THE UNIVERSITY OF MELBOURNE

FACULTY OF BUSINESS & ECONOMICS

Melbourne Institute Working Paper Series

Working Paper No. 10/12

The Impact of Computer-Assisted Interviewing on Interview Length

Nicole Watson and Roger Wilkins

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

The Impact of Computer-Assisted Interviewing on Interview Length*

Nicole Watson and Roger Wilkins
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Melbourne Institute Working Paper No. 10/12

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 978-0-7340-4270-5

May 2012

* The research reported on in this paper was supported by an Australian Research Council Discovery Grant (#DP1095497). This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and view reported in this paper, however, are those of the authors and should not be attributed to FaHCSIA or the Melbourne Institute.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

Computer-assisted personal interviewing (CAPI) offers many attractive benefits over paper and pencil interviewing. There is, however, mixed evidence on the impact of CAPI on interview *length*, an important survey outcome in the context of length limits imposed by survey budgets and concerns over respondent burden. In this paper, recent data from a large, nationally representative panel study is used to investigate CAPI's impact on interview length. A key feature of our analysis is that, through use of both experimental and quasi-experimental evidence, we examine the roles played by specific factors which, while typically associated with CAPI, vary in their extent and nature from study to study. We find that effects very much depend on how CAPI is implemented: the hardware and software adopted, the extent and nature of the dependent data introduced, and even interviewer workloads, can all have large influences on the CAPI impact—a finding that helps explain the conflicting results from previous studies. Overall, we conclude that, absent dependent data, CAPI will almost certainly increase interview lengths. However, the potential reductions in interview length from dependent data are very large, such that even modest levels of dependent data can lead to net reductions in interview lengths.

Keywords: Computer-assisted methods, interview length decomposition, dependent data, interviewer composition, learning effects

1. Introduction

Computer-assisted personal interviewing (CAPI) offers many attractive advantages over paper and pencil interviewing (PAPI) and most large ongoing face-to-face surveys using PAPI methods are likely to eventually shift to CAPI. Its advantages include improved and more complex routing, automated checking of inconsistencies in the data with the respondent as they occur, simultaneous data entry, reduction in lags between interviews and data delivery, and capture of system information, such as section timestamps, during the interview (de Leeuw, Hox and Snijkers, 1995). Further, in panel surveys, CAPI facilitates much greater use of data collected in previous waves, thereby improving data quality (Jäckle, 2008). Despite these apparent advantages, it is important to have a good understanding of the impacts that a move from PAPI to CAPI has on survey outcomes, both to enable survey managers to make appropriate implementation and budgeting decisions, and to ensure appropriate interpretation of the resulting data.[1]

Relatively few (published) studies have investigated the impacts of CAPI on survey outcomes, although there has emerged a consensus view on the effects of the introduction of CAPI on four aspects of data quality. First, a change from PAPI to CAPI does not significantly affect response rates (Martin et al., 1993; Baker, et al. 1995; Lynn, 1998; Laurie, 2003; Nicoletti and Peracchi, 2003; Schräpler et al., 2006; Watson and Wilkins, 2011) or attrition rates (Martin et al., 1993; Schräpler et al., 2006). Second, the vast majority of respondents are ambivalent about the change in mode and very few negative reactions are recorded (de Leeuw, et al., 1995; Martin et al. 1993; Watson and Wilkins, 2011). Third, interviewers are generally positive about the change and appreciate the more professional look it gives them (de Leeuw, et al., 1995; Banks and Laurie, 2000; Watson and Wilkins, 2011). Finally, almost all previous studies recorded a lower rate of missing data in CAPI compared to PAPI due to the elimination of routing errors (for example, de Leeuw, 1995; Laurie, 2003; Schräpler, et al. 2006; Watson and Wilkins, 2011).

There is considerably less evidence, and less consensus, on the impact of CAPI on other survey outcomes, including the proportion of 'don't know' or 'refused' responses, the length of responses to open-ended questions, and interview lengths. In this paper, we focus on the issue of interview length and how it is affected by survey mode, use of dependent data and interviewer experience. The presumption is that, all else equal, reductions in interview length are desirable, reducing survey costs and respondent burden, or making it possible to increase survey content for a given budget.[2] Understanding how CAPI can affect interview lengths is therefore clearly important.

The limited number of existing studies of the impact CAPI has on interview length provide conflicting evidence. Martin et al. (1993) and Fuchs et al. (2000) found interviews took 12 to 17 per cent longer to administer via CAPI, while Baker et al. (1995), Lynn (1998) and Jäckle (2008) each present evidence of a 16 to 20 per cent decrease in interview length. Martin et al. (1993) also reports on another study finding no evidence of a difference in interview lengths. The reasons suggested for longer durations include slower speed of typing compared with writing, differences in 'looping' strategies for repeated question sequences (for example, for questions in respect of each child in the household) (Fuchs et al, 2000), and interviewer inexperience with CAPI (Martin et al., 1993). The reasons given for reducing interview time

[1] Indeed, it is also valuable to understand CAPI-PAPI differences in survey outcomes for a *new* study.

[2] Of course, in principle, reductions in time spent on each question may have adverse effects on data quality, and so this also needs to be assessed. Indeed, in a separate paper (Watson and Wilkins, 2012), we investigate the effects of the introduction of CAPI on data quality in the HILDA Survey and find no adverse effects.

under CAPI include automation of the use of dependent data that had previously been provided on paper (Baker et al., 1995) and greater use of dependent data in a proactive fashion (Jäckle, 2008). The implication from these findings is that implementation details, such as type of computer interface, looping strategies, interviewer training and use of dependent data, all affect the impact of CAPI on interview length. However, little is known based on these studies about the relative importance of these factors.

There is also very little known about how quickly interviewers adapt to new technology. Indeed, Baker et al. (1995) and Caviglia-Harris et al. (2012) are possibly the only two studies to have examined differences in learning effects between CAPI and PAPI. Baker et al. (1995) found that CAPI interview lengths continued to decline over time, while PAPI interview lengths were relatively stable. Caviglia-Harris et al. (2012), by contrast, found there were no differences in learning effects between modes, although they were comparing surveys with different content conducted at different time points. An important limitation of both studies is that they considered interview duration as a function of the length of time in the field, whereas it is the number of interviews conducted by the interviewer that is likely to be the more relevant factor for capturing learning effects.

To investigate the impacts of CAPI on interview length, we draw on new evidence from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, a large-scale nationally representative longitudinal survey conducted in the market research arena. This is perhaps the only evaluation since the late-1990s to investigate the effect of CAPI on interview lengths, and therefore provides new evidence relevant to the contemporary context.[3] In particular, the use of computers and information technology, and the nature of the technology itself, have changed considerably since the earlier evaluations. Furthermore, since the HILDA Survey is longitudinal, we can also consider the impact of 'feed forward' (dependent) data, which can be a very important feature associated with the introduction of CAPI, and which cannot be investigated in cross-sectional surveys.

The HILDA Survey is a valuable source for investigating the extent and nature of the effects of CAPI on interview lengths, providing both experimental and quasi-experimental evidence. Specifically, we use data from an experiment conducted on the longitudinal 'dress rehearsal' sample in Wave 7 (2007), together with data from Waves 6, 9 and 10 (conducted in 2006, 2009 and 2010, respectively). The Wave 7 dress rehearsal sample, which comprises approximately 1,300 respondents (one-tenth the size of the main sample), had approximately half of the respondents assigned to CAPI and the other half assigned to PAPI, facilitating comparison of CAPI to PAPI within an experimental setting. Waves 6, 9 and 10 provide further evidence on the impact of CAPI in the form of pre- and post-CAPI data. Waves 6 and 10 provide a quasi-experiment by virtue of having almost identical questionnaires. Similarly, it is also possible to identify the effects of dependent data through comparisons of questionnaire sections in Waves 9 and 10 which have the same content, but which had additional dependent data introduced in Wave 10. Both the experimental data provided by the Wave 7 dress rehearsal and the quasi-experimental data provided by Waves 6 and 10 are furthermore used to investigate how the 'learning effects' of interviewers—as captured by reductions in interview lengths—differ across the two survey modes.

The paper proceeds as follows. The potential effects of CAPI and dependent data on interview length are canvassed in Section 2, before describing the data and research design in Section 3. Section 4 examines the impact of CAPI on overall interview length, while Section

[3] As noted, Caviglia-Harris et al. (2012) compare two waves with different questionnaire content and therefore are not able to ascertain the impact of CAPI on interview lengths.

5 considers impacts on individual questionnaire sections, and is thereby able to provide insights on the effects of specific mechanisms by which CAPI can affect interview durations—most notably, the use of dependent data. We consider differential interviewer learning effects by mode in Section 6. In Section7, we draw together the findings of Sections 4 to 6 by presenting a decomposition of the total change in mean interview length between Waves 6 and 10, estimating the contributions made by the introduction of dependent data, changes to interviewer composition and interviewer learning effects, and 'other' factors associated with the introduction of CAPI. Section 8 concludes.

2. Potential impacts of CAPI on interview length

As the existing mixed evidence on the effects of CAPI on interview length would suggest, impacts depend on the way in which CAPI is implemented. One potentially important source of impacts on interview times is the ability to include dependent data—that is, where pre-existing information on respondents is used (fed forward). This dependent data can be used in either a proactive or reactive fashion. Proactive dependent data is used to tailor questions based on known characteristics of respondents, such as sex, age and place of birth, or to remind the respondent of a previous answer or ask them if they are still in the same situation. Reactive dependent data is used to confirm a change after they have reported a different status from that reported in a previous wave. Proactive dependent interviewing is likely to reduce interview times, since it removes the need to answer some questions and also lowers respondent cognitive burden, making some questions easier to answer. Reactive dependent interviewing is likely to increase interview length, since additional questions are asked if there is an inconsistency (Jäckle, 2008).

In addition to proactive dependent interviewing, several other factors typically present under CAPI are likely to reduce interview times, including automation of questionnaire routing and automatic generation of context-dependent text. Automated routing means that interviewers do not need to work out the relevant skips, which can also improve data quality, while automated context-dependent text (such as a child's name or text appropriate to a person with two or more jobs) mean that interviewers do not have to work out what text they should use for specific situations.

While proactive dependent data, automated routing and automated context-dependent text would seem to make CAPI clearly faster, several factors potentially contribute to longer interview lengths in CAPI than in PAPI. First, inconsistent or unusually high or low answers can be queried (and resolved) under CAPI. Such queries, which may result in the interviewer being required to either type in an explanation for the response or work back through the questionnaire to correct some previously entered data, will increase interview times. Second, a CAPI system will not progress to the next question until the current question is correctly filled in, which may slow interviewers down (although presumably this improves data quality). Third, in PAPI, experienced interviewers often read ahead and are asking the next question of the respondent while they are finishing off filling in the answer to the last question. They cannot do this in CAPI. Fourth, many interviewers can write faster than they can type, although the increase in interview lengths due to this factor will depend on both the manner of entry (for example, stylus or keyboard) and the number of questions requiring written text for responses. Finally, the need for interviewers to learn and become comfortable with the CAPI technology will tend to increase interview length, at least in the first wave that an interviewer uses CAPI.

There are, furthermore, other factors that have ambiguous implications for interview lengths, such as how many questions are displayed on each screen, how questions in grids on the paper questionnaire are programmed into CAPI, the type of computer device used, the

manner of entry of responses (for example, keyboard or touch screen), and indeed the nature of the survey.

It is thus unclear *ex ante* whether the interview lengths for CAPI should be shorter or longer than PAPI, with the outcome depending on a variety of factors that will vary from survey to survey. It is perhaps not surprising, therefore, that previous studies have had mixed results regarding the impact of CAPI on interview length: findings will depend on the net effect of the particular combination of factors present in the study.

3. Data and evaluation strategy

The HILDA Survey is a household-based panel study which collects information on families, incomes and labour market activities on an annual basis (Wooden and Watson, 2007). The survey began in 2001 with nearly 7,700 household interviews and 14,000 individual interviews from a representative sample of people living in private dwellings. This wave 1 panel is followed over time and interviews are conducted with all members of the household aged 15 and over. The majority of the questionnaire content is stable over time, although special questionnaire modules are included in each wave. For the first eight waves, the interviewer administered questionnaires were completed using PAPI, but from Wave 9 CAPI has been used. In anticipation of the shift to CAPI, an experiment was conducted in 2007 using the longitudinal dress rehearsal sample, which is one-tenth the size of the main sample and is used to test survey content and methods prior to administration on the main sample.[4]

There are four survey instruments administered by the interviewer, plus a Self-Completion Questionnaire given to the respondent to complete themself. Household-level information is collected in the Household Form (HF) and Household Questionnaire (HQ). The HF establishes who is part of the household and collects some basic details of household members, while the HQ contains questions on child care arrangements and housing. Individual-level information is collected by personal interview with all persons in the household aged 15 years and over. The interview covers a wide range of topics, including education, employment, income, relationships, and health. The Continuing Person Questionnaire (CPQ) is used for those who have been interviewed previously and the New Person Questionnaire (NPQ) is completed by individuals when they are interviewed for the first time. The NPQ includes most of the same questions as the CPQ, but also asks for background information, such as place of birth and marital history.

For this analysis of interview lengths, we draw on data from Waves 6, 9 and 10 of the main sample together with the Wave 7 dress rehearsal sample. This combination of sources allows us to isolate the roles played by specific factors in determining interview length. First, data from Waves 6 and 10 can be used to conduct a 'before-after' analysis of the impact of CAPI, since Wave 6 was administered by PAPI and Wave 10 by CAPI, and both were very similar in terms of questionnaire content, as the rotating wealth module was included in both of these waves. Second, a split-sample test of CAPI was conducted in the Wave 7 dress rehearsal. This provided the same conditions for both samples in all aspects other than survey mode, allowing isolation of the impact of CAPI through comparison of interview lengths across the two modes. Finally, comparisons between Waves 9 and 10, which were both CAPI administered, provide information on the impact of dependent data on interview lengths. In

[4] In this experiment, the face-to-face interviewers, rather than households, were randomly assigned to either CAPI or PAPI delivery. Respondent assignment was nonetheless effectively random, and indeed Watson and Wilkins (2011) show that the demographic composition of the CAPI and PAPI samples is not significantly different.

particular, the amount of dependent data used was expanded between Waves 9 and 10 and, since all of the questionnaire sections were timed, durations of sections with equivalent content, but differences in dependent data, can be compared across the two waves.

Table 1 provides a summary of the survey mode, use of dependent data, special modules and timestamp information included in each wave. We have listed three modes in Table 1 because the particular CAPI setup in the Wave 7 dress rehearsal was different to that ultimately adopted in Waves 9 and 10. In the Wave 7 dress rehearsal, the interviewers used laptops with either an external mouse or an inbuilt touchpad. In Waves 9 and 10, the interviewers used a tablet and stylus.[5] However, the software used, Confirmit[6], remained the same, and indeed the scripts developed for the Wave 7 dress rehearsal formed the basis of the scripts used in Waves 9 and 10.

Use of dependent data has progressively increased across the four waves. A limited set of data was pre-printed on the Household Form for Waves 6 and 7—name, sex, date of birth and date of last interview. In the CAPI component of the Wave 7 dress rehearsal, two additional pieces of information were also brought forward from the last interview: whether the respondent was employed; and the number of employers they had. Wave 9 further carried forward the respondent's contact details (collected for tracking purposes), eliminating the need to re-enter details that had not changed since the previous interview. Wave 10 again extended the dependent data to include legal and social marital status at the time of last interview, whether the respondent owned a home previously, and the name, age, date of birth and sex of the respondent's own children.

The bottom panel of Table 1 shows that timestamp and interview duration information allows durations to be determined for every section of the HQ across all four waves, but the level of detail available varies somewhat for the PQs. In Wave 6, only the duration of the entire PQ is available. In Wave 7, timestamps enabled durations to be determined for the new sections introduced in Wave 7, but not for the existing sections. From Wave 9, durations are available for each individual section of the PQ. Further, where the interview was conducted on paper (in Waves 6 and the paper component of the Wave 7 dress rehearsal), we also have the interviewer's estimate of the overall length of each PQ and HQ to help validate the estimate obtained from the timestamps.

Examination of the timestamps revealed that, while most of the data appeared sensible, problems existed for some interviews under both PAPI and CAPI. When interviewers recorded the time on the paper questionnaire, they sometimes made mistakes with the 24-hour format they were required to use. It was also found that interviewer estimates of length had fewer extreme values than calculated using the timestamps. Under CAPI, timestamps sometimes generate unlikely durations when the interviewer suspends the interview and comes back to it later. Taking into account these findings, it was decided that the interviewer estimates of duration would be used for Wave 6 and the PAPI part of the Wave 7 dress rehearsal, while the sum of the section durations based on the timestamps would be used for the CAPI interviews. Moreover, if any section did not have the start and end times on the same day, or the duration was calculated to be more than 90 minutes, the section was excluded from the analysis.

[5] The main reason for the change in hardware was that the fieldwork provider was changed between Waves 8 and 9 (see Watson, 2010), although the dress rehearsal experience was also a factor.
[6] See www.confirmit.com.

Table 1: Survey features and interview duration information available in each wave

	Wave 6	Wave 7DR	Wave 9	Wave 10
Mode				
PAPI	✓	✓ (split sample)		
CAPI – laptop		✓ (split sample)		
CAPI – tablet			✓	✓
Dependent data				
Name, sex, date of birth	✓	✓	✓	✓
Date of last interview	✓	✓	✓	✓
Employment (Section C/D)		✓	✓	✓
Children (Section G)				✓
Marriage / living with (Section H)				✓
Home ownership (Section J)				✓
Tracking (Section T)			✓	✓
Special module				
Wealth	✓			✓
Retirement		✓		
Health			✓	
Duration information available				
Total PQ	✓	✓	✓	✓
Sections of PQ		Some (AA_H, K, T)	✓	✓
Total HQ	✓	✓	✓	✓
Sections of HQ	✓	✓	✓	✓

Note: DR=Dress Rehearsal, PQ=Person Questionnaire (continuing or new), HQ=Household Questionnaire.

The number of interviews conducted and the mean interview times are presented in Table 2. For the most part, the restrictions on the timestamps resulted in less than 1 per cent of cases being excluded, although up to 9 per cent of cases were excluded for the Wave 7 dress rehearsal CAPI sample. The mean interview length of the CPQ was 31.5 minutes for a face-to-face interview using paper methods in Wave 6, whereas in Wave 10, when the wealth module was repeated in CAPI, this fell to 30.0 minutes—a 5 per cent decrease. There was, however, no difference between Waves 6 and 10 in mean HQ interview length. (Note that the overall length of the HQ is much longer in wealth years because of the additional household-level questions asked.) A very different picture emerges from the Wave 7 dress rehearsal: PQ interviews conducted by PAPI took on average 32.0 minutes whereas by CAPI they took 41.6 minutes, an increase of 9.6 minutes, or 30 per cent. The HQ also took longer with CAPI than PAPI in the Wave 7 dress rehearsal—1.7 minutes, corresponding to a 32 per cent longer interview.

Table 2: Number of interviews and mean length by wave and survey mode, Waves 6, 7DR, 9 and 10

		Number interviewed			Number with non-missing lengths			Mean length (minutes)		
Wave	Mode	CPQ	NPQ	HQ	CPQ	NPQ	HQ	CPQ	NPQ	HQ
6	PAPI - face-to-face	11,260	761		11,258	761		31.5	37.5	
	PAPI - phone	800	84		799	84		27.4	33.9	
	All methods	12,060	845	7,133	12,057	845	7,129			11.8
7DR	PAPI - face-to-face	538	40		536	40		32.0	40.4	
	CAPI - face-to-face	625	34	340	588	31	336	41.6	48.5	7.0
	PAPI - phone	118	18		115	17		35.1	39.5	
	PAPI			366			366			5.3
	All methods	1,281	92	706	1,239	88	702			
9	CAPI - face-to-face	11,349	738		11,331	730		34.8	44.3	
	CAPI - phone	1,098	116		1,097	114		35.1	43.2	
	All methods	12,447	854	7,227	12,428	844	7,148			7.5
10	CAPI - face-to-face	11,615	742		11,596	740		30.0	39.9	
	CAPI - phone	1,084	87		1,079	86		29.5	40.4	
	All methods	12,699	829	7,305	12,675	826	7,168			11.8

Note: DR=Dress Rehearsal, CPQ=Continuing Person Questionnaire, NPQ=New Person Questionnaire, HQ=Household Questionnaire.

4. Effects of CAPI on overall interview length

Interview durations can be affected by a range of factors other than survey mode, most notably the characteristics of the respondents that determine the route they take through the questionnaire. Duration can also be affected by the respondent's ability to understand the questions, retrieve the relevant information, judge the accuracy of the response, and then select and deliver an answer (Tourangeau et al., 2000, p.7). We therefore take a regression approach to estimating the effect of CAPI on overall interview length, thereby controlling for many of these factors.

The models are estimated on pooled data from Waves 6, 9 and 10, together with the Wave 7 dress rehearsal. Included in the models for the PQs are variables for the characteristics of the respondents that determine which path through the questionnaire they take: employment status (two dummies), education level (two dummies), marital status, whether they have children and, if they have children, the number of children they have. The model for the HQ contains a similar set of variables, but excludes employment status and marital status of the respondent, includes variables for children living in the household rather than the respondent's own children, and adds variables for whether the household owns the dwelling, rents it, is involved in a rent-buy scheme, or lives there rent free. To allow for other factors which may affect the interview process, variables are additionally included (in all models) for age (five dummies), sex, English-speaking ability and presence of a long-term health condition. It is expected that these variables will capture differences in the speed with which the respondent is able to understand and answer the questions, although some of the variables—especially age and sex—may also affect which parts of the questionnaire are relevant to the respondent.

Table 3: Regression coefficients for interview lengths

	CPQ			NPQ			HQ		
Variable	Coeff	SE	Sig	Coeff	SE	Sig	Coeff	SE	Sig
Age (base=15-17 years)									
18-24	3.55	0.32	***	8.00	0.80	***	-0.04	0.53	
25-44	3.95	0.33	***	9.68	0.96	***	0.41	0.52	
45-64	5.02	0.34	***	10.71	1.29	***	0.69	0.52	
65-74	4.73	0.39	***	5.38	2.31	**	0.44	0.53	
75-99	5.33	0.41	***	9.51	2.99	***	0.83	0.53	
Female	0.65	0.11	***	1.46	0.49	***	-0.07	0.07	
Limited or no English	0.37	0.30		2.38	1.42	*	0.38	0.21	*
Employment status (base=employed)									
Unemployed	-1.68	0.31	***	-2.47	0.87	***			
Not in labour force	-9.13	0.15	***	-7.27	0.61	***			
Long-term health condition	3.45	0.13	***	5.59	0.64	***	0.31	0.08	***
Education (base=university or higher)									
Year 12 or Cert 3 or 4	-0.45	0.16	***	-1.93	0.87	**	-0.14	0.10	
Year 11 or below	-1.44	0.15	***	-2.87	0.85	***	-0.43	0.09	***
Has no children	-1.48	0.20	***	-2.21	0.98	**			
Number of own children	0.61	0.05	***	1.11	0.29	***			
Marital status (base=married)									
Defacto	2.53	0.19	***	1.27	0.90				
Separated	3.94	0.34	***	4.95	2.12	**			
Divorced	4.16	0.23	***	2.24	1.68				
Widowed	2.93	0.28	***	-0.60	2.53				
Never married	2.83	0.21	***	0.97	1.02				
New interviewer	2.29	0.14	***	1.56	0.61	***	0.56	0.09	***
Within-wave experience	-0.24	0.01	***	-0.23	0.06	***	-0.08	0.02	***
Within-wave experience squared	3.2E-03	2.5E-04	***	3.1E-03	1.1E-03	***	1.8E-03	6.0E-04	***
Within-wave experience cubed	-1.9E-05	1.8E-06	***	-1.8E-05	6.7E-06	***	-1.9E-05	7.9E-06	**
Within-wave experience 4^{th} power	3.5E-08	4.0E-09	***	3.2E-08	1.4E-08	**	6.2E-08	3.3E-08	*
Wave and mode (base=W6 face-to-face PAPI)									
W6 phone PAPI	-4.54	0.39	***	-5.37	1.40	***			
W7 face-to-face PAPI	-0.23	0.48		-0.54	2.02				
W7 face-to-face CAPI	8.81	0.46	***	9.57	2.25	***			
W7 phone PAPI	1.93	1.03	*	-1.28	3.02				
W9 face-to-face CAPI	2.91	0.14	***	6.15	0.64	***			
W9 phone CAPI	1.86	0.35	***	2.33	1.27	*			
W10 face-to-face CAPI	-1.22	0.14	***	2.83	0.64	***			
W10 phone CAPI	-3.25	0.35	***	0.76	1.45				
No child in household							-2.73	0.26	***
Number of children in household							1.95	0.22	***
Number of children in household squared							-0.22	0.04	***
Housing tenure (base=own)									
Rent							-3.00	0.08	***
Rent-buy scheme							-0.86	1.66	
Rent free							-1.98	0.20	***
Wave and mode (base=W6 PAPI)									
W7 PAPI							-6.79	0.29	***
W7 CAPI							-5.15	0.30	***
W9 CAPI							-4.72	0.09	***
W10 CAPI							0.08	0.09	
Constant	32.34	0.44	***	40.62	1.98	***	13.69	0.60	***
N	38,312			2,597			22,089		
Adjusted R-squared	0.206			0.359			0.335		

Notes: CPQ=Continuing Person Questionnaire, NPQ=New Person Questionnaire, HQ= Household Questionnaire, *, **, and *** indicate statistically significant difference at the 10, 5 and 1 per cent levels, respectively.

We have also allowed for two types of interviewer effects on duration. First, a dummy variable is included for interviewer 'survey status', which indicates whether the interviewer is new to the HILDA study in the current wave. Second, we include a fourth order polynomial for the cumulative number of interviews conducted for the wave as at the current interview, which we label 'within-wave experience'. The latter set of variables captures interviewer within-wave learning about the questionnaire, which would be expected to reduce interview durations. Finally, we include a series of dummy variables that specify the wave and mode of interview (whether face-to-face or phone and whether PAPI or CAPI).[7]

Table 3 presents the regression results. Coefficient estimates for respondent characteristics are broadly consistent with expectations, with respondents who are older, have long-term health conditions and/or have poor English having longer interview durations. Effects associated with interviewer characteristics are likewise as expected: new interviewers tend to take longer, and there is evidence of a within-wave learning effect for all interviewers. The shape of the polynomial fitted implies a learning effect across the first 40 interviews of about 5 minutes and then the average interview time seems to stabilise until around 140 interviews where the interview durations begin to decline again.[8]

The regression results provide several distinct pieces of evidence on the effect that interview mode has on interview length. First, Waves 6 and 10 had very similar questionnaire content, thereby providing a quasi-experiment on the effects of CAPI. For the CPQ regression, the estimate on the variable 'W10 f2f by CAPI' implies that CAPI on average reduced CPQ interview duration by 1.2 minutes, holding all else constant. This is almost certainly an underestimate of the impact of CAPI on CPQ length, since the Wave 10 CPQ contained additional questions to those included in the Wave 6 CPQ—in particular, questions on salary sacrifice and non-cash benefits. In contrast to the CPQ, the estimate for the NPQ implies CAPI increased interview length by 2.8 minutes (which is likely to be larger than the true impact of CAPI because of the presence of additional questions in Wave 10). This may in part reflect the fact that there is no dependent data in the NPQ, and therefore less scope for CAPI to reduce interview durations. For the HQ, the estimate implies no difference in length between CAPI and PAPI.

The evidence from the Wave 7 CAPI experiment contrasts markedly with the evidence from comparison of Waves 6 and 10. In the Wave 7 dress rehearsal, the CPQ took 9.4 minutes longer under CAPI, the NPQ took 10.3 minutes longer and the HQ took 1.6 minutes longer. Two main factors are likely to explain the contrast. First, the mouse-dependent nature of Confirmit, when implemented in the laptop environment adopted for the Wave 7 CAPI trial, may have slowed the interview more than other CAPI software may have. In Wave 10, a stylus was used on a tablet, which is much closer to the pen and paper environment with which interviewers were familiar. Second, the Wave 10 interview included more dependent data in the CPQ, which is likely to have shortened interviews.

Counter to expectations, and in contrast to the findings for face-to-face interviews, we find that the CPQ and NPQ interviews conducted by phone both show increases in length between Waves 6 and 10. However, this appears to be due to two very experienced telephone

[7] It is not known whether a household interview was undertaken face-to-face or by phone, so we can only include variables for whether it was undertaken by PAPI or CAPI.

[8] The decline at this high end is likely due to a small number of very experienced interviewers being given additional work as they work quickly. A separate polynomial for new and experienced interviewers was tested, but the simpler specification performed reasonably well. Note that further investigation of interviewer learning effects is undertaken in Section 6.

interviewers undertaking 40 per cent of all telephone interviews in Wave 6. These two interviewers had interview lengths on average 10.5 minutes lower than the other interviewers for the CPQ and 16.3 minutes lower for the NPQ. By comparison, in Wave 10, the share of the telephone interviews undertaken by the two interviewers with the most interviews was only 16 per cent; moreover, their mean interview lengths were no different to the rest of the interviewers.

5. Effects of CAPI on section lengths

In this section, we give more consideration to the mechanisms by which CAPI impacts on interview durations, in particular focusing on the impact of dependent data. We do this by comparing across waves the lengths of specific questionnaire sections with differing amounts of dependent data. Our first set of results use data from Waves 9 and 10, and we limit our investigation to only those sections where the questions are identical or almost identical—thus excluding Section K (on health and lifestyle issues) and Section J (on personal assets and debts). We use the same regression model specification here as was used in the models of overall interview duration presented in Table 3, but restrict the sample to Waves 9 and 10, and therefore necessarily exclude the survey mode dummies for Waves 6 and 7.

Table 4 presents the mean difference between Waves 9 and 10 in the duration of each questionnaire section after controlling for differences in respondent and interviewer characteristics. Separate models are estimated for each CPQ and NPQ section, and separate estimates are presented for face-to-face and phone interviews. These 'adjusted' mean differences are obtained from a set of dummy variables included in each model that distinguish both survey mode and wave of interview. The omitted dummy is 'W9 face-to-face', so that the estimate for face-to-face interviews is simply the coefficient on the dummy 'W10 face-to-face', while the estimate for phone interviews is the coefficient on 'W10 phone' minus the coefficient on 'W9 phone'. Table 4 also provides a brief description of the (relatively minor) questionnaire changes to each section between Waves 9 and 10, as well as a description of the dependent data added in Wave 10.

Focusing first on the results for face-to-face interviews, we find that for several of the sections where there had been no change to the questionnaire or the use of dependent data between Waves 9 and 10 (Sections A and B), the interview durations have also not changed. Further, where a few additional questions were included in Section E (on other labour market activities), we see an expected increase in Wave 10.

Surprisingly, interview lengths for the sections on current employment (C) and for those not employed (D) fell, by 0.8 and 0.3 minutes respectively, even though there were very minimal changes to the questionnaire and there was no additional use of dependent data. This decline is most likely due to the time taken with the industry and occupation questions to record appropriate descriptions for occupation and industry. Analysis of the length of the text recorded at Wave 9 versus Wave 10 indicated a small but significant reduction in the number of characters recorded for occupation title (1.5 characters) and tasks (2.6 characters), but an increase for industry (0.8 characters). These differences in text length are small and therefore do not account for the overall time decrease. An alternative explanation is that interviewers had become more proficient in writing with the stylus by Wave 10, leading the character recognition software to more accurately interpret their writing, and thus reducing the amount of time they spent on these questions. This explanation is also consistent with the results for Section F, which contains many write-in boxes where the interviewer records the amount of income received from different sources. For this section, despite the addition of 14 quite

complicated questions, only partially offset by the removal of three questions, the adjusted mean duration of the section decreased by 0.2 minutes.

For the sections that incorporated new dependent data in Wave 10, we see a reduction in duration for Section G (children) by 0.4 minutes and an increase in Section H (relationships) by 0.1 minutes. In Section G, the interviewer no longer had to write the name, sex and age of pre-existing children, so it is sensible that time savings were made. The increased duration of Section H suggests that, for marital and cohabitation status, it takes longer to remind the respondent of their answer in the previous interview than it does to get them to recall the information (but presumably it improves consistency across waves).

Table 4: Comparison of section lengths in Waves 9 and 10

Section	N	Adjusted mean difference in length (W10 - W9)						Questionnaire changes in W10	Dependent data added in W10
		Face-to-face			Phone				
		Coeff	SE	Sig	Coeff	SE	Sig		
Continuing Person Questionnaire (CPQ)									
A Education	24,932	0.01	0.02		-0.13	0.06	**	No change	None
B Employment status	24,973	-0.01	0.01		-0.01	0.03		No change	None
C Current employment	16,143	-0.75	0.07	***	-0.46	0.20	**	Minor change - added C7 (available to work now)	None
D Not employed	8,792	-0.28	0.06	***	-0.63	0.21	***	Essentially no change – removed one option at D3 (unemployment duration)	None
E Other labour market activity	24,939	0.30	0.04	***	0.33	0.11	***	Added 5 questions on job-related discrimination	None
F Income	24,944	-0.23	0.05	***	-0.67	0.15	***	Removed F30 (government bonus payments) and F34 and F35 (credit cards); added 14 questions on non-cash benefits	None
G Children	24,953	-0.44	0.03	***	-0.33	0.09	***	No change	Names, sexes and ages of own children
H Relationships	24,979	0.08	0.01	***	-0.07	0.03	**	Essentially no change – one word change to H16 (defacto relationships)	Marital status and cohabitation status
New Person Questionnaire (NPQ)									
AA Country of birth & language	1,645	0.09	0.08		-0.21	0.21		No change	None
BB Family background	1,647	-0.22	0.14		-0.05	0.38		No change	None
A Education	1,647	-0.17	0.10	*	0.15	0.26		No change	None
B Employment status	1,650	0.00	0.02		-0.02	0.04		No change	None
C Current employment	938	-0.34	0.23		0.39	0.61		Changes as per CPQ	None
D Not employed	712	-0.33	0.31		0.08	0.84		Changes as per CPQ	None
E Other labour market activity	1,646	0.51	0.16	***	0.73	0.42	*	Changes as per CPQ	None
F Income	1,648	0.16	0.14		-0.24	0.39		Changes as per CPQ	None
G Children	1,648	-0.13	0.09		0.32	0.23		No change	None
H Relationships	1,649	-0.02	0.06		-0.19	0.17		Changes as per CPQ	None

Notes: The regression models include controls for age, sex, marital status, English ability, employment status, long-term health condition, education, children in the household, housing tenure, whether the interviewer is new to the study, and how many individual interviews the interviewer has completed as of the time of the interview. *, **, and *** indicate statistically significant difference at the 10, 5 and 1 per cent levels, respectively.

Examining CPQ estimates for the interviews conducted by phone, we see similar effects as evident for face-to-face interviews. The notable exceptions are the sections on education (A) and relationships (H). These differences are not readily explicable.

For the NPQ sample (lower panel of Table 4), most estimates are not statistically significant, which is consistent with the fact that there is no dependent data in the NPQ, and therefore no scope for reductions in interview time from this source. However, it should also be noted that the relatively small size of the NPQ sample is likely to be a factor in the absence of statistically significant estimates. Of the few estimates that are statistically significant, only one—that for face-to-face interviews in Section A—is qualitatively different from the CPQ estimate, and it is only of marginal significance.

Further evidence on the role played by dependent data is available in the tracking section (Section T) of the questionnaire. For this analysis, we draw on data from not only Waves 9 and 10, but also the Wave 7 dress rehearsal, which collected timestamps before and after this section (but not other sections). The results are presented in Table 5. We find that compared to the paper version of the tracking section, the laptop based CAPI (with no feed forward data) took 0.5 minutes longer in the CPQ when conducted face-to-face. When dependent data was introduced in Wave 9, there was a reduction over the paper-based method by a sizeable 2.6 minutes, which is about 8 per cent of the total length the CPQ without the dependent data in Section T. We also see a further reduction from Wave 9 to Wave 10, despite there being no changes to either the questions asked or the data that was brought forward from the last interview. Anecdotal evidence from the fieldwork company indicate that in Wave 9 there was a greater focus on making sure the tracking information was complete and correct, since it was the first time tracking information had been fed forward. Once this editing work had been done, there was less to update in Wave 10. Nevertheless, we also see a drop in the time taken to complete Section T for the NPQ, where all new contact information needs to be collected, implying that at least part of the reduction in time taken to complete Section T was due to improvements in interviewer competency with the CAPI console and stylus.

Table 5: Effect of mode change on the tracking section lengths (Waves 7DR, 9 and 10)

	CPQ			NPQ		
Variable	Coeff	SE	Sig	Coeff	SE	Sig
Wave and mode (base=W7 face-to-face PAPI)						
W7 face-to-face CAPI	0.49	0.14	***	0.46	0.69	
W7 phone PAPI	-0.39	0.29		-1.11	1.38	
W9 face-to-face CAPI	-2.56	0.11	***	0.18	0.53	
W9 phone CAPI	-2.42	0.13	***	-0.18	0.59	
W10 face-to-face CAPI	-2.88	0.11	***	-0.14	0.53	
W10 phone CAPI	-3.07	0.12	***	-0.53	0.60	
N	26,029			1,702		
Adjusted R-squared	0.088			0.044		

Notes: The regression models include controls for age, sex, marital status, English ability, long-term health condition, education, number of children, marital status, whether the interviewer is new to the study, and how many individual interviews the interviewer has completed as of the time of the interview. *, **, and *** indicate statistically significant difference at the 10, 5 and 1 per cent levels, respectively.

6. Interviewer learning effects under CAPI

To explore the ways in which interviewers adapt to different interviewing modes, for each wave we estimate a regression model of the determinants of CPQ interview duration that includes a set of dummy variables for within-wave experience—that is, the number of interviews conducted by the interviewer in the current wave as at the current interview: 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-74, and 75-300. We furthermore distinguish between interviewers new to the HILDA Survey in that wave and interviewers who had conducted interviews in one or more previous waves by interacting the within-wave experience dummies with dummies to identify new and experienced interviewers each wave. In addition, for the Wave 7 dress rehearsal, separate estimates are produced for the PAPI and CAPI samples.

The key estimation results from these models are summarised by Figures 1 and 2, which present graphs of 'standardised' mean predicted CPQ interview durations by level of within-wave experience. The mean predicted duration is calculated by evaluating the predicted duration for each within-wave experience dummy, holding other explanatory variables fixed at mean values. This predicted duration is then 'standardised' by setting the predicted duration to 30 minutes for experienced interviewers with within-wave experience of between 50 and 74 interviews, and rescaling all other predicted durations accordingly (so that they maintain the same relativity with respect to experienced interviewers with 50-74 interviews in each wave).

The graphs provide clear indications that learning effects of interviewers differ by survey mode. Figure 1 compares mean predicted durations of new and experienced interviewers in Wave 6 (conducted via PAPI) and Wave 10 (conducted via CAPI). For the interviewers who were new to the survey, the CAPI method enables quicker learning of the survey process and after only 15 or so interviews, the average duration of their interviews are reasonably close to the experienced interviewers. When the interviewers are working with the paper-based methods, their interview durations are still around 4 minutes longer by the time they reach 30 interviews, and they never achieve parity with the experienced interviewers. Even after 50 interviews, they remain an average of 2.5 minutes longer. The learning effect of the experienced interviewers is smaller and, in contrast to new interviewers, is not significantly different between PAPI and CAPI.

In Figure 2 we compare mean predicted durations across waves for CAPI methods only. Compared to Wave 10, we see that there were much greater learning effects in Wave 9, when the experienced interviewers were learning the tablet and Confirmit technology together with the new sections of the Wave 9 questionnaire. This 2.5 minute difference between Waves 9 and 10 evaporated by around the 15th interview. By comparison, in the Wave 7 dress rehearsal CAPI test, the experienced interviewers took the first 24 or so interviews to grasp the technology, after which their average interview duration dropped by about 5 minutes.

As a robustness check of the results presented in Figures 1 and 2, we repeated the analysis restricting the sample to 'high–yield' interviewers, who we define as interviewers who completed at least 75 interviews that wave. Essentially, this removes any effect of interviewers who were struggling with the interview process and perhaps were taking a long time to complete the interviews. It also removes interviewers who may not have experienced enough interviews to have completed their 'learning'. The results were, however, very similar to those reported in Figures 1 and 2.

To summarise, the tablet-based CAPI method helps new interviewers 'get up to speed' with the questionnaire and interview process much quicker than occurs with paper-based methods.

For the experienced interviewers, once they have experienced CAPI for one wave, the within-wave learning effects in CAPI are quite similar to those evident for paper-based methods.

Figure 1: Effects of within-wave interviewer experience on CPQ interview duration, Waves 6 and 10

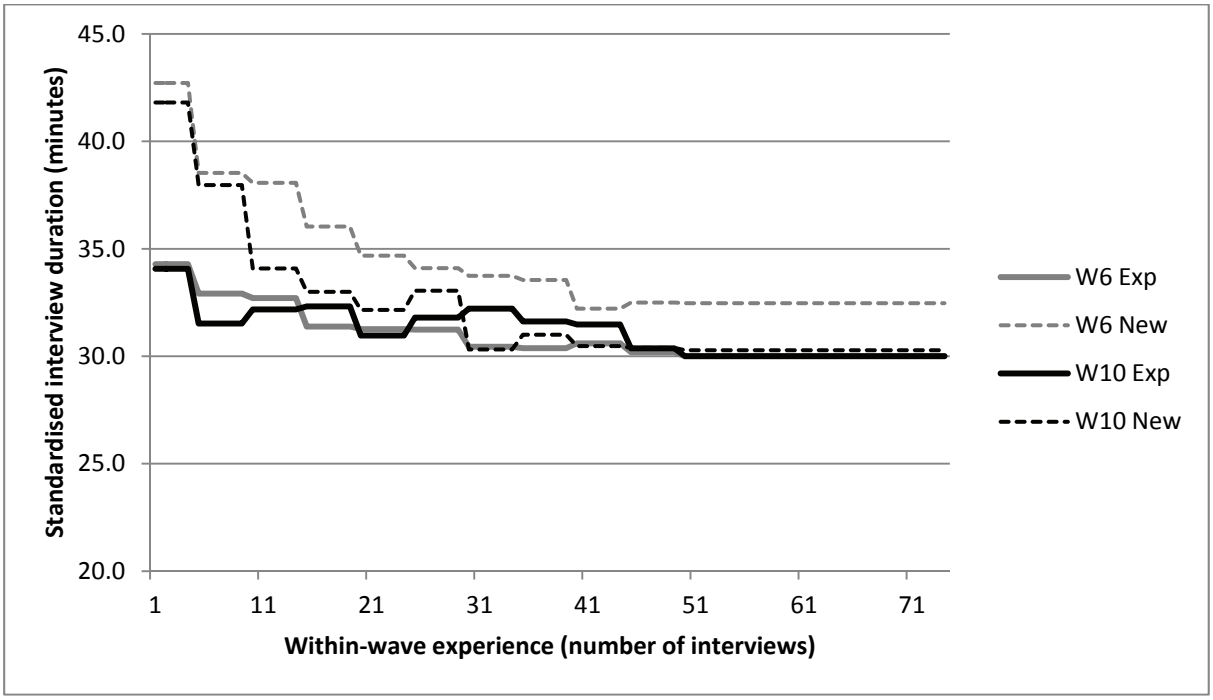
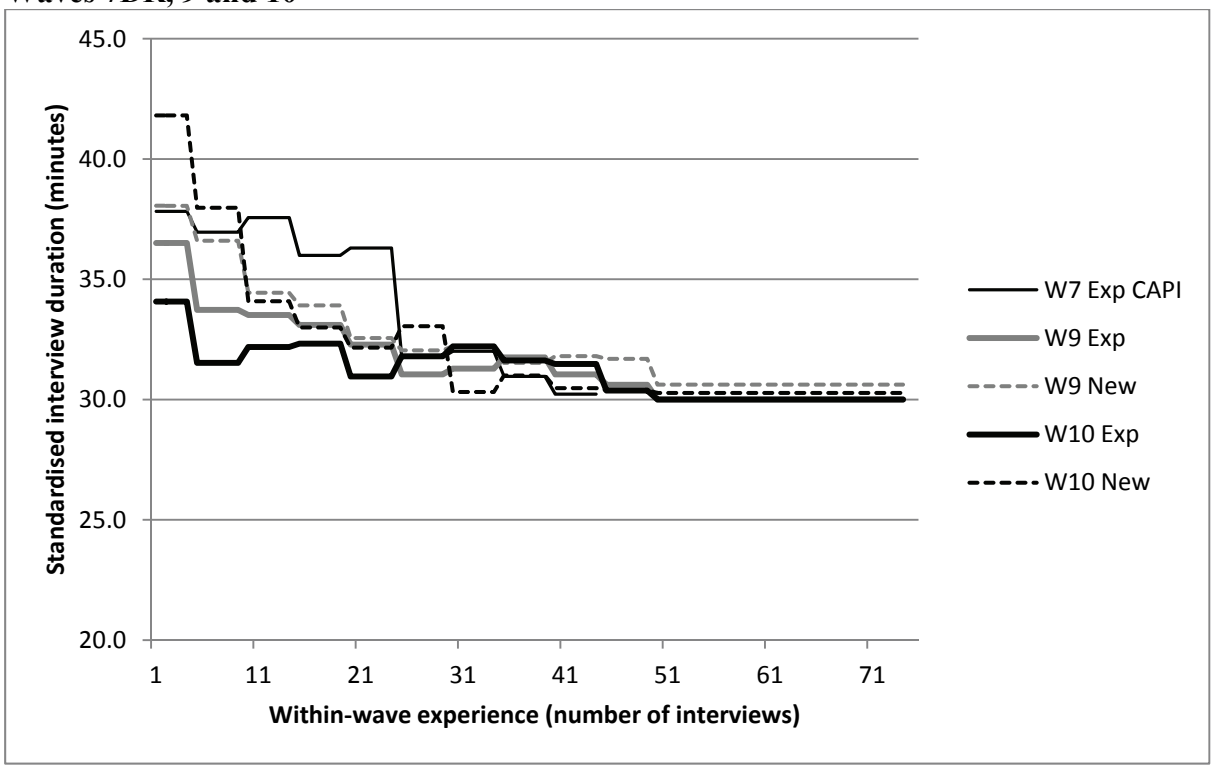


Figure 2: Effects of within-wave interviewer experience on CPQ interview duration, Waves 7DR, 9 and 10



7. Decomposition of the difference in interview lengths between CAPI and PAPI

The results of the preceding analyses can be used to quantify the roles played by specific factors in determining the overall effect of the introduction of CAPI on interview duration. In particular, by comparing the total CPQ duration in Wave 10 with the CPQ duration in Wave 6, and drawing on the evidence presented in Sections 5 and 6, we can decompose the change in duration resulting from the move from PAPI to CAPI, as implemented in Wave 10, into 'dependent data', 'interviewer learning', 'interviewer composition' and 'other' change of mode effects.

Table 6 presents the results of this decomposition exercise, showing the estimated effect of each component and the source of the estimate. The upper panel draws on the results obtained in Section 5 to ascertain the effects of the introduction of dependent data. It shows that using dependent data reduced CPQ interview duration by 3.2 minutes, with most of the reduction coming from the feeding-forward of contact information in the tracking section. Some further reductions were also made in section on children, although these were partly offset by an increase in time in the relationship section.

The effects of changes to interviewer learning and the composition of interviewers were calculated by decomposing the following expression for the difference between Waves 10 and 6 in the mean *predicted standardised* interview length:

$$\bar{l}_{10} - \bar{l}_{6} = \sum_{g=1}^{2}\sum_{i=1}^{12} p_{ig,10}\bar{l}_{ig,10} - \sum_{g=1}^{2}\sum_{i=1}^{12} p_{ig,6}\bar{l}_{ig,6} \qquad (1)$$

where g indicates interviewer survey status (experienced or new), i describes within-wave experience (in 12 categories as defined in Section 6), $p_{ig,w}$ is the proportion of interviews undertaken by interviewers in Wave w with survey status g and within-wave experience in category i, and $\bar{l}_{ig,w}$ is the mean predicted standardised interview length for the same group.

Equation (1), by expressing the mean interview length in each wave as a weighted average of mean interview lengths across groups defined by interviewer survey status and within-wave experience, can be decomposed into four components as follows:

$$\bar{l}_{10} - \bar{l}_{6} = \sum_{i=1}^{12} p_{i1,10}\left(\bar{l}_{i1,10} - \bar{l}_{i1,6}\right) + \sum_{i=1}^{12} p_{i2,10}\left(\bar{l}_{i2,10} - \bar{l}_{i2,6}\right) + \sum_{g=1}^{2}\sum_{i=1}^{12} \bar{l}_{ig,6}\left(p_{ig,x} - p_{ig,6}\right) + \sum_{g=1}^{2}\sum_{i=1}^{12} \bar{l}_{ig,6}\left(p_{ig,10} - p_{ig,x}\right) \qquad (2)$$

where $p_{ig,x} = p_{ig,10}{}^{\sum_1^{12} p_{ig,6}}\Big/_{\sum_1^{12} p_{ig,10}}$.

The first two terms combined estimate the change in the mean interview length between Waves 6 and 10 due to different learning effects of interviewers. They do this by calculating the change in the mean predicted standardised interview length that would occur if the interviewer composition (in terms of both survey status and within-wave experience) was the same in Wave 6 as the actual composition in Wave 10. The first term gives the contribution made by differences in learning by experienced interviewers and the second term gives the contribution made by differences in learning by new interviewers.

The third term of Equation (2) estimates the difference in mean interview length due to changes in the distribution of within-wave experience—that is, due to changes in the proportions of interviews conducted by interviewers with each of the 12 levels of within-wave experience (as at the time of the interview). The variable $p_{ig,x}$ is a rescaling of $p_{ig,10}$ such that the proportion of interviews conducted by new (as opposed to experienced) interviewers is held constant at Wave-6 levels, which allows us to isolate the effects of changes in the distribution of within-wave experience. The mean interview length of each of the 24 groups defined by survey status and within-wave experience is also held constant at the Wave-6 level, which means the effect is evaluated at Wave-6 interview lengths.

The fourth term estimates the difference in lengths due to change in the split between new and experienced interviewers, also evaluated at Wave-6 interview lengths. This is identified by holding constant the proportion of interviews in each of the 24 within-wave experience and survey status groups and, as in the third term, holding constant the mean interview length of each survey status by within-wave experience group.

The decomposition provides some useful information about the role of CAPI and various workforce decisions have on overall interview lengths. In terms of the differential learning effects between PAPI and CAPI, we find a slight increase in mean standardised interview length due to the experienced interviewers (by 0.1 minutes), whereas new interviewers contributed a decrease in overall length (by 0.2 minutes).[9] Changes in the distribution of within-wave experience acted to increase the mean interview length by 0.8 minutes. This was the result of an increase in the size of the interviewer workforce between Waves 6 and 10 from 140 to 155, which meant that a higher proportion of interviews were conducted while the interviewers were still in their learning phase (given that the total number of interviews was similar in the two waves). Balanced against this, however, the change in the proportion of interviews conducted by new interviewers—specifically, a reduction from 24 per cent in Wave 6 to 13 per cent in Wave 10—acted to decrease the mean interview length by 1.0 minute. Thus, the net effect of changes to learning effects and the composition of the interviewers between Waves 6 and 10 was to remove 0.3 minutes from the mean interview length.

The last panel of Table 6 simply reproduces the reduction in the mean overall interview length between Waves 6 and 10, which was 1 minute. Since the sum of the changes across all components must equal this overall change, we can infer the net role played by other factors as equal to the difference between the sum of the known components and the actual overall change. This residual difference is 2.5 minutes, meaning that changes between Waves 6 and 10 other than the introduction of dependent data, changes in interviewing learning and changes in interviewer composition acted to increase the mean CPQ interview duration by 2.5 minutes. Since very little changed between Wave 6 and Wave 10 other than the introduction of CAPI, we can be reasonably confident that this increase is CAPI-related—that is, due to characteristics of CAPI (as implemented in Wave 10) other than dependent data and differences in interview learning and composition.

[9] While the interview lengths for new interviewers only have declined by 1.9 minutes, the proportion of interviews conducted by new interviewers is only 0.13, so the overall effect is small.

Table 6: Decomposition of change in CPQ interview length between Waves 6 and 10

Component	Effect (mins)	Source
Dependent data		
Contact details (Section T)	-2.9	Table 5^
Children (Section G)	-0.4	Table 4^
Relationships (Section H)	0.1	Table 4^
Interviewer learning		
Experienced interviewers	0.1	First term of Equation 2, using information from Graph 1 in conjunction with information on the composition of interviewers.
New interviewers	-0.2	Second term of Equation 2, using information from Graph 1 in conjunction with information on the composition of interviewers.
Interviewer composition		
Within-wave experience	0.8	Third term of Equation 2, using information from Graph 1 in conjunction with information on the composition of interviewers.
Survey status	-1.0	Fourth term of Equation 2, using information from Graph 1 in conjunction with information on the composition of interviewers.
Total difference	-1.0	Table 3^
Other aspects of CAPI (obtained by subtraction)	2.5	

Notes: ^Estimate obtained from the specified table that provided a split between face-to-face interviews and phone interviews. For Wave 10, 8 per cent of interviews were conducted by phone, so an overall estimate is obtained by taking 0.92 of the face-to-face estimate and 0.08 of the phone estimate.

This decomposition makes it clear that there are several factors that affect the interview lengths when we move from paper-based interviewing to computer-assisted interviewing. The use of dependent interviewing can have a large effect and changes in the learning effects for interviewers can be quite small when aggregated across new and experienced interviewers. Had we only used a small amount of dependent data and kept the same composition of our interviewer workforce, the introduction of CAPI into HILDA may well have led to an increase in CPQ interview length, of the order of about 8 per cent.

8. Conclusion

A clear message from the analysis undertaken in this paper is that the effects of introducing CAPI into an existing study on interview length very much depend on how CAPI is implemented. The hardware and software adopted, the extent and nature of dependent data introduced, and even interviewer workloads, can all have large impacts on interview durations experienced under CAPI. This is perhaps most starkly illustrated by the experience in the HILDA Wave 7 dress rehearsal experiment, where the particular combination of hardware and software adopted led to the concerning result that the introduction of CAPI increased interview duration by 30 per cent in the individual interview. This was significantly

larger than reported by other studies that have found that CAPI interviews were longer, where the increase was of the order of 12 to 17 per cent (Martin el al., 1993; Fuchs et al. 2000). By examining the difference in interview lengths between Waves 6 and 10, which had almost identical questionnaire content, we are able to conclude that this increase in the Wave 7 experiment was the net outcome of the hardware-software combination used, greater learning difficulties experienced by the interviewers, and very limited use of dependent data.

The reduction in interview length between Waves 6 and 10 can be explained in large part by the use of dependent data, which reduced the mean CPQ interview duration of 31.5 minutes in Wave 6 by 10 per cent. We furthermore find, drawing on Wave 9 in addition to Waves 6 and 10, that the dependent data which produced the greatest reductions in interview length was for questions that require writing. In particular, the feeding forward of contact details in the 'tracking' section (Section T) acted to reduce interview lengths by an average of 2.9 minutes; and feeding forward children's names in Section G also had a sizeable impact, reducing interview lengths by an average of 0.4 minutes. It therefore appears that reductions in interview length from using dependent interviewing come mainly from reducing the amount of writing required of interviewers.

A further beneficial effect of CAPI we find is that new interviewers are able to get up to speed with the questionnaire and interview process much quicker with CAPI than with paper questionnaires. These gains were in our case eliminated by slightly lower rates of learning for the experienced interviewers, leading to only a 0.3 per cent reduction in length overall due to differences in learning effects. We did, however, find that interviewers appeared to improve their writing skills with the tablet over the initial waves, as evidenced by the time reductions experienced between Waves 9 and 10 in the sections relating to employment (Sections C and D – jointly 0.6 minutes), income (Section F – 0.2 minutes), and tracking (Section T – 0.3 minutes).

We also note that it can be important to account for compositional changes in the interviewer workforce when assessing the effects of CAPI introduction. We find that, on average, interviewers need to interview approximately 50 respondents in order to achieve stable interview lengths. Larger interviewer workloads will thus reduce overall respondent burden. In the case of the HILDA Survey, an expansion of the workforce coincident with the introduction of CAPI resulted in a reduction in workload sizes, which acted to increase the mean CPQ interview duration by 3 per cent. Nevertheless, this was completely offset by a shift in interviewer composition in Wave 10 towards more experienced interviewers.

Finally, while the individual contributions of other CAPI features to the overall effect of CAPI on interview lengths are not identified in this paper, we estimate that their combined effect was to increase interview length by 8 per cent. This is likely to reflect the effects of features such as automated confirmation of unexpected responses, slower entry of text, and loss of interviewer ability to 'read ahead'. The net reduction in overall interview length as a result of moving from PAPI to CAPI, as implemented in Wave 10, was therefore only 3 per cent.

9. References

Baker, R.P., Bradburn, N.M., and Johnson, R.A. (1995), 'Computer-assisted Personal Interviewing: An experimental evaluation of data quality and costs', *Journal of Official Statistics*, Vol. 11, No. 4, pp. 413-431.

Banks, R. and Laurie, H. (2000), 'From PAPI to CAPI: The case of the British Household Panel Survey', *Social Science Computer Review*, Vol. 18, No. 4, pp. 397-406.

Caviglia-Harris, J., Hall, S., Mullan, K., Macintyre, C., Bauch, S.C., Harris, D., Sills, E. Roberts, D. Toomey, M., and Cha, H. (2012),'Improving household surveys through computer-assisted data collection: Use of touch-screen laptops in challenging environments', *Field Methods*, Vol. 24, No. 1, pp. 74-94.

de Leeuw, E., Hox, J., and Snijkers, G. (1995), 'The effect of computer-assisted interviewing on data quality', *Journal of the Market Research Society*, Vol. 37, No. 4, pp. 325-344.

Fuchs, M., Couper, M. and Hansen, S.E. (2000), 'Technology effects: Do CAPI or PAPI interviews take longer?', *Journal of Official Statistics*, vol. 16,no. 3., pp. 273-286.

Jäckle, A. (2008), 'Dependent Interviewing: Effects on Respondent Burden and Efficiency of Data Collection', *Journal of Official Statistics*, Vol. 24, No. 3, pp. 411-430.

Laurie, H. (2003), 'From PAPI to CAPI: Consequences for data quality on the British Household Panel Survey', Working Papers of the Institute for Social and Economic Research, No. 2003-14, University of Essex, Colchester.

Lynn, P. (1998), 'Data collection mode effects on responses to attitudinal questions', *Journal of Official Statistics,* Vol. 14, No. 2, pp. 1-14.

Martin, J., O'Muircheartaigh, C. and Curtice, J. (1993), 'The use of CAPI for attitude surveys: An experimental comparison with traditional methods', *Journal of Official Statistics*, Vol. 9, No. 3, pp. 641-661.

Nicoletti, C. and Peracchi, F. (2003), 'Survey response and survey characteristcs: Micro-level evidence from ECHP', Economics and Statistics Discussion Paper No. 15/04.

Schräpler, J., Schupp, J., and Wagner, G.G. (2006), 'Changing from PAPI to CAPI: A longitudinal study of mode-effects based on an experimental design', DIW Discussion Paper No. 593, Berlin.

Tourangeau, R., Rips, L., and Rasinski, K. (1990), *The Psychology of Survey Response*, Cambridge: Cambridge University Press.

van Bastelaer, A., Kerssemakers, F., and Sikkel, D. (1988), 'Data collection with hand-held computers: Contributions to questionnaire design', *Journal of Official Statistics*, Vol. 4, No. 2, pp. 141-154.

Watson, N. (2010), 'The impact of the transition to CAPI and a new fieldwork provider on the HILDA Survey', HILDA Discussion Paper Series No. 2/10, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N. and Wilkins, R. (2011), 'Experimental change from paper-based interviewing to Computer-Assisted interviewing in the HILDA Survey', HILDA Discussion Paper Series No. 2/11, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Wooden, M. and Watson, N. (2007), 'The HILDA Survey and its Contribution to Economic and Social Research (So Far)', *The Economic Record*, vol. 83, no. 261, pp. 208–231.