not the same for all interviewers, and that part of the interviewer variance can be explained by these differences (Hox 2010). This is an indication that interviewers did not deal with comparable groups of respondents. By adding these variables into the model we can control for at least part of the differential composition of the respondent groups interviewed by each interviewer. It is nevertheless possible that after controlling for these variables there are still differences between the respondent groups. However, we assume that these remaining differences are not, or are only weakly, related to the interview length.

After controlling for the aforementioned relevant respondent characteristics, new estimates can be used in order to calculate a new intra-interviewer correlation: $\rho_{int|c}$:

$$\rho_{int|c} = \frac{\sigma^2_{uI|c}}{\sigma^2_{u|c} + \sigma^2_{e|c}}$$

The formula shows that the intra-interviewer correlation is the proportion of total variance of a variable explained by the interviewer after controlling for respondent characteristics. Table 3 shows the intra-interviewer correlations for each country (R5). These are the proportion of variance in interview length explained by the differences between interviewers, after controlling for respondent characteristics. The differences between countries in terms of the intra-interviewer correlations are considerable. For the Czech Republic, the differences between interviewers explain 15% of the variance in interview length, whilst for Israel this figure is 62%. High percentages (>50%) are also observed for Hungary, Russia and Poland. Unlike these countries, interviewers have less impact (<25%) in Portugal, Finland and Sweden. Although the differential composition of respondent groups may still be responsible for part of the differences between interviewers, these results clearly indicate different fieldwork activities in each country and differences between interviewers within countries. To validate the results obtained by using data from ESS round 5, the same intra-interviewer correlations were calculated for data from round 4 (R4, control variables: age, education, number of not applicable questions and asking for clarification). In 14 countries the same fieldwork organization was involved (another organization in Belgium, Bulgaria, Germany, France and United Kingdom). For round 4, there are also large differences between countries and the differences between interviewers are significant in all countries. In 12 countries the intra-interviewer correlations (IIC) is higher in round 4 than in round 5. In countries with the same fieldwork organization there are considerable differences (increase or decrease) between both rounds. Despite

these differences, the results for round 4 support the general trend in the results for round 5: differences between countries and differences between interviewers within countries.

It is notable that although interviewer effects were not expected, and whilst effects from the respondent characteristics in Table 2 were expected, in all countries the differences between interviewers explain much more of the variability in interview length than do the respondent characteristics presented in Table 2. The intra-interviewer correlation and the mean number of interviews per interviewer (m) are the two components of the interviewer design effect [$\text{def}_{int} = 1 + \rho_{int}(m - 1)$]. The interviewer design effect is the extent to which the variance in the sample mean of a simple random sample is increased due to interviewer variability. In all countries, the variance inflation for the variance of the sample mean of interview length is substantial. This means that the estimate of the mean interview length in each country is less precise than intended.

There are only two available interviewer characteristics to explain the observed interviewer variability: the interviewer's workload and the order of the interviews. For both variables, a negative effect on interview length is expected. Following Olson and Peytchev (2007), the interviewer order is expressed as a sequential number of the interviews conducted by each interviewer and encompasses the interviewer's experience over the field period of a survey. The interviewer's workload is the total number of realized interviews in the fifth round of the ESS. In most countries (14 out of 19), interviewers are paid per completed interview.[1] It can be assumed that in these countries, interviewers try to optimize their cost-benefit analysis by spending less time on each interview so that more can be completed. It should be noted that no other information about the interviewer's workload is available. In some countries, interviewers may work on more than one survey and the ESS related workload is not necessarily an interviewer's whole workload.

The coefficients presented in Table 3 are the fixed partial effects of interview order and the interviewer's workload, added to the model including respondent characteristics. Nearly all the fixed effects of interview order and interviewer's workload are negative, 14 effects of interview order and 8 effects of workload are statistically significant (p<0.05). This general pattern of fixed effects of these two interviewer characteristics mainly supports expectations. Interviewers proceed through an interview faster when they have already carried out other interviews (negative effect of order) and interviewers with a higher workload interview faster (negative effect of workload). There is no clear evidence that payment schemes for interviewers have any impact on the effect of workload.

Table 3 also illustrates the proportion of variance at the interviewer level explained by interviewer characteristics in addition to respondent characteristics. This shows the difference between interviewer-level residual variance (σ^2_{ulc}) in the model with and without the interviewer characteristics, proportional to the interviewer-related variance in the model without these characteristics. Once again, the differences between countries are remarkable. In Belgium, Switzerland,

Spain and Slovenia both characteristics explain more than 10%; in Germany, Denmark, Poland, Russia and Sweden the proportions are in the range between 5% and 10%. In some countries we get an unexpected negative proportion. Snijders and Bosker (1994) explain why adding group-level variables to the model can have this result. The main reason is that adding group-level variables to the model can increase the estimate of the residual variance. In this situation, the proportional reduction in variance can be negative.

The next step in the analysis of interviewer effects on interview length is to evaluate the impact of the observed differences in interview length between interviewers on the data quality. Do interviewers with different mean interview lengths obtain data of differing quality? Because it is not the intention of this paper to answer this important question in depth, we only present the results of a first, basic analysis. This is limited to a descriptive analysis of the relationship between the mean speed of interviewing and the percentage of 'Don't Know' answers during the interview. For this analysis, the mean interview length for each interviewer was calculated. Based on the quantiles of this distribution in each country, five groups of interviewers were distinguished: from slow (quantile 5 = 20% of interviewers with the highest mean interview length), to fast (quantile 1 = 20% of interviewers with the lowest mean interview length). This variable (speed of interviewing) categorizes the differences in mean interview length between interviewers and was used as an independent variable in a one-way analysis of variance. The dependent variable in this analysis of variance was the percentage of 'Don't Know' answers (DK) during an interview (100× [number of questions with DK answers/total number of questions with a substantive answers, a refusal or a don't know answer]). The percentage of DK answers is considered as a general data quality indicator and a high percentage is an indication of respondent satisficing behaviour. The results in Table 4 describe the relationship between the percentage of 'Don't Know' answers and the speed of interviewing. Although the proportions of explained variance are small, there is a significant relationship in most countries. This means that there are significant differences between the five 'speed of interviewing' groups with regard to the mean percentages of 'Don't Know' answers. There seems to be no clear (linear) pattern in this relationship. Once again the differences between countries are striking. There is no country with a clear positive linear relationship. The most common trend is a U-shaped relationship (e.g. Switzerland, Germany, Denmark and Spain). In some countries there is a negative trend (e.g. Bulgaria, Hungary and Russia). However, whilst these results illustrate the potential impact of the differences between interviewers in speed of interviewing on the data quality, it is clear that further analysis is needed.

---

[1] In Finland, Norway and Spain interviewers are paid per hour. In Sweden they have a regular fixed salary and in Switzerland there is a combination of 'paid per hour' and 'paid per completed interview'.

Table 3: Evaluation of interviewer effects on interview length by country

| | n interviewers | Mean workload | Intra interviewer correlation | | def_{int} | R-square Int. | Fixed effect order | Fixed effect workload |
			R5	R4				
Belgium	126	13.52	0.41	0.49	6.11	0.11	-0.10*	-0.27*
Bulgaria	234	10.40	0.42	0.70	4.99	0.01	0.06	-0.28*
Switzerland	57	26.42	0.42	0.60	11.69	0.23	-0.03	-0.24*
Czech Republic	442	5.40	0.15	0.74	1.64	0.01	-1.22*	1.76*
Germany	193	15.70	0.42	0.13	7.10	0.05	-0.30*	-0.15
Denmark	90	17.53	0.30	0.69	6.04	0.08	-0.22*	-0.24
Spain	67	28.13	0.31	0.32	9.48	0.11	-0.18*	-0.20*
Finland	128	14.70	0.20	0.64	3.77	-0.02	-0.25*	0.09
France	151	11.44	0.43	0.17	5.53	0.03	-0.13*	-0.23
United Kingdom	172	14.08	0.46	0.21	7.04	0.03	-0.45*	-0.09
Hungary	183	8.53	0.60	0.76	5.49	0.01	-0.15*	-0.14
Israel	93	24.67	0.62	0.89	15.75	0.01	-0.05	-0.15
Netherlands	158	11.58	0.41	0.30	5.33	0.01	-0.11	-0.22
Norway	106	14.60	0.34	0.67	5.57	-0.01	-0.14*	0.02
Poland	170	10.30	0.55	0.38	6.16	0.06	-0.30*	-0.40*
Portugal	75	28.67	0.17	0.32	5.79	-0.01	0.07*	-0.06
Russia	344	7.54	0.59	0.24	4.87	0.05	-0.25*	-0.37*
Sweden	123	12.20	0.24	0.48	3.72	0.05	-0.11	-0.25
Slovenia	67	20.94	0.40	0.87	9.02	0.16	-0.16*	-0.25*

* $p<0.05$

5 Conclusion and discussion

Although the basic question raised in this article is rather simple, the answer to the question 'How long will it take?' is more complex and starts with: 'It depends'. Much variability has been observed in interview length and the results make clear that the interview length in the fifth round of the ESS depends on the country in which the survey was organized, the situation of the respondent and in particular, who carried out the interview. The results make clear that it is relevant and necessary to investigate interview length and the impact on data quality resulting from differences in the speed of interviewing between interviewers.

Although no differences between countries and interviewers were expected, the observed differences are considerable and are an indication that the standardization of fieldwork activities in general and the basic principle of standardized interviewing in a cross-national survey in particular have not been optimally implemented.

In some countries (especially those using PAPI for data collection) the measurement of interview length is imprecise. The observed variance in interview length between countries seems to be an indication that in each country there is an 'own survey' culture, with specific practices (e.g. the selection and training of interviewers, monitoring and control of fieldwork activities, etc.). Apparently the central specifications for participating countries in the ESS are not sufficient to eliminate a country-specific approach to fieldwork activities. The differences between countries are much more important than the differences between languages. More detailed fieldwork monitoring in each country is necessary in order to understand and explain these differences.

Some respondent characteristics have an impact on interview length. The results support the obvious assumptions that fewer applicable questions reduce the interview length, interviewing older respondents takes more time and the interview duration increases when a respondent more frequently asks for clarification. It is remarkable that in some countries these relevant respondent characteristics explain only a small proportion of the variation in interview duration. This lack of explanatory power (e.g. countries in which respondent characteristics can explain less than 5% of the variance in interview length) can be considered as a negative aspect of data quality and is further reason to evaluate in greater detail the data collection and processing procedures used in these countries.

In all countries, the differences between interviewers explain a significant and substantial part of the variance in interview length. Because respondents are not randomly assigned to interviewers, it is still possible that, after controlling for relevant respondent characteristics, part of this explained variation is caused by area effects and by the different composition of the respondent groups interviewed by each interviewer. However, it is very unlikely that these clustering effects are responsible for the entire variations in interview length. The order of the interview and the interviewer's workload both have a negative effect on interview length. The analysis of the interviewer variance in interview length can also be considered as a general and indirect evaluation of interviewer performance during fieldwork. Certainly some interviewers carry out their interviews quickly, whilst others are slow. In itself, this should not be a problem. However, it could be assumed that this difference in interviewer behaviour and deviation from the key principle of standardized

Table 4: Mean percentage of 'Don't Know' answers by speed of interviewing

	Slow 1	2	3	4	Fast 5	F	R-square
Belgium	0.51	0.36	0.41	0.31	0.37	3.37**	0.01
Bulgaria	2.09	2.31	2.10	1.89	1.65	5.22**	0.01
Switzerland	0.83	0.67	0.61	0.88	1.09	6.10**	0.02
Czech Republic	1.09	1.23	1.25	1.13	1.27	0.93	0.00
Germany	0.94	0.93	0.84	1.12	1.11	3.85**	0.01
Denmark	0.59	0.43	0.41	0.54	0.53	1.67	0.00
Spain	0.89	0.65	0.46	0.61	0.67	8.70**	0.02
Finland	0.52	0.35	0.38	0.49	0.46	1.25	0.00
France	0.43	0.42	0.26	0.39	0.45	4.68**	0.01
United Kingdom	1.42	1.32	2.13	1.47	2.07	12.72**	0.02
Hungary	1.36	1.28	1.17	1.09	1.04	2.47*	0.01
Israel	1.19	1.54	1.96	1.42	0.97	17.09**	0.03
Netherlands	0.66	0.56	0.60	0.57	0.73	1.52	0.00
Norway	0.23	0.31	0.39	0.39	0.18	2.80*	0.01
Poland	1.92	1.62	1.22	1.35	1.16	8.37**	0.02
Portugal	1.19	1.58	1.73	1.97	1.51	14.12**	0.03
Russia	2.55	2.69	2.44	2.09	1.93	8.58**	0.01
Sweden	0.80	0.76	0.66	0.58	0.80	1.26	0.00
Slovenia	1.23	1.16	1.03	1.23	2.05	9.87**	0.03

The header "Speed of interviewing" spans columns 1–5.

* $p<0.05$; ** $p<0.01$

interviewing has an effect on the quality of the obtained answers. The first results of the analysis of the effect of differences in speed of interviewing on the percentage of 'Don't Know' answers clearly illustrate the possible impact of these differences on data quality. However, the results make clear that more research is needed into the relationship between the speed of interviewing and the quality of the resulting data. In the meantime, it can be suggested that the observed differences in interview length are not ideal and that greater effort is needed in order to reduce the differences between interviewers and countries. Permanent investment in training, monitoring and follow-up of interviewers in each participating country is necessary. Based on the results, we recommend a systematic and detailed evaluation of the data quality from relatively short and long interviews, and fast and slow interviewers. It seems worthwhile to use the measurement of interview length as an auxiliary variable in the assessment of data quality.

References

Bassili, J. (1996). The "How" and "Why" of Response Latency Measurement in Survey Research. In N. Schwarz & S. Sudman (Eds.), *Answering Questions: Methodology for determining cognitive and communicative processes in survey research* (p. 319-346). San Francisco: Jossey-Bass Publishers.

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

Bryk, A. S., & Raudenbus, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park: Sage Publications.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. S. Leinhardt (Ed.), *Sociological methodology 1981* (p. 389-437). San Francisco: Josse-Bass.

Fowler, F., & Mangione, T. (1990). *Standardized survey interviewing: Minimizing interviewer-related error*. Beverly Hills, CA: Sage Publications.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. New York: Wiley.

Hox, J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research*, *22*, 300-318.

Hox, J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). New York: Routledge.

Japec, L. (2005). Interviewer burden and its effects on data quality in the Swedish part of the European Social Survey (ESS). In L. Japec (Ed.), *Quality Issues in Interview Surveys, Some Contributions.* Stockholm: Department of Statistics Stockholm University.

Olson, K., & Peytchev, A. (2007). Effect of interviewer experience on interviewer pace and interviewer attitude. *Public Opinion Quarterly*, *72*, 273-286.

Presser, S., Couper, M., Lessler, J., Martin, E., Martin, J., Rothgeb, J., et al. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly*, *68*, 109-130.

Snijders, T., & Bosker, R. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, *22*, 342-363.