8 Interview, Module, and Question Length in SHARE

Hendrik Jürges

8.1 Introduction

The length of an interview is a major concern in survey research in several respects. First, respondent compliance is generally better for short rather than long interviews. One of the main reasons given by respondents for lack of participation is that they do not have time, and one of the first questions about the interview asked by respondents is "how long does it take?". Second, longer surveys are more expensive than shorter surveys, because they are more tedious to be programmed and tested and because they use more of the interviewers' time.

Despite these concerns, the SHARE interview is long. Considering the breadth of the topics involved in ageing, there was no way to make SHARE a short interview. Following experiences from HRS and ELSA, we decided to set a target average of 80 minutes for a single household interview and a target of 120 minutes for a couple interview. As is shown below, we actually came quite close to that target in most countries. However, this target was not easy to reach. Between each preparatory step (UK Pilot, All-country Pilot, and Pretest, see Chapter 2), the instrument was shortened by a considerable margin by eliminating a large number of questions across all fields and disciplines.

8.2 Measuring interview length using Blaise audit trails

There are two ways at our disposal to measure interview length in SHARE. One way involves the use of conventional timestamps, recording the system time of the laptop at the beginning and end of each module. A second way is to use data from Blaise audit trail files, in everyday language called "keystroke files". In these files, Blaise records for each interview each single keystroke made by the interviewer together with the exact time when it was made. With the help of these files it is not only possible to reconstruct the course of entire interviews, it is also possible to analyse the time interviewers spend on each single question. For the purpose of the present analysis, the Blaise audit trail data was converted by CentERdata to Stata and SPSS files containing the length in seconds spend on each questions asked in SHARE 2004. The data are not yet available with release 1 but we intend to make available in future releases. As is illustrated below, they will provide potentially interesting data for researchers interested in survey methodology.

Besides the unprecedented amount of detail provided by Blaise audit trails, the keystroke files also have the advantage of allowing a more precise measurement of overall interview length. In case of computer breakdown, only single questions instead of whole modules or even interviews have missing or corrupt time stamps. Keystroke files also provide a valuable diagnostic tool. First, they allow to identify questions which take an unusual long time to be answered. This can provide hard evidence for potential difficulties respondents face when answering particular questions. Second, they allow to detect questions or interviews in which interviewers have not followed protocol.

It should be noted that interviewers and respondents sometimes report considerably longer interview lengths than what is measured by the keystroke data. There are several reasons for this divergence. First, interview time felt might be different from actual time spend. Second, and more importantly, the keystroke files

do only measure the actual interview length, which does not include time spent on gaining entrance in the respondent's home, setting up the computer, closing down the computer, suspending the interview, etc. Third, we do not measure the time to complete the drop-off questionnaire. Although SHARE did require interviewers to assist respondents only when they had difficulty completing the drop-off, interviewers often waited at the respondents' homes to collect the completed drop-off. Evidence from the pilots suggests that this adds another 15 to 20 minutes to the average interview length.

8.3 Total interview length

The distribution of total interview length by number of person interviews in a household is shown in Table 8.1. The overall average length was 66.6 minutes in a single respondent household and 109.9 minutes in a multi-respondent household (two, three, and four respondent interviews are collapsed into one category; three and four respondent interviews were actually very rare). The overall average is thus below our target of 80 and 120 minutes, respectively. Table 8.1 also shows that there is considerable divergence both within and between countries. In France, Greece, and Switzerland, we were actually very close to our target. The shortest interviews were made in Austria, Spain, and Italy. In fact, more than one fourth of all single-respondent interviews in these countries were shorter than 40 minutes, i.e. shorter than half of our target time. The longest interviews were conducted in Denmark and Sweden. Nearly 25 percent of all single-respondent interviews took more than 100 minutes in these two countries.

Table 8.1 Household interview length in minutes, by country and number of interviews

	Single Interviews				Multiple Interviews			
Country	Mean	Q25	Median	Q75	Mean	Q25	Median	Q75
Austria	47.9	31.8	44.4	61.8	74.6	52.7	68.5	92.3
Switzerland	77.7	58.3	72.2	94.3	121.0	88.5	115.8	148.6
Germany	63.3	49.5	60.7	75.2	105.0	83.3	104.1	122.2
Denmark	84.8	70.2	83.9	97.9	141.5	119.3	140.1	163.9
Spain	47.1	34.3	45.0	57.6	81.7	58.9	79.5	101.0
France	77.3	62.0	76.8	91.6	120.9	94.7	119.7	142.0
Greece	77.7	61.3	75.3	91.5	121.0	94.6	120.2	142.1
Netherlands	69.9	53.4	68.5	85.7	110.4	85.2	106.1	132.9
Italy	53.4	39.9	51.4	64.8	88.5	67.7	83.6	106.5
Sweden	84.9	64.3	78.8	99.8	133.1	105.1	127.5	155.3
Total	66.6	47.4	64.1	82.9	109.9	81.7	106.7	134.2

In the following table, we show how interview length varies with respondent age. In a multi-respondent household, age is defined as the age of the primary respondent, or where no primary respondent is defined (such as in France), the age of the first respondent. The overall pattern for all types of households is slightly U-shaped. For instance, interviews in multiple respondent households in the youngest age group (50-54) and in the oldest age group (80 and over) lasted on average 7 to 8 minutes longer than interviews in the middle age groups (60 to 69). We see the result of two counteracting forces: At younger ages, in particular when still working, respondents go through detailed questions in the employment and pensions module, but take less time on health questions. Older respondents take more time on health questions.

Moreover, the lower age limit for completing the walking speed test was 75 years, adding a little less than 3 minutes on average to each individual interview.

Table 8.2 Household interview length in minutes, by age group and number of interviews

Age Group	Single respondents	Multiple respondents
50-54	66.2	115.4
55-59	68.8	111.3
60-64	65.1	106.9
65-69	64.6	107.2
70-74	64.4	109.0
75-79	68.2	110.4
80+	70.0	114.8

8.4 Module length

The average length of each SHARE 2004 module is shown, by country, in Table 8.3. The longest module overall is Employment & Pensions (EP). The module took on average 8.9 minutes to complete, again with considerable variation across countries. Danes needed on average 14.5 minutes, while Spaniards only needed 5.2 minutes (note, however, that these differences are unconditional, ie. differences in labour force participation are not controlled for). The second longest module was Physical Health (PH) with on average 6.9 minutes. Cross-national variation is smaller than in the EP module. It took Austrians only 4.8 minutes to complete this module, whereas Greek respondents needed on average 8.6 minutes.

Table 8.3 Module length in minutes, by country

Module	AT	CH	DE	DK	ES	FR	GR	IT	NL	SE	Total
Coverscreen	2.2	1.0	2.5	2.7	2.6	1.6	4.1	2.3	2.7	2.8	2.4
Cover- Main	0.3	0.6	0.4	0.6	0.2	0.7	0.7	0.2	0.5	0.4	0.4
Demographics	2.6	4.3	3.3	3.9	2.5	3.5	4.1	2.7	2.9	3.9	3.3
Physical Health	4.8	6.9	6.6	8.3	5.9	7.7	8.6	6.3	6.1	8.1	6.9
Behavioral Risks	1.3	2.4	1.9	3.0	1.2	2.4	2.4	1.3	2.1	2.9	2.1
Cognitive Function	4.8	7.4	6.3	7.3	5.1	6.6	7.3	5.7	6.6	6.9	6.3
Mental Health	1.9	3.3	2.8	3.3	2.3	3.6	3.5	2.8	2.8	3.5	3.0
Health Care Use	3.3	5.5	4.7	5.6	3.4	6.5	5.8	3.8	5.7	5.1	4.9
Employment/Pens.	5.9	11.5	8.3	14.5	5.2	10.0	8.9	6.0	8.8	12.3	8.9
Grip Strength	1.3	2.4	2.0	2.5	1.6	2.1	2.5	1.8	2.2	2.3	2.0
Walking Speed	1.6	3.4	2.2	3.6	2.2	3.6	3.1	2.1	3.2	2.5	2.6
Children	3.5	4.9	3.9	5.8	3.9	4.8	4.1	3.5	5.0	5.9	4.5
Social Support	1.3	2.2	1.9	3.0	1.2	2.2	2.1	1.4	2.2	2.6	2.0
Financial Transfers	1.3	2.3	2.0	2.2	1.0	2.1	2.6	1.3	1.6	2.5	1.8
Housing	2.2	3.6	2.7	4.1	2.6	3.7	3.6	2.5	3.2	3.9	3.2
Household Income	0.6	0.7	0.7	0.8	0.5	0.8	0.8	0.6	0.6	0.7	0.7
Consumption	2.0	3.5	2.7	3.7	1.9	4.2	2.6	2.4	3.8	5.0	3.2
Assets	2.2	3.4	3.2	4.8	1.7	4.4	2.7	1.8	3.3	5.9	3.3
Activities	1.0	1.9	1.2	2.1	0.5	1.5	1.3	0.7	1.6	2.1	1.4
Expectations	2.6	5.3	4.2	4.6	2.8	5.1	4.7	3.6	4.3	4.9	4.1
Interviewer	1.0	1.7	1.1	1.5	1.3	1.8	1.4	1.5	1.2	1.2	1.3

The shortest module was CM (Coverscreen (Main); 0.4 minutes), a module that sets up the structure of the interview in multi-respondent household by determining who is the financial, family, and housing respondent. The module also sets a couple of parameters internally for all households (not visible to the interviewer), such as the interview country and whether pre-Euro options are called for amount questions.

There is another module shorter than one minute, which is Household Income (HH), consisting of a few questions regarding the income of household members who were not part of the interview.

An initial concern was the length of the physical measurements, grip strength (GS) and walking speed (WS). Most professional interviewers are not used to make such measurements and great care was taken to train them properly on protocols. As it turned out, measurements were indeed feasible and added little to the overall respondent burden (in fact, respondents and interviewers often expressed relief about the fact that the strenuous question-answer routine was interrupted). The grip strength test took on average 2 minutes to be completed and gait speed was completed in 2.6 minutes on average.

Table 8.4 Length of PH (physical health) module in minutes, by self-reported health and country

	Self-reported health				
Country	Very good	Good	Fair	Poor	Very Poor
Austria	3.4	4.5	5.6	6.2	7.0
Switzerland	5.7	6.5	9.5	10.9	10.8
Germany	5.0	5.9	7.5	7.7	9.0
Denmark	6.3	7.8	10.1	11.4	12.5
Spain	4.0	4.8	6.7	7.5	7.9
France	5.5	7.0	9.3	9.9	11.1
Greece	6.9	8.7	9.2	9.9	10.0
Italy	4.7	5.4	6.9	8.2	8.7
Netherlands	4.7	5.7	7.6	8.6	8.4
Sweden	5.8	7.7	9.5	10.8	12.6
Total	5.4	6.3	7.9	8.7	9.7

Above, we saw that total interview length is U-shaped in age. One reason is that older respondents need more time to answer health questions. Table 8.4 shows the length of the physical health module by self-reported health status. Respondents who report to be in very good health need on average 5.4 minutes to complete this module – again with some cross-national variation. In contrast, respondents in very poor health needed nearly 10 minutes to complete the module.

8.5 Question length

When researchers design a new questionnaire, they estimate the length of the interview by rule of thumb: on average four answers (or "ticks") fit in one minute. In fact, this rule of thumb proved to be quite accurate, and it has given us valuable guidance during the development process. Table 8.5 shows the question length in SHARE, by module. Overall, the keystroke data contain information on more than 6 million questions asked. Across all modules, the average length of a question was 13 seconds, or put the other way around, we asked 4.6 questions per minute.

The average question length in a module also gives some indication of the difficulty of a module, because difficult questions need more time to be asked (and answered, for that matter). The shortest average tick length is found in the Interviewer Observations module (6.2 seconds). In this module, interviewers record observations during the interview, e.g. respondents' willingness to cooperate or interference of third persons. This module does not involve the respondent, so that tick length in this module measures the time to read and answer a question without a respondent, providing some lower bound for the question length in a personal interview.

Table 8.5 Average Tick Length in Seconds, by Module

Module	Mean	StdDev	N
Coverscreen	8.8	17.4	305,876
Cover-to-Main	14.1	20.4	17,744
Demographics	9.3	12.8	475,039
Physical Health	16.0	22.0	580,518
Behavioral Risks	13.2	13.8	209,196
Cognitive Function	21.8	27.8	388,111
Mental Health	10.8	10.9	356,340
Health Care Use	14.2	19.2	454,659
Employment	12.9	16.8	910,314
Grip Strength	16.1	32.4	168,356
Walking Speed	17.8	41.1	40,359
Children	9.4	10.1	454,046
Social Support	12.0	12.0	219,081
Financial Transfers	11.5	12.1	157,192
Housing	11.5	14.1	251,214
Household Income	11.3	15.4	52,633
Consumption	23.5	33.6	123,373
Assets	14.3	15.8	128,683
Activities	13.6	17.8	240,381
Expectations	14.9	24.0	368,697
Interviewer Observations	6.2	10.9	280,582
Total	13.0	19.1	6,182,394

Table 8.6 lists the 15 questions with the longest average length. The longest questions were CF010 and CF008. This was to be expected because of the nature of the questions. CF010 is the verbal fluency test. The Blaise application gives each respondent exactly 60 seconds to name as many different animals as possible, setting a lower physical limit of one minute to the question length. CF008 is the question that asks respondents to recall as many words as possible from a list of ten words. Here, interviewers were instructed to allow up to one minute for recall.

Table 8.6 Distribution of tick times - the 15 longest questions

Question	Mean	Q10	Q25	Q50	Q75	Q90	N
CF010	80	68	70	74	82	96	21,912
CF008	74	43	54	68	87	111	21,952
CO005	51	6	13	32	64	112	15,630
PH006	49	13	24	39	62	95	22,305
GS001	49	4	9	26	74	127	22,243
PH048	48	11	22	40	65	96	22,294
CF016	46	15	27	42	59	81	21,936
CO002	41	10	19	33	52	81	15,607
PH010	41	11	20	33	52	82	22,302
HC002	39	12	20	31	48	73	22,260
EP078	39	6	12	23	46	88	17,963
GS006	36	4	5	15	51	101	20,350
CV001	35	2	3	11	53	86	19,140
PH011	34	7	13	25	44	72	22,298
EX024	31	2	3	16	39	73	22,217

Another question that took long to be answered was CO005 (51 seconds on average): "*Please look at card 31.Thinking about the last 12 months: about how much did your household spend in a typical month on all goods and services, including groceries, eating out, telephone and everything else?*" This sometimes called "one-shot consumption question" obviously placed a large burden on many respondents. Roughly half of them needed more than half a minute, and one quarter even needed a minute or more to think about the question and answer to it. However, as remarkable as the average length is the wide dispersion. The fastest 10 percent needed 6 seconds or less (possibly including a considerable proportion of "don't know").

As a final example, it took respondents on average 49 seconds to answer PH006. Again, this is not surprising. PH006 asks for diagnosed diseases to be chosen from a list of 14 different conditions.

8.6 Summary

This chapter gives some basic assessment of the net length of a typical SHARE interview. The overall average was 67 minutes for a single-respondent household and 110 minutes for a multi-respondent household. Despite the complexity of the survey instrument, we managed to keep respondent burden within reasonable limits. During the development process of SHARE, considerable effort has been made to stay within these limits. In fact, the first questionnaire draft tested in the UK in September 2002 took about 100 and 150 minutes, respectively, despite the fact that we randomly dropped a few modules from each interview. The necessary redesign of the questionnaire that followed was considered painful by most of the researchers involved. Each of them saw exciting questions being taken away from the questionnaire. However, they also often cleverly restructured their respective modules, leading to the rather lean and focused final version of the survey instrument.

The chapter also demonstrated the use of some particularly valuable information recorded by the Blaise CAPI instruments: audit trail files. With the help of these files it is possible to analyse modules and even single question in great detail, for instance by using the time needed to answer individual questions as an objective measure of their difficulty. This provides valuable complementary information to the interviewer reports at debriefings.