The Reliability of Coding Occupational Descriptions: measurement issues in a CAPI panel survey

Heather Laurie

and

Nick Moon
NOP World

ISER Working Papers
Number 2003-13

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Acknowledgement:

Readers wishing to cite this document are asked to use the following form of words:

> **Laurie, Heather and Moon, Nick. (2003) 'The Reliability of Coding Occupational Descriptions: Measurement Issues in a CAPI Panel Survey', *Working Papers of the Institute for Social and Economic Research*, paper 2003-13. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

This paper reports the results of a re-coding experiment carried out using occupational descriptions from the 1999 round of the British Household Panel Survey (BHPS) and the impact of CAPI screen design for the longitudinal consistency of coding occupational descriptions. The BHPS is a household panel of some 5,500 households with an annual round of data collection since 1991. The BHPS is conducted by the Institute for Social and Economic Research based at the University of Essex, UK and NOP Research Group, London. A source of error in any survey is to be found in the coding of verbatim responses post fieldwork. In a longitudinal panel survey, the concern is not only with the reliability of coding at one point in time but also with reliability over time, where respondents may report the same job in different ways at different years of the survey. In a CAPI survey, screen design may also influence what is recorded by interviewers with subsequent effects for coding reliability.

The screen design for collecting occupational descriptions on the BHPS provides two text boxes for interviewers to record the verbatim response. Some interviewers interpret this screen as one box for the job title and the other for the job description while others use one text box only for both elements. Couper and Conrad (2000) have recently reported that the reliability of occupational coding is affected by an interaction between the length of the verbatim response recorded by interviewers and the difficulty of language used in the response. Using 1500 occupational descriptions from the BHPS, we first coded the responses using the job title recorded by the interviewer in the first text box only. We then double-coded the same cases using both the job title and the full description of the work involved recorded by the interviewer in the second text box. The analysis compares the cross-sectional coding results for the 'short' and 'long' descriptions. It then assesses the longitudinal consistency of the occupational coding for respondents who have not changed jobs between years of the survey for the whole sample and for the cases with 'short' and 'long' descriptions. Two propositions are tested. First, that the longitudinal consistency will be greater where long descriptions were used for coding and secondly that there will be no significant differences in the longitudinal consistency of coding between interviewers using one text box only and those who split the responses over two text boxes. The implications for interviewer training and CAPI screen design to reduce measurement error are discussed.

1. Introduction

The British Household Panel Survey is a national household panel survey of over 10,000 individuals in some 6,000 households in Britain which is carried out by the Institute for Social and Economic Research (ISER) based at the University of Essex. ISER is an interdisciplinary academic social research unit primarily funded by the Economic and Social Research Council of the UK. One of the main tasks of ISER is to manage the BHPS and provide high quality panel data for the wider user community including academic researchers, government departments and independent social policy research institutes. The interviews for the BHPS are conducted by NOP Research, a research organisation based in London, UK who are contracted by ISER to carry out the fieldwork and initial editing and coding of the data. The BHPS commenced in September 1991 and returns to re-interview panel members on an annual basis using face to face interviews. The sample was drawn from the small users file of the Postcode Address File and covers non-institutional residences in England, Wales and Scotland. The panel is currently preparing for the twelfth round of data collection. This paper uses data from waves 8 and 9 of the BHPS collected in 1998 and 1999.

2. Coding of occupations

The original title of the centre that runs the BHPS was the ESRC Research Centre on Micro-Social Change in Britain, and this shows the importance of social change to the survey. This is not the place for a discussion on the importance of employment-related social class to the status and behaviour of households, but it is undoubtedly the case that occupational mobility is a very significant factor in changing household circumstances. Occupational mobility is relatively easy to measure when people change employer, but is more problematic when people remain with the same employer but change the nature of the work they do. This means that the coding of respondent occupation is a vital part of the survey.

Occupational coding schemes have changed in Britain over the years, but the principal form of occupational coding throughout the life of the BHPS has been the Standard Occupational Classification (SOC), developed by the UK government's social survey arm. It is a hierarchical 3-digit coding system, with occupations classified into ten major groups, each of which is sub-divided into up to ten minor groups, with the minor groups being sub-divided into the final codes.

Coding of SOC was a laborious process, involving looking up job titles in the index of a reference manual, and then cross-referencing the resulting code against other factors such as employment versus self-employed to produce the final code. Since the reference books were not updated very frequently, there was always a considerable chance that some jobs would simply not be found in the index.

The basis of the information requirement for SOC is a pair of open-ended questions asked of respondents:

"What is your exact job title?"
"Could you please tell me what your job actually involves?"

Respondents who are not working are often asked about their last main job. There are also other questions that may be used to determine the exact SOC code, such as whether the respondent is employed or self-employed, the industry the respondent works in, whether the respondent has management or supervisory responsibilities, and so on.

NOP has for many years had a specialist team of coders responsible for most of NOP's occupational coding, including industrial classification, and they have built up a very high level of familiarity with the processes.

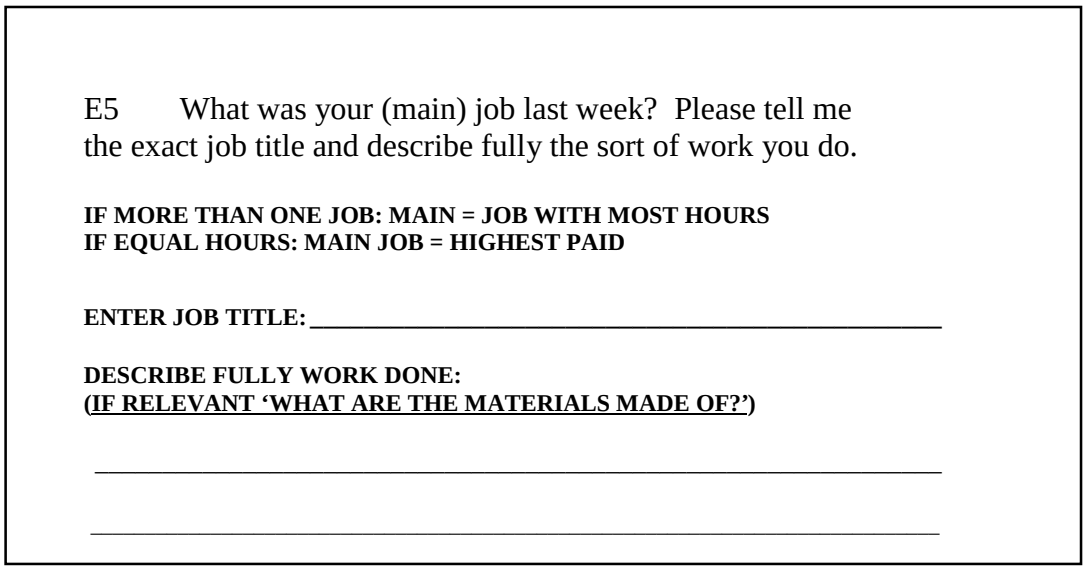
In the 1990s Peter Elias and colleagues at the University of Warwick developed a computerised method of conducting SOC coding, called Computer Assisted Standard Occupational Classification, or CASOC. This involves the coders entering the text recorded by the interviewer at the verbatim question onto a PC. The computer then interrogates the program's built-in database of job titles, and offers one or more possible codes to the coder. The coder is also offered more information about what the offered codes mean, and has the opportunity to move up, down, or sideways in the hierarchical system, looking not just at alternative codes, but also at the job descriptions that go with them. In principle this is not very different from what the coders did before in leafing back and forth through the coding manuals, but the use of the computer, and in particular its ability to make intelligent judgements, makes the coders' task in practice much easier.

The other key aspect of CASOC is that it does not merely provide the 3 digit SOC code, but also generates a further 3 characters. These additional characters can be used by the CASOC program to generate codes in alternative coding schemes such as Hope-Goldthorpe, KOS and others.

3. CAPI screen design and survey error

For the first eight years, the BHPS was conducted using pen and paper questionnaires. Because it was known from the start that broadly the same questionnaire would be used every year, a very high level of effort was put into making it as easy as possible for interviewers to use. There is much more white space on the page than is customary in most questionnaires, and interviewers consistently said how much they enjoyed using the BHPS questionnaire compared with other surveys they worked on. Below is an example of the part of the questionnaire involving the respondent's current employment. It will be noticed that the two separate questions - one on job title and one on work done - run together in the questionnaire, while there are two physical spaces to write in the answers. Interviewers were given a clear view of their task, because the first answer space was labelled "Enter job title", and the second came with an instruction "Describe fully work done"

Figure 1 – Paper Questionnaire Main Occupation Question



E5 What was your (main) job last week? Please tell me the exact job title and describe fully the sort of work you do.

IF MORE THAN ONE JOB: MAIN = JOB WITH MOST HOURS
IF EQUAL HOURS: MAIN JOB = HIGHEST PAID

ENTER JOB TITLE: _____

DESCRIBE FULLY WORK DONE:
(IF RELEVANT 'WHAT ARE THE MATERIALS MADE OF?')

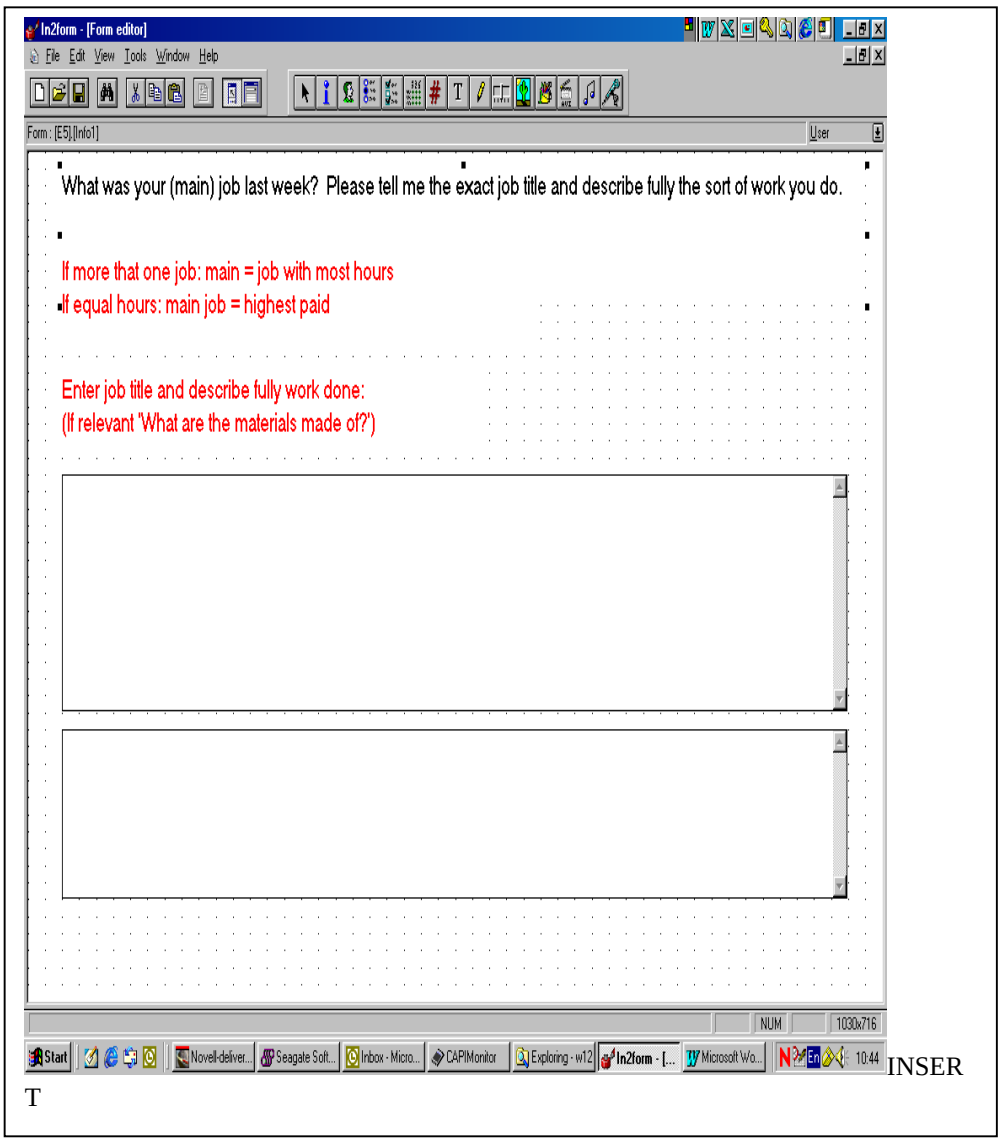
_____

_____

In year 9, after a lengthy period of testing, the BHPS switched to Computer-Aided Personal Interviewing (CAPI), using the In2uitive software. As well as the usual principles if usability, one of the key principles of the development of the CAPI script was that, since the same interviewers tended to work on the survey year after year, the CAPI questionnaire should look on screen as much like it did on paper as possible. The screen-shot below from the CAPI version of the same employment question shows how similar it is to the paper version, but also shows how it differs in one key respect.

The second main guiding principle of the questionnaire design process was that there should be a constant "look" to the questionnaire. Too often CAPI questionnaires – like Web questionnaires – are inconsistent in terms of completing the task: the answer categories may appear in different positions on the screen, or the action necessary to move to the next question may vary. We therefore tried to ensure that the question wording always appeared at the very top of the screen, with the answer categories below, with the exact layout determined by whether it was a precoded or write-in, whether it was single answer or multiple answer and so on. This is why the labels that appeared alongside the write in spaces on the paper questionnaire were not present on the CAPI screen.

Figure 2 – CAPI Questionnaire Main Occupation Question

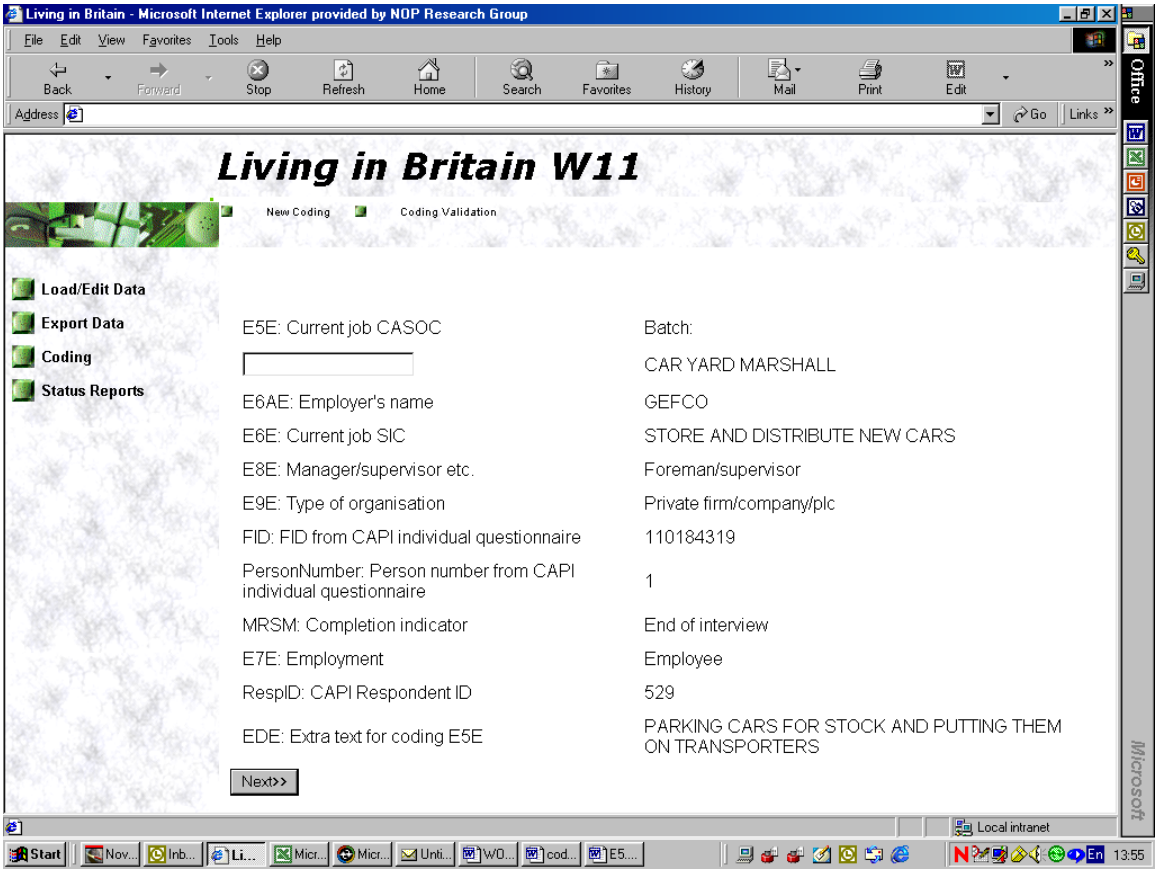


Interviewers were thus faced with a two-part question, and two text boxes to write in, but no specific instruction on how to do so. As is explained below, they proved to use them in very different ways, which proved to be problematic for the coding process.

Aside from the questionnaire design issues, the switch to CAPI also raised issues for the coding process. We had realised from the start that there were potential problems with on-line coding of data fed straight from the CAPI program, in comparison with coding of paper-based questionnaires. Coders working with a paper questionnaire can see the whole page on which the answer was written, which can provide useful contextual information, and can if necessary look through the whole of the rest of the questionnaire to look for information that may help them find the right code.

On-line coders see only the verbatim text written in at the open-ended question itself, without any contextual information at all, and this can make their task more difficult. IN developing a bespoke coding system for the BHPS, we tried to minimise the difficulties by providing contextual information on the same screen as the verbatim text. Below is a screen shot from the coding screen for the current occupation question.

Figure 3 – Occupation Coding Screen



Another of the worries about CAPI when it was first widely introduced into surveys, especially in the case of keyboard-free handheld computers, was that interviewers would

write less at open-ended questions when they had to use an unfamiliar keyboard, or an even more unfamiliar touch screen, than when they used to write longhand on a piece of paper.

When NOP first switched to CAPI we ran a number of parallel tests, one of which was to compare the number of words on regular open-ended questions carried on the NOP Omnibus survey. This showed that there was virtually no difference between paper and CAPI, and was a significant factor in the decision to make the move to CAPI.

Similarly, when the BHPS moved to CAPI, the average number of words recorded by interviewers on verbatim items was similar to that recorded on paper. We were particularly concerned that interviewers may truncate responses on occupational descriptions which could potentially lead to differences in the coding of SOC.

While we observed no significant differences in the average number of words used by interviewers compared to the paper questionnaire, we did observe an unexpected difference in the way that interviewers used the screen for recording occupational descriptions on the CAPI questionnaire. Some interviewers used the two boxes we had provided as we had expected with the second box being used only when the description exceeded the 250 characters allowed by the first box. In the majority of cases the 250-character limit in the first box was sufficient and the second box was not used at all. Other interviewers interpreted the two boxes as requiring the job title to be entered in the first box and the description of what was done in the job in the second box.

There was considerable variation by interviewer in the use of the boxes. Of 236 interviewers, 105 used the second box at least once but there is a core of 34 interviewers who use it for most of their cases. It should, therefore be borne in mind that the majority of recodings involving the long form will be drawn from a relatively small proportion of all interviewers, and there is almost bound to be some interviewer effect.

We realised this only through a compounding error where the programmer responsible for setting up the on-line coding system output only the text from the first box on screen for coding. This meant that a significant proportion of the occupations were only being coded with a minimal two or three word job title and no explanatory information. The error was spotted while coding was being completed and rectified but we used this as an opportunity to compare the cases where interviewers had used one box only to record the occupational description and those where two boxes had been used. We also carried out a series of recoding exercises to determine the reliability of the coding depending on the length of the description and how the information had been recorded by the interviewer, that is using one box only or using the two boxes on the CAPI screen.

We expected that the coding where two boxes had been used would be more reliable in a code/recode test as the coder would, on average, have more information to use during coding. In addition to the cross-sectional reliability of coding, we also expected that the longitudinal consistency of occupational codes for people who had not changed job in the

year between interview points would be greater where two boxes were used by interviewers.

4. Comparing short and long occupational descriptions

The average number of words used by interviewers when recording the occupational description was seven words. Across all responses recorded, this average remained the same regardless of whether they used one box only or spread the description over two boxes on the screen. Comparing the interviewers who used one box only and those who used two boxes, we see that interviewers using both boxes entered descriptions that were on average two words longer than did those using just one box. For the interviewers who used both boxes, the description in the first box was typically very short, mostly just the job title of a few words. The average number of words entered in box one for these cases was just three words. When the error in not outputting the data for coders was made, this means the coders were using just the job title for coding with no or limited explanatory information about the job.

Table 1 Average number of words entered by number of boxes used

	Average number of words entered
Interviewer used box 1 only	7
Interviewer used box 1 and box 2	9
All box 1 responses	6
Box 1 for those also using box 2	3
All box 2 responses	6
All responses box 1 and box 2	7

4.1 Cross-sectional code/recode tests

The first recode we carried out was a straightforward recode across all cases. The responses from the second box were used in the 1149 cases where additional information was present for the recode only, with the original code using the information in the first box only. In terms of the overall sample, the reliability of the code/recode is as expected, producing an overall difference of 5.4% at the three digit level. Table 2 shows the comparison of SOC codes by the recode at the one digit level, with relatively high levels of reliability across all occupational groups.

Table 2 Occupational code by recode, all cases at one digit level.

	Recode (%)									
SOC code	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
1	**97.1**	0.3	0.7	0.7	0.1	0.4	0.4	0.2	0.5	*886*
2	0.6	**97.9**	1.2	0.1	0.3	0.1	--	0.2	1.6	*592*
3	0.8	1.2	**96.0**	0.7	0.1	0.4	0.2	0.4	0.9	*689*
4	1	0.3	0.6	**97.5**	0.1	0.3	1	0.9	0.5	*1038*
5	0.2	0.2	0.6	0.2	**98.4**	--	0.6	1.6	0.9	*700*
6	--	--	0.4	--	--	**98.2**	--	0.2	0.2	*729*
7	0.2	--	--	0.7	--	--	**97.3**	0.2	--	*483*
8	0.1	--	0.4	--	0.7	0.1	--	**96.5**	0.7	*561*
9	--	--	--	0.2	0.1	0.4	0.4	--	**94.7**	*539*
Total	*888*	*577*	*682*	*1033*	*685*	*737*	*486*	*567*	*561*	*6217*

When we split the sample into those where only one box was used by the interviewer and those where two boxes were used, there is only a 1% difference for those using one box only and a 26.5% difference for those using two boxes when we code to the three digit level. The reliability for the cases where one box only was used is one hundred percent at the one digit level across all SOC categories. The cases where the interviewer used both boxes are far less reliable on the recode than for the sample as whole. It seems that the variability in coding across the whole sample is due entirely to the cases where two boxes were used (see table 3).

Table 3 Original SOC code/recode by number of screens used (one digit level)

		Recode (%)																	
		1		**2**		**3**		**4**		**5**		**6**		**7**		**8**		**9**	
	N boxes	***1***	***2***	***1***	***2***	***1***	***2***	***1***	***2***	***1***	***2***	***1***	***2***	***1***	***2***	***1***	***2***	***1***	***2***
SOC																			
1		**100**	**84.6**		109		3.8		3.5		0.9		2.3		2.1		6		3
2			3.0	**100**	**88.7**		6.1		0.5		1.8		0.8		--		1		8.9
3			4.1		--	**100**	**79.5**		3.5		0.9		2.3		1		2.1		5
4			5.3		6.6		3.0	**100**	**86.9**		0.9		1.5		5.2		5.2		3
5			1.2		1.9		3.0		1	**100**	**89.9**		--		3.1		9.4		5
6			--		0.9		2.3		--		--	**100**	**90**		--		1		1
7			1.2		--		--		3.5		--		--	**100**	**86.6**		1		--
8			0.6		--		2.3		--		4.6		0.8		--	**100**	**79.2**		4
9			--		--		--		1		0.9		2.3		2.1		--	**100**	**70.3**
Total		*719*	*169*	*471*	*106*	*550*	*132*	*835*	*198*	*576*	*109*	*607*	*130*	*389*	*97*	*471*	*96*	*460*	*101*

Note: All of the cell sizes off the diagonal are less than 10 cases.

Given the higher levels of variability for the cases where the interviewer used two screens, we might conclude that the additional information in the longer description in itself leads to the coder having to make finer level distinctions and more decisions about the choice of the correct code. The more decisions a coder has to make the greater the scope for differences in interpretation which might affect the final code chosen. To try and disentangle what was happening during the coding of these longer descriptions, we carried out a second recode exercise on approximately half of the cases where two boxes were used.

For 600 of the cases where interviewers had used both screens to record the occupational description, a second re-coding exercise was carried out. For these 600 cases we therefore have the original code done using just the first box short description and a code / re-code validation using the long description from both boxes. This allows us to make comparisons of the difference between codes set out in Table 4 below.
In all comparisons, the percentage difference is greater at the three-digit level than at the one digit level as would be expected. We first recoded the cases where two boxes had been used but using just the text in the first box. At the three-digit level there was a difference of 21% and at the one digit level this fell to 12%. Long recode 1, between the original code and the recode using both boxes, shows a 25.4% difference at the three-digit level and a 15.2% difference at the one digit level. Long recode 2 shows a 20.6% difference at the three-digit level and a 12% difference at the one digit level. The greatest difference is with the code/recode test between the two long descriptions where two boxes were used by the interviewer. On this recode, there is a 31.5% difference at the three-digit level and an 18.5% difference at the one digit level

Table 4: Difference between short and long re-codes where two screens used by interviewer

	% difference 3 digit level	% difference 1 digit level
Short code/short recode	21.0	12.0
Short code/ long recode 1	25.4	15.2
Short code/ long recode 2	20.6	11.9
Long recode 1/long recode 2	31.5	18.5

Table 5 shows the code/recode at the one digit level. It is noticeable that some occupational categories are more reliable on the recode than others, suggesting that the types of jobs described were unambiguous and relatively easy to code correctly. Occupations in major group 5 (craft and related occupations) and major group 6 (personal & protective service occupations) had a code/recode reliability at the one digit level similar to the sample as a whole. The lowest reliability was in major group 1 (managers and administrators). The other occupational categories once again a low recode reliability compared to the sample as a whole. The longer descriptions where the

interviewer had used two boxes were clearly less reliable on the recode test. However, the question remained as to whether this was due to the longer, more complex description itself or to some form of interaction between the information recorded by the interviewer and how this was used during coding by the coders.

Table 5 SOC code/recode at the one digit level for cases where two screens used.

	SOC Recode (%)								
SOC code	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	**81.3**	6.6	4.5	3.7	1.6	--	6.3	1.9	--
2	3.6	**86.9**	--	1.9	--	--	--	--	--
3	4.5	6.6	**85.1**	3.7	1.6	--	--	1.9	--
4	4.5	--	1.5	**90.7**	--	--	4.2	--	2.4
5	3.6	--	3	--	**95.2**	--	--	7.7	--
6	0.9	--	3	--	--	**98.1**	--	--	--
7	0.9	--	1.5	--	--	--	**89.6**	--	4.9
8	--	--	1.5	--	1.6	1.9	--	**84.6**	7.3
9	0.9	--	--	0.9	--	--	--	3.8	**85.4**
Total	*112*	*61*	*67*	*108*	*63*	*54*	*48*	*52*	*41*

Note that the cell sizes are small with all off diagonal being less than 10 cases.

We then looked at the actual descriptions recorded for the 179 cases where differences had occurred between the original SOC code and the first recode exercise for the cases where information was recorded in both boxes. We hoped to try and establish which of the codes the coders had used was the correct code or the 'best' code given the available information.

It is clear that the screen design where the interviewer could choose to use either one box or two for entering the occupational description had an affect on the recode reliability of the occupational coding. In some cases, the additional information provided by interviewers led to an improvement on the original SOC code as the duties of the job were clarified, allowing the correct code to be assigned. In other cases, the additional information in the second box did not lead to a better code but seems to have introduced some ambiguity into the coding process.

4.2 Longitudinal comparisons

The second area we wanted to explore was what the effect of having used either one box or two would be on the longitudinal consistency of occupational coding between waves of the panel survey. In theory, those who have not changed jobs in the year between interview points, should remain in the same occupational group even though in reality this is often not the case. Respondents may describe the job they are doing differently at

each year, interviewers may vary in the amount of information they record, and there will be some variability between coders.

We selected the cases in the 1999 round of the BHPS who reported no job change in the previous year and were interviewed at both 1998 and 1999. For these cases, there was a 31.5% difference at the three-digit level between the code assigned in 1998 and the code at 1999 while at the one digit level, this fell to 18.7%. These rates of difference are seen at each pair of years of the BHPS as the occupational description is collected at each year rather than being fed-forward as a prompt (see Table 6).

Table 6: Comparison 1998 and 1999 SOC codes for those with no job change

	1999 SOC (%)									
1998 SOC	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
1	**80.5**	5.2	4.1	5.3	1.7	2.2	5.1	1.1	1.6	*445*
2	3.9	**83.3**	8.5	1.4	1.7	1.1	--	1.8	6.1	*353*
3	5.3	6.5	**76.6**	2.7	2.3	1.9	0.5	1.1	2.8	*347*
4	5.3	1.5	4.4	**84.9**	2.6	1.6	6.0	1.4	1.6	*574*
5	1.4	1.2	1.2	0.7	**78.1**	0.3	1.4	9.9	0.8	*288*
6	0.5	0.9	2.3	0.3	0.3	**85.8**	--	0.7	5.3	*344*
7	2.7	0.6	0.3	3.1	0.7	1.6	**81.9**	0.7	1.6	*224*
8	0.2	0.6	2.0	0.9	10.3	0.8	1.4	**78.7**	2.4	*280*
9	0.2	--	0.6	0.7	2.3	4.7	3.7	4.6	**77.6**	*243*
Total	*437*	*324*	*342*	*584*	*302*	*365*	*216*	*282*	*246*	*3098*

When we compare the 1998 code with the 1999 recoded occupations for these cases the same difference at the one digit level (18.7%) is found as with the initial 1999 code. Given the higher levels of difference of the recode for the cases using two boxes in the 1999 data, we might also have expected to see higher levels of difference in the longitudinal consistency of coding for those where two boxes and longer descriptions were used at 1999. As can be seen in table 7 there was no difference in the longitudinal consistency of the occupational codes depending on whether one box or two was used by the interviewer at the second interview in 1999. For the longitudinal comparison, the length of the description does not appear to have any effect at all. It may be the case that any effect is swamped by the differences that already exist due to the way that respondents describe their job to interviewers when re-interviewed in 1999. (Table 7)

Table 7: Longitudinal consistency of SOC at one digit level by number of boxes used in 1999 (respondents in same job interviewed at both years)

	One box at 1999	**Two boxes at 1999**
	%	%
1998/9 codes		
Same	81.7	80.2
Different	18.3	19.8
1998 code/1999 recode		
Same	81.6	80.2
Different	18.4	19.8
N	*2487*	*592*

5. Implications for interviewer training and CAPI screen design

One conclusion to be drawn from this exercise which has clear implications for interviewer training, is that more information is not necessarily better for the reliability of occupational coding. Simply adding extra information can confuse rather than help the coder unless it is relevant information that helps clarify the description and lead to more accurate coding. We have spent some time at interviewer briefings trying to explain to interviewers how difficult the coders' task is when coding occupations, and have given examples of some of the fine distinctions that coders need to make, and of job descriptions that are clearly inadequate. These exhortations to interviewers have tended to be in the form of encouragement simply to write longer descriptions. It is now clear that we need to stress the need to include only information that is relevant, and which helps to distinguish the respondent's job from another, similar, one.

As far as screen design in a CAPI questionnaire, the design of the screen can lead interviewers to use it in different ways. In this case the screen design does seem to have contributed an element of unreliability, at least as far as the code/recode of occupations at the cross-sectional level. In trying to improve data quality by allowing plenty of space for interviewers to enter enough detail for coders, we inadvertently introduced greater variability.

It is likely that this is a combination of the length of the description, ambiguities within the description, and what the coder chose to read on their screen when the text was displayed for them. If it is the case that the coders are more likely to use just the first box information for coding even when there is additional information available on their screen, then they will be coding using a very short description of an average of three words. This in itself could lead to the wrong code being assigned. If on the recode, the

second coder used all the available information then you would expect a fairly high level of difference between codes.

It is thus clear that it was unreasonable to expect all interviewers to behave the same way when presented with two text boxes and instructions of the kind used so far on the CAPI implementation of the BHPS. We therefore need to add more explicit questions to make it absolutely clear what we want interviewers to write where.

6. Conclusion

Both of the propositions we set out to test in the abstract are not proven. The proposition that coding reliability would not be affected by the number of boxes used by interviewers is not supported by the data. The variability for the cases where interviewers used two boxes rather than one is greater than where one box only was used. The second proposition, that the longitudinal consistency would be greater where two boxes were used, was also not supported. We found no difference for the longitudinal consistency of the occupational codes.

When we first discovered the ambiguity in coding, our first thought was that if we could instruct interviewers better on what to write where we would get a consistent approach, and this would in turn lead to a more consistent approach from coders. At the very least, by combining the answers from the two boxes into a single variable, the splitting of responses across one or two boxes would become irrelevant, as the coder would see the same information either way.

However, the somewhat counter-intuitive findings revealed in this paper suggest we should be turning our attention elsewhere. If the addition of extra information can make coding more variable, rather than less, then the problem seems to lie in the nature of the information being collected. We should therefore turn our attention to better interviewer training on the kinds of information that is helpful to coders.

Bibliography

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, Sage Publications Inc.

Campanelli, Pamela C., Moon, N.(1994) Computer aided occupational class coding *Paper presented at the 1994 Conference of the World Association for Public Opinion Research, Danvers, Massachusetts*

Couper, M. And Burt, G. (1994) 'Interviewer attitudes toward computer-assisted personal interviewing (CAPI) ' *Social Science Computer Review*, 12 (1), 38-54.

Conrad, F.P., Couper, M.P (2000) Coding open-ended reports: classifying occupation in the Current Population Survey. *Proceedings of the Fifth International Conference on Logic and Methodology, Cologne, 2000*

Edwards, B., Bittner, D., Edwards, W.S., and Sperry, S. (1993) CAPI effects on interviewers: A report from two major surveys' Paper presented at the Census Bureau Annual Research Conference, Washington (mimeo)

Hansen, S.E., Beatty P., Couper, M.P. (2000)The Effect of CAI Screen Design on User Performance. *Proceedings of the Fifth International Conference on Logic and Methodology, Cologne, 2000*