Data Collection Mode Effects on Responses to Attitudinal Questions

Peter Lynn[1]

A study comparing two data collection modes – pencil and paper (PAPI), and computer-assisted personal interviewing (CAPI) – is reported. The study was part of the *British Social Attitudes* survey, an annual survey involving personal interviews, consisting mainly of attitude and opinion questions, in respondents' homes. In 1993, half the sample was assigned to each mode. This is the first large-scale study in the U.K. to permit direct assessment of CAPI/PAPI mode effects. Little evidence is found of mode effects on propensity to respond, or to agree to be reinterviewed. There is, however, strong evidence of effects on response patterns. CAPI respondents are more likely to choose extreme categories to answer scales and less likely to answer ''don't know.'' There are also differences in scale means. Interview length is slightly shorter with CAPI.

Key words: Attitude questions; computer assisted interviewing; data collection modes; interview length; response rates; response effects.

1. Introduction

This article assesses the effect of a change from the traditional Pencil-and-Paper Interviewing (PAPI) method to Computer-Assisted Personal Interviewing (CAPI), on an annual repeated cross-sectional survey of attitudes – the *British Social Attitudes* survey. It builds upon and extends the pioneering work of Martin, O'Muircheartaigh and Curtice (1993). The data analysed were collected during the 1993 round of the survey, when a random half of the sample was assigned to CAPI, and the other half to PAPI. The experiment was mounted primarily to assess the feasibility and efficiency of CAPI for this particular survey. This led to the conclusion that the survey should switch to CAPI. Since 1994 it has been carried out entirely in CAPI. However, the experiment also offers the opportunity to examine data collection mode effects on response patterns, and this is the main focus of this article. The article also looks at effects on cooperation rates, interview length, and costs.

2.   Computer-Assisted Personal Interviewing

CAPI is a data collection mode which involves the interviewer entering responses directly into a portable lap-top computer; a paper version of the questionnaire is no longer necessary. The nature of CAPI is described and its main advantages and disadvantages discussed by Weeks (1992) and Martin and Manners (1995). Despite the rapid growth of CAPI in recent years, only limited research has been published on its effects on survey data. There seems to be widespread agreement that interviewers do not have particular difficulties accepting, or coping with, CAPI (Manners 1987; van Bastelaer, Kerssemakers and Sikkel 1988; Olsen 1992; Martin, O'Muircheartaigh and Curtice 1993) and may indeed prefer it to PAPI (Baker 1990; Weeks 1992). And it has been demonstrated that CAPI reduces the amount of missing data (Sebestik et al. 1988; Olsen 1992; Martin, O'Muircheartaigh and Curtice 1993). Other findings relate to cooperation rates (Manners 1987; van Bastelaer, Kerssemakers and Sikkel 1988; Weeks 1992; Martin, O'Muircheartaigh and Curtice 1993; Baker, Bradburn and Johnson 1995), response effects (Baker, Bradburn and Johnson 1995; Nicholls, Baker and Martin 1997) and interview length (Couper, Groves and Kosary 1989; Martin, O'Muircheartaigh and Curtice 1993; Baker, Bradburn and Johnson 1995). But nearly all the research into CAPI effects has been carried out on large, continuing, government-sponsored surveys, mainly soliciting factual information. The one exception is Martin, O'Muircheartaigh and Curtice (1993), which investigated attitudinal items on ad hoc (one-time) surveys, but with rather small samples. It was largely the results of that study which encouraged the research team on the *British Social Attitudes* survey to experiment with CAPI.

3.   Study Design

The sample design for the survey (Jowell et al. 1994) involved selecting 176 postal sectors from throughout Britain with probability proportional to number of postal addresses, and then 28 addresses within each selected sector. A postal sector is a contiguous geographical area containing about 2,500 addresses, on average. The sampling frame was the Postcode Address File (Lynn and Lievesley 1991). At each address, the interviewer made a random selection of one individual aged 18 or over for interview, using an amended version (Lynn 1992) of a method first proposed by Kish (1949). This selection was done on the doorstep using PAPI methods. A random half of the 176 sectors were assigned to CAPI, and the other half to PAPI. This design provided a good test of the *feasibility* of CAPI, as within each of the half-samples the survey could be administered exactly as it would be if the full survey were carried out in that mode. But it is not an ideal design to test the *effect* of CAPI on the survey data, because CAPI effects are confounded with area effects and interviewer effects. (For the most part, one sector was covered by one interviewer. The allocation of interviewers to PSUs was largely determined by geographical location.)

The mode effects study is also sub-optimal for other reasons. The CAPI sample was issued in two stages in order to reduce the number of lap-top computers that were needed. The first 44 sectors (CAPI wave 1) were issued to interviewers, who had five weeks to complete their fieldwork. The computers were then transferred to a different set of interviewers to work on the other 44 CAPI sectors (CAPI wave 2). Fieldwork for CAPI wave 2 took around seven weeks. The 88 PAPI sectors, on the other hand, were all issued at one

time, and fieldwork continued over a period of around three months. Also, reissues (addresses that are issued to a second interviewer, in order to boost response rate, after the first interviewer has failed to make contact or recorded a ''weak'' refusal) amongst the CAPI wave 1 sample were carried out in PAPI (because the interviewers in those sectors no longer had lap-top computers, as they had been transferred to wave 2 interviewers). In the analysis below of effects on responses to questions, interviews are classified by the mode in which they were actually carried out, rather than the mode to which they were originally assigned.

4.   Effects on Unit Non-response

It is of interest to know whether CAPI induces different co-operation behaviour from PAPI. If CAPI alters sample members' propensity to respond, then this could affect the nature of nonresponse bias in the survey estimates. The extent to which bias is affected will depend both on the level of nonresponse, and the way it is distributed across sample sub-groups. Previous studies have failed to identify any significant overall response rate differences between CAPI and PAPI (Manners 1987; van Bastelaer, Kerssemakers and Sikkel 1988; Weeks 1992; Martin, O'Muircheartaigh and Curtice 1993; Baker, Bradburn and Johnson 1995; Nicholls, Baker and Martin 1997), though in all but one study the direction of the difference was the same (lower response with CAPI) and in a couple of them the magnitude of the difference was close to achieving significance. In the current study, it was assumed that mode of interviewing would be unlikely to affect the propensity for the interviewer to achieve contact at a sampled address. Refusal rates, calculated on the base of addresses at which contact was made, have therefore been compared. There were slightly more refusals amongst the CAPI sample (Table 1), though the difference, taking into account the clustered nature of the sample, does not achieve significance (two-tailed $P = 0.18$).

However, as explained above, this simple comparison confounds the effect of mode of interviewing with the effect of differing fieldwork periods. The shorter fieldwork period for the CAPI sample may have reduced the response rate, as interviewers may have more often had to accept a refusal rather than backing off and trying again on another day. This comparison therefore provides limited information, and it should not be interpreted as evidence that CAPI *per se* induces more refusals than PAPI. It is also argued (Martin and Manners 1995) that major changes to fieldwork procedures may produce an

Table 1.   Refusal rates

	Refusals	Productive response	Base (addresses at which contact made)
CAPI wave 1	31.9%	64.4%	1,024
CAPI wave 2	32.9%	62.9%	979
CAPI total	32.4%	63.7%	2,003
PAPI	30.0%	67.6%	2,030

Note: The small number of cases neither classified as a refusal nor a productive response were mainly cases where the sample member was unable rather than unwilling to respond (inadequate English spoken, too ill, temporarily away from home, etc.) plus a few miscellaneous nonresponse categories.

initial slight reduction in response rates, while the interviewers and field staff are adjusting to the new demands, but that this effect often proves to be transitory, disappearing within a year or so.

This study provides no evidence that overall response rate was affected by mode of data collection, and it is relevant to note that the response rates in subsequent years, when the survey was carried out entirely in CAPI, were not notably different from those in earlier – PAPI – years.

5. Comparison of Sample Composition

In order to carry out the analyses of mode effects on responses reported in Section 6 of this article, it was desirable to control for any differences in sample composition between the CAPI and PAPI samples. Such differences could have arisen due to sampling variance or differential nonresponse. The intention was therefore to identify any significant differences in composition between the two samples and to remove those differences by weighting.

The profiles of the two achieved samples were therefore compared in terms of demographic and classificatory variables. We must of course assume that the variables investigated are not subject to any mode-induced response effect – in other words that respondents would give the same answer to each question whether interviewed in CAPI or PAPI. Cross-tabulations reveal few differences between the two samples. Table 2 lists variables for which a significant difference is observed ($P < 0.05$). The particular categories for which the difference is significant are marked with an asterisk. Table 3 lists selected variables for which no significant difference was observed. The biggest difference

Table 2.   Comparison of the achieved sample profiles – significant differences

Variables exhibiting significant differences		CAPI %	PAPI %	Difference	Standard error
Household type	Single person 60+	9.7	7.2	2.5	* 1.1
	Single person 17–59	6.3	4.9	1.4	0.8
	2+ pensioners only	10.2	12.0	−1.8	1.2
	2 adults: at least one under 60	20.7	25.2	−4.5	* 1.6
	3+ adults: at least one under 60	19.9	17.7	2.2	2.2
	with children	33.3	33.0	0.2	2.2
Marital status	Married or cohabiting	63.5	69.6	−6.1	* 2.9
	Never married	15.7	14.5	1.1	1.0
Ethnic origin	Indian/Pakistani	3.2	1.5	1.6	* 0.7
Normally read daily newspaper	Yes	60.2	64.2	−4.0	* 2.0
Household tenure	Rented: LA/New Town	17.4	19.3	−1.9	1.9
	Rented: HA/Other	14.7	10.5	4.2	* 1.4
Household income	Less than £6,000 p.a.	24.6	21.1	3.6	2.0
(Base)		(1,320)	(1,625)		

Note: * denotes $P < 0.05$.

is in respect of household structure. The PAPI sample contains fewer people in single-person households (12.1%, compared with 16.0% in the CAPI sample). It is equally single persons aged over 60 and younger single people who are under-represented in the PAPI sample, compared with the CAPI sample. These differences are reflected in the marital status profile of the two samples. The PAPI sample contains a larger proportion of people who are married or living as married (69.6%; 63.5%). The CAPI sample also contains slightly larger proportions of people of Indian and Pakistani origin, people who do not normally read a daily morning newspaper at least three times a week, people who live in housing association or privately rented property, and people whose gross household income is less than £6,000 per annum. Variables which did *not* exhibit differences between the two samples were gender, age, activity status, housing type, car ownership, age at which education was completed, highest educational qualification obtained, socio-economic group, Registrar General's social class, and the Goldthorpe-Heath class schema.

The observations of differences in the sample composition suggest that differential unit nonresponse may be at work here. For example, CAPI may have depressed response amongst those married or living as married, while encouraging response amongst those living alone. The effects were not large, but the effect of CAPI on the distribution of non-response across sub-groups is perhaps an area worthy of further investigation. Only very limited information has been provided by studies published to date.

6. Effects on Responses to Questions

There is more than one valid way of comparing response distributions between the CAPI and PAPI samples. Here, two ways are considered. A simple comparison between the two achieved samples (each weighted to adjust for differential selection probabilities) will be

Table 3. *Comparison of the achieved sample profiles – similarities*

Selected variables exhibiting no differences		CAPI %	PAPI %	Difference	Standard error
Gender	Female	55.5	55.5	−0.1	2.2
Age	Under 35	32.1	31.3	0.8	1.6
	60+	23.8	24.3	−0.5	1.5
Activity status	Full-time paid work	51.0	50.1	0.9	1.8
	Retired	17.0	16.8	0.1	1.1
Left education	Aged 15 or less	41.9	43.4	−1.5	2.2
Housing type	Semi-detached house	29.8	34.7	−4.9	4.0
Highest ed. qualification	'A' level or above	31.6	31.0	0.6	2.5
RG's social class	I or II	21.0	21.8	−0.9	1.8
Goldthorpe-Heath schema	Working class	33.5	32.5	1.0	2.5
Household car ownership	Yes	76.0	74.5	1.5	2.4
Telephone in accommodation	Yes	90.7	90.2	0.5	2.0
(Base)		(1,320)	(1,625)		

Notes: The figures in this table have based on data that have been weighted to correct for differential selection probabilities; for each variable, cases for which a substantive answer was missing (don't know/not answered) have been excluded from the calculations. * denotes $P < 0.05$.

referred to as the *unadjusted comparison*. This will confound nonresponse effects (mode, area, fieldwork period, interviewer) with response effects (mode, area, interviewer). Most of these effects are likely to manifest themselves in the form of variance rather than bias. The exceptions are likely to be nonresponse bias induced by fieldwork period and mode, and response bias induced by mode. The unadjusted comparison might therefore provide a reasonable estimate of the *net* CAPI effect (due to both nonresponse and response), but muddied by any effect of fieldwork period on nonresponse.

The *adjusted comparison* attempts to control for nonresponse effects, so that response effects due to CAPI can be separated out. This is done by applying a weighting to eliminate the differences identified in Section 5 above between the distributions of the two achieved samples. Obviously, some of the variables which exhibit different profiles between the two samples are correlated. The adjustment weighting must take this into account, so an AID analysis was carried out, using *SPSS CHAID for Windows*. The analysis identified household size, in three categories, as the variable which provided the most powerful discrimination between the two samples. Altogether, the analysis defined ten categories which highlighted the differences between the samples. Amongst two and three-person households, newspaper readership, type of housing, and household income were important discriminants; amongst households with four or more people, the Goldthorpe-Heath class schema, and gender were important.

CAPI sample members were weighted by the ratio of PAPI respondents to CAPI respondents within each of the ten categories. This method is, of course, likely to control out any genuine effects of CAPI on nonresponse, as well as effects of interviewers, areas, fieldwork period, and sampling variance. The *adjusted comparison* therefore provides estimates of the effect of CAPI on response distributions *for a given achieved sample* (specifically, for the sample achieved with PAPI). But there is no guarantee that the same sample would actually be achieved.

Many of the *British Social Attitudes* questions require the respondent to choose an answer category from a scale. For example, the answers offered might be: ''agree strongly; agree slightly; disagree slightly; disagree strongly.'' This example is a 4-point scale, as there are four answers to choose from. The questionnaire also includes 3-point and 5-point scales.

The CAPI and PAPI respondents have been compared in terms of the responses given to each scale question. The analysis concentrated on those questions which formed part of the interviewer-administered questionnaire. Respondents were also given a self-completion booklet to complete, but the data resulting from this were not examined as part of the present study. A few questions which were largely factual in nature were excluded from the analysis, leaving only those which were deemed to be ''opinion'' or ''attitude'' questions. There were 90 such questions included in the analysis. Of these, 17 were asked of the whole sample, and the other 73 only of a random half of the sample.

Past research led us to expect that there may be differences in response for items of a sensitive nature. CAPI seems to elicit greater admission of socially undesirable behaviour (Beckenbach 1994; Baker, Bradburn and Johnson 1995; Nicholls, Baker and Martin 1997). (Though in terms of eliciting admission of sensitive behaviour, use of self-completion rather than interviewer-administered questionnaires seems to be more important than use of computer rather than paper – see Tourangeau and Smith 1996.) By extension, CAPI

may also elicit greater admission of socially undesirable *opinions*. This could manifest itself in differences in scale means. Similarly, Martin, O'Muircheartaigh and Curtice (1993) speculate that the apparent perceived greater confidentiality of CAPI may have caused the higher level of admissions of extreme views which they observed. Their study also examined the proportion of ''don't know'' responses to both knowledge and opinion questions, and found no significant differences. To address these issues, the current study compares responses to each question in three ways – first, in terms of the mean response; second, in terms of the proportion of respondents choosing one of the two extreme categories; and third, in terms of the proportion of respondents answering ''don't know.'' In addition, for 3-point and 5-point scales, the middle response category and ''don't know''s were combined to provide a comparison of the proportion of respondents giving a ''neutral'' answer. For all of these comparisons, tests of significance were carried out using complex standard errors for differences in means, calculated using the CLUSTERS programme (Verma and Pearce 1986). This takes into account the effects of stratification, clustering, and weighting on the standard errors, and uses a Taylor-expansion method of calculation.

The results are summarised in Table 4, and discussed in the following sections of this article.

6.1. Scale means

In order to compare mean responses to scale questions, each scale has been treated as an interval measure. For example, the points on a 4-point scale are assigned values of 1, 2, 3 and 4.

Under the unadjusted comparison, the mean score was significantly different ($P \leq 0.05$) between the CAPI and PAPI respondents for 15 out of the 90 questions (17%). Indeed, for 9 questions (10%), the mean was significantly different at the 0.99 level ($P \leq 0.01$), and for 4 questions (4%), the mean was significantly different at the 0.999 level ($P \leq 0.001$). Clearly, then, the number of questions exhibiting large differences in means is much greater than would be expected by chance if mode of interviewing actually had no effect on mean score. So it seems that CAPI has an effect on where respondents place themselves on scale questions. But it is not easy to identify the nature of the effect, as the nature of the questions, and the response categories offered, vary greatly. For example, CAPI respondents appear to be less satisfied with the way GP services are run, less likely to

Table 4. Differences in response to scale questions
The percentage of variables exhibiting differences between the CAPI and PAPI samples

	Scale means		Extreme categories		Don't knows		Neutral responses	
Comparison	Un	Adj	Un	Adj	Un	Adj	Un	Adj
$P \leq 0.05$	17	20	14	14	53	57	38	38
$P \leq 0.01$	10	7	6	7	39	36	27	23
$P \leq 0.001$	4	3	2	2	19	22	12	11

Base = 90 scale questions; Un = unadjusted, Adj = adjusted.

believe that hospital in-patients would only be allowed home when they were really well enough to leave, and more likely to believe that there is a lot of prejudice against protestants in Northern Ireland. CAPI respondents were also more likely to think that traffic congestion, and increased traffic in the countryside, were not problems, and were more likely to trust either a Stormont government or an Irish government to act in the best interests of Northern Ireland.

In general, therefore, it can be concluded that mode of data collection appears to have an effect on the responses to scale questions, but that there is no easily discernible pattern in the direction of the effect. It certainly does not seem to be the case that there are any differences in the question format, or the appearance of subsequent filtering, between the two modes, that could have caused such effects. The nature of the effect might warrant further research, perhaps testing explicit sociological theories of response behaviour.

Under the adjusted comparison, there are actually a greater number of questions for which the difference in means is significant at the 0.95 level (18 or 20%), although the numbers significant at the 0.99 and 0.999 levels decrease to 6 (7%) and 3 (3%) respectively. So, although some of the differences observed under the unadjusted comparison *may* have been artefacts of nonresponse effects, there appear also to be considerable CAPI-induced response effects on mean scores to scale questions.

6.2. Use of extreme categories

For 13 of the 90 questions (14%) there was a significant difference ($P \leq 0.05$) between the CAPI and PAPI samples in the proportion of respondents choosing one or other of the two extreme categories (unadjusted comparison). However, the nature of the effect was again not consistent. There were seven questions to which CAPI respondents gave more extreme answers, and six to which they gave less extreme answers. However, five of the questions (6%) were significantly different at the 0.99 level ($P \leq 0.01$), and these five were all ones to which CAPI respondents gave *more* extreme answers. Overall, across the 90 items, the mean percentage of respondents choosing an extreme category was 34.0% of those interviewed by CAPI sample and 33.7% of those interviewed by PAPI. This difference ($+0.3\%$) is not large, but if the analysis is restricted to 5-point scales, for which the extreme categories may seem more extreme than for smaller scales, the difference increases to $+1.3\%$ (across 39 items).

Under the adjusted comparison, there are again 13 questions (14%) which exhibit a significant difference at the 0.95 level, although two of these are questions which did not reach significance under the unadjusted comparison. There are now six questions (7%) significant at the 0.99 level – the same five as under the unadjusted comparison, plus one other.

Thus, CAPI appears to increase the propensity of respondents to give extreme answers to some questions. This result supports the findings of Martin, O'Muircheartaigh and Curtice (1993), who also reported that CAPI respondents are more likely to choose the extreme categories to scale questions. The explanation may lie in the suggestion above that CAPI may produce greater admission of sensitive information. Expressing an extreme opinion may be sensitive. The effect may be greater for questions addressing issues which are themselves sensitive.

6.3. Use of ''Don't know''

The proportion of respondents who answer ''don't know'' is generally small. For most of the 90 questions analysed here, less than 6% answered ''don't know,'' and for many the proportion was less than 2%, although there are a few questions for which the proportion was higher – notably a set of six questions on Northern Ireland issues, where the proportion ranged from 11% to 21% (PAPI sample figures).

CAPI respondents were much less likely than PAPI respondents to give a response of ''don't know.'' The mean percentage of ''don't know'' responses, across the 90 items, was 3.2% for the CAPI sample and 4.8% for the PAPI sample (unadjusted comparison). Under both comparisons, just over half the questions analysed showed a significant difference ($P \leq 0.05$) between the two samples in the proportion of ''don't know'' responses. For all but three of these questions, there were fewer ''don't know''s with CAPI. For the other three there were more, but these three were a single battery of consecutive questions about satisfaction with aspects of the NHS, whereas the others included a wide range of question forms and topics. So it seems clear that CAPI reduced the propensity to answer ''don't know.'' Indeed, over a third of the questions exhibited a significant difference at the 0.99 confidence level, and about one in five at the 0.999 level, with little difference between the two comparison methods.

However, nearly half the questions which exhibit a significant difference are 3-point or 5-point scale questions. It is sometimes thought that the middle category on such scales may be tantamount to a ''don't know'' response. Analysis revealed that many of these questions showed significant differences in the proportions giving the middle category response, and these differences were always in the opposite direction to the difference in the proportion of ''don't know''s. When ''don't know'' and middle category responses for 3-point and 5-point scales are combined to form a single ''neutral response'' category, two-thirds of the significant differences disappear. The result is that overall, under both comparisons, 38% of questions show significant differences between CAPI and PAPI ($P \leq 0.05$) in the proportion of neutral responses, compared with over 50% which showed differences in the proportion of ''don't know''s alone. And for all but two of these questions, the CAPI respondents gave *fewer* neutral responses.

So, it appears that CAPI tends to induce use of the middle category rather than ''don't know'' for 3-point and 5-point scale questions. But there are also 26 4-point scales for which CAPI elicits significantly fewer ''don't know''s. With these questions, it is less obvious what responses are given instead. Six of these questions exhibit a significant difference in mean response, and four show a significant difference in the proportion of extreme responses, of which only two have significantly *more* extreme responses under CAPI (one of these four is also one of the six with a difference in means). The other 17 show no differences in either of these respects. So the tendency not to use the ''don't know'' category may affect the distribution of responses to 4-point scales in some cases, but in general the ''extra'' responses seem to be distributed across the categories in much the same way as others.

One hypothesis is that some respondents perceive that the computer requires one to answer, and to answer with one of a proffered set of responses, whereas a person writing things down with pencil and paper might be more permissive towards unwillingness (or

inability) to answer in those terms. However, this finding – fewer ''don't know''s in CAPI – does not reflect that of the earlier study (Martin, O'Muircheartaigh and Curtice 1993), which found no difference. The explanation may lie in the detail of the CAPI implementation. In the earlier study, ''don't know'' was listed on screen as one of the response alternatives, with an associated code to be entered by the interviewer (usually ''8''). This is directly equivalent to the PAPI implementation on both studies. In the current study, a specific key (the left square bracket: ''['') was designated the ''don't know'' key. So ''don't know'' did not appear as one of the response options on the main screen, but the instruction ''[: Don't know'' was always displayed on an instruction line at the foot of the screen. On neither study was ''don't know'' offered explicitly to the respondent as a possible response. Though this subtle difference between the two implementations could not have affected the answers given, it may have influenced some interviewers' probing and coding of vague answers. Two other studies (Olsen 1992; Bergman et al. 1994) have suggested that small changes in the appearance to the interviewer of a question and its answer categories could induce apparent mode effects. Indeed, Bergman et al. (1994) present an example where a change in layout of answer categories appears to have produced a decrease in the proportion of ''middle-category'' answers – not dissimilar to the current finding. Perhaps this provides a salutary reminder of the sensitivity of data to the exact nature of the questioning and recording. This applies to any kind of survey, but a change of data collection mode may inadvertently provide extra opportunities for questioning and recording methods to change.

7. Other Effects

7.1. Interview length

Couper, Groves and Kosary (1989) and Martin, O'Muircheartaigh and Curtice (1993), whose fieldwork took place in 1989 and 1990, both found that interviews took longer in CAPI. However, Weeks (1992) suggests that ephemeral factors, which may now no longer apply, may have caused ''early research'' (not cited) to find that CAPI increased interview length. Baker, Bradburn and Johnson (1995) reported a considerable *reduction* in interview length with CAPI. The current study also found that interviews were quicker with CAPI. In general, CAPI's potential to reduce interview length may depend on the nature of the survey. The ability to define and record equivalent measures of interview length in the two modes is also an issue. There are particular reasons why the *British Social Attitudes* interview might be speeded up by CAPI. There are very few textual answers to be entered. Also, CAPI renders unnecessary some check questions which require (PAPI) interviewers to undertake careful coding, often involving referring back to earlier pages.

The mean length of interview, as recorded by interviewers, was 45.5 minutes in CAPI, and 54.0 minutes in PAPI. 13% of CAPI interviews took 30 minutes or less, compared to 2% of PAPI interviews; at the other extreme 11% of CAPI interviews and 16% of PAPI interviews took over an hour. However, there may be a mode effect on the way interview length is recorded. Interviewers using PAPI wrote in the time at the start and end of the interview. From this they calculated and recorded interview length. Data from many

surveys show that interviewers tend to enter times as multiples of five minutes. CAPI interviewers were also asked to key interview length, but embedded in the text of the instruction were the start and end times (as logged by the computer's internal clock – interviewers were not asked to enter start and end times), so most interviewers may have simply entered the difference between the times displayed. This surmise appears to be backed up by the very high degree of agreement between the interview length recorded by the interviewer and the length calculated internally by the computer. Consequently, it seems likely that the CAPI interview lengths are more accurately recorded than the PAPI lengths. It should also be recognised that interview length is not a perfectly-defined concept. With PAPI it is possible for interviewers to enter the start time (on the front of the questionnaire) as soon as the sample member consents to the interview. Some minutes may then pass – for example while seating and survey materials are arranged – before the first question is asked. The CAPI interviewer, on the other hand, cannot enter the start time until he or she has switched on the computer, loaded the programme, and moved to the first question. Indeed, this is one of the advantages of CAPI – it gives the researcher more control over the interview situation.

Thus, the observed difference of 8.5 minutes may be an over-estimate of the time saving induced by CAPI, but it does seem likely that CAPI brings *some* saving in time. Of course, we cannot be sure whether a shorter interview, with less time spent on each question on average, is good or bad news from the point of view of validity and reliability of responses.

7.2. Respondents' willingness to be reinterviewed

There was no evidence that the experience of being interviewed in CAPI had an adverse effect on respondents' willingness to participate in surveys again. At the end of the interview, all respondents were asked whether they would mind if they were included in a similar survey in a year's time: 89.7% of CAPI respondents, and 87.6% of PAPI respondents said that they would not mind. Indeed, this suggests that CAPI might actually increase the proportion of respondents willing to be reinterviewed (two-tailed $P = 0.08$, although this does not allow for any systematic interviewer effects), a finding which is consistent with Martin, O'Muircheartaigh and Curtice (1993).

7.3. Interviewer and respondent reactions and survey administration

Interviewers generally reacted favourably to CAPI, and reported that respondents did likewise. Office administration procedures were very different with CAPI, but there were no major problems once staff had got used to the new systems. Also, significant savings were made in data keying and editing. These are now widely accepted characteristics of CAPI (Nicholls, Baker and Martin 1997) and do not require further discussion here.

7.4. Costs

The effects of interviewing mode on data quality and survey administration should not be considered in isolation from survey costs. Marginal improvements may not be worthwhile if the price is too high. Conversely, marginal disadvantages may be acceptable if a significant cost saving is achievable (as this saving could be re-deployed to improve data quality

in other ways). CAPI requires initial expenditure on hardware and software, and other set-up costs. But it is no longer necessary to print, distribute, and store paper questionnaires. And there will also be significant savings in data keying and editing. The extent to which the extra initial costs have to be borne by the individual survey in question, or can be spread across a number of surveys, will be organisation-specific. But in general, large-scale continuing surveys are the most likely to benefit from net savings under CAPI (Martin and Manners 1995). One-off surveys (term common in North America is one-time survey) are the least likely to benefit from CAPI, particularly if the sample size is small. The *British Social Attitudes* survey series falls between these two extremes: it is regular, but not continuing, and has a medium-sized annual sample. In 1993, the use of CAPI almost certainly increased survey costs. But net savings may accrue in future years, as routines become established, and questionnaire modules will not require re-programming. It is expected that, within an institution that has invested heavily in CAPI, the long term survey running costs should be no more with CAPI than with PAPI. This will be monitored.

8.   Conclusions

In some respects, CAPI clearly brings improvements to the survey process and to the survey data. The occurrence of item missing data, due to the interviewer inadvertently omitting to ask the question, or to record the response, can be eliminated. Some administrative procedures and the process of data handling and editing, can be greatly simplified.

On the other hand, the analysis presented here appears to show clear CAPI/PAPI response effects on attitudinal scale questions. CAPI increases the propensity to give ''extreme'' responses, and there are common effects on mean scores. Also, CAPI may reduce the propensity to give ''don't know'' answers, though this could be an effect of the particular CAPI implementation on this survey. For 3-point and 5-point scales, where the proportion of ''don't know''s is reduced, the proportion of middle-category responses is often simultaneously increased. But overall there still remains a reduction in the proportion of neutral answers given.

However, although these effects are statistically significant, and we may therefore judge that they are probably ''real,'' this in itself does not imply that the effects are necessarily undesirable. There are two issues here. First, even if it has been shown that the two modes produce *different* response behaviour, there is no evidence as to which mode has greater reliability or validity. It should not be assumed that PAPI is ''right,'' just because that is the data collection mode in which researchers have put their faith in the past. The second issue is that of whether the observed mode effects would make any difference to the conclusions drawn by analysts of the data. This issue is particularly relevant where data collected by the two modes are combined into a single analysis, for example when comparing data collected in two different years, where PAPI was used in the first year and CAPI in the second. The strongest mode effect is on the proportion answering ''don't know,'' but the proportion itself is usually very small, so the average propensity to answer ''don't know'' would have to increase many-fold before the substantive categories were significantly affected. The ''don't know'' responses are in any case excluded from many analyses, on the grounds that they are not of substantive interest. Data users should

be aware of possible mode effects on analyses where the proportion of ''don't knows'' could affect the results, but it seems that few analyses fall into that category. What remains, then, is a concern that a change in data collection mode may artificially alter scale means, or the proportion giving an answer on one particular side of a scale. There is some evidence that scale means may be affected by data collection mode, and this is something of which researchers should be aware.

Other findings of the current study were a shorter interview length in CAPI, no significant difference in survey response rate, and, if anything, a slightly greater willingness to be reinterviewed after a CAPI interview. The latter two findings are in accordance with those of earlier studies. The conclusion regarding interview length is that the effect of CAPI may be highly dependent on the nature of the interview.

9. References

Baker, R.P. (1990). What We Know About CAPI: Its Advantages and Disadvantages. Paper presented at the American Association of Public Opinion Research Conference, Lancaster, PA, 17–20 May.

Baker, R.P., Bradburn, N.M., and Johnson, R.A. (1995). Computer-assisted Personal Interviewing: An Experimental Evaluation of Data Quality and Cost. Journal of Official Statistics, 11, 413–431.

Beckenbach, A. (1994). Computer-assisted Questioning: The New Survey Methods in the Perception of the Respondents and Methodological Implications on Data Quality. Paper presented at the World Sociology Congress (Sessions of the Research Committee on Logic and Methodology), Bielefeld, Germany, 18–23 July.

Bergman, L.R., Kristiansson, K.-E., Olofsson, A., and Säfström, M. (1994). Decentralised CATI Versus Paper and Pencil Interviewing: Effects on the Results in the Swedish Labour Force Surveys. Journal of Official Statistics, 10, 181–195.

Couper, M., Groves, R.M., and Kosary, C. (1989). Methodological Issues in CAPI. Proceedings of the American Statistical Association, Section on Survey Research Methods, 349–354.

Jowell, R., Curtice, J., Brook, L., and Ahrendt, D. (1994). Technical Details of the Survey. British Social Attitudes: The 11th Report, Jowell et al. (eds.), Aldershot: Dartmouth Publishing Company.

Kish, L. (1949). A Procedure for Objective Respondent Selection Within the Household. Journal of the American Statistical Association, 44, 380–387.

Lynn, P. (1992). Applications of Computers in Sample Design and Selection. In Survey and Statistical Computing, A. Westlake et al. (eds.), Amsterdam: Elsevier.

Lynn, P. and Lievesley, D. (1991). Drawing General Population Samples in Great Britain. London: SCPR.

Manners, T. (1987). The First Laptop Computer Project in SSD. Survey Methodology Bulletin, 21, 25–34.

Martin, J. and Manners, T. (1995). Computer Assisted Personal Interviewing in Survey Research. In Information Technology for the Social Scientist, R. Lee (ed.), London: UCL Press.

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, 9, 641–661.

Nicholls, W.L., Baker, R.P., and Martin, J. (1997). The Effect of New Data Collection Technologies on Survey Data Quality. In Survey Measurement and Process Quality, L. Lyberg et al. (eds.), New York: Wiley and Sons.

Olsen, R.J. (1992). The Effects of Computer-Assisted Interviewing on Data Quality. European Science Foundation Network on Household Panel Studies, Working Paper no. 36.

Sebestik, J., Zelon, H., DeWitt, D., O'Reilly, J., and McGowan, K. (1988). Initial Experiences with CAPI. Proceedings of the Fourth Annual Research Conference, U.S. Bureau of the Census, 357–365.

Tourangeau, R. and Smith, T.W. (1996). Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Content. Public Opinion Quarterly, 60, 275–304.

Van Bastelaer, A., Kerssemakers, F., and Sikkel, D. (1988). Data Collection with Hand-held Computers: Contributions to Questionnaire Design. Journal of Official Statistics, 4, 141–154.

Verma, V. and Pearce, M. (1986). CLUSTERS: A Package Program for the Computation of Sampling Errors for Clustered Samples. International Statistical Institute Research Centre Technical Report no. 131, ISI, Voorburg.

Weeks, M.F. (1992). Computer-Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications for Survey Operations. Journal of Official Statistics, 8, 445–465.