A Practical Application of Audit Trails

Rob Bumpstead, Office for National Statistics

Introduction

As Computer Assisted Interviewing (CAI) has become the industry standard in the collection of official survey statistics, interest in mechanisms for monitoring the behaviour of those interacting with these systems has grown. Initial attention was focused on monitoring interviewers as they administered CAI questionnaires. More recently, this interest has extended to studying respondent behaviour in various modes of Computer Assisted Self-Interviewing (CASI).

This paper describes the initial stages of a small-scale project, carried out by the Social Survey Division of the Office for National Statistics (ONS), to evaluate the utility to the organisation of the monitoring method known as an audit trail, which has become available in Blaise in recent years. An ONS survey of children and adolescents in the care of local government authorities, known as the survey of Looked After Children (LAC) was chosen as the vehicle for this trial. The Blaise instrument[1] for the survey was designed to collect information about the mental health of children and included a substantial computer assisted self-completion section relating to sensitive or illegal behaviour. The main interest of the trial was to monitor respondent behaviour in the Audio Computer Assisted Self-Interview (A-CASI) section of the interview.

This paper may be of particular interest to those considering use of the audit trail tool, and its associated methodology, for the first time. The emphasis of this paper is upon the practicalities of using an audit trail, and an evaluation of the insights that may be gained by its use. While the work of the project is still on-going, and results incorporating final data are awaited, this paper seeks to explore some of the issues surrounding the capabilities, benefits and difficulties of using audit trails and audit trail data from Blaise.

Capabilities of the Blaise audit trail

The Blaise audit trail provides a record of the values of fields and the movements between fields. As the user enters a field, the value of that field and the time are recorded. When the user leaves the field another record of the value and time is made. In addition, the audit trail can also record use of certain functions, for example, help, save or the use of a searchtag. In this way a detailed, timed record of item response and navigation through the instrument is collected.

There are a number of potential uses that this information may be put to. Perhaps the principal use is methodological; to study navigation through an instrument, use of function keys, speed of data entry, and other aspects of respondent or interviewer interaction with a CAI instrument. What the Blaise audit trail cannot do is record *everything* the respondent or interviewer enters. So if, for example, a respondent begins a question response then deletes their initial answer and enters a different response and then leaves the field, only the final response will be recorded by the audit trail. The audit trail does not produce what is commonly referred to as a 'key-stroke file', that is, a file which details every key-stroke action in the CAI interaction. The limitations of the Blaise audit trail approach are considered later in this paper.

Other potential uses of the audit trail are more prosaic. An audit trail may be used to help debug problems in the Blaise code by tracking the flow of a very complex instrument. The audit trail also offers a means of

[1] The instrument was written in Blaise for Windows 4.2

recovery when data has been corrupted or lost: using the audit trail it may be possible to recreate such cases. Utilisation of audit trails for these purposes are outside the scope of this paper.

Using audit trails on CASI and A-CASI instruments

The 2001 survey of mental health of children in state care was the latest in a decade's work by Social Survey Division in the field of mental health survey research. The questionnaire for this survey is based upon that of previous survey of mental health among children living in private households, conducted in 1999. This 1999 survey used a combination of CAPI and CASI when interviewing young people. The interview included a number of problematic topics such as troublesome and violent behaviour, drug use and sexual activity. The ability of self-interview methods to increase reporting of sensitive, undesirable or criminal behaviour has been well established in the survey literature and therefore, a range of questions relating to such behaviour were included in the CASI section of the interview. (For example, Bradburn 1991; de Leeuw and Nicholls 1996; Tourangeau 1996)

The 2001 survey presented additional problems for the researcher. Much of the subject matter was similar to the previous survey and so a self-interview section was highly desirable. However, the children in state care were much more likely than their counterparts in private households to have problems with literacy and other learning difficulties. In addition, the prevalence of behavioural problems and problems with concentration among the sample population was also known to be very high. These factors could militate against the successful implementation of CASI, and so it was decided to use A-CASI. It was hoped that the audio option would enable a greater number of young people to complete the section, and that the added stimuli would help to engage the respondents. In addition, the heightened sense of privacy that A-CASI is sometimes theorised to provide, may further encourage disclosure of sensitive behaviour. Respondents certainly appear to prefer this mode of self-completion when compared with CASI (O'Reilly 1994).

Audit trail data could potentially play a useful role in assessing the success of these methods. The research team gathered as much information as possible about the impact of A-CASI; they questioned both the interviewers and respondents about their observations and experiences of this section of the interview. (Gatward 2001). However, an audit trail could provide much more information about the child's navigation through the instrument, their ability to use the keyboard and other special key functions available to them.

Perhaps most importantly the audit trail would be a means to ascertain whether, as well as providing answers, the children in the study were actually *listening* to the questions. The Blaise multi-media function provides the capability to complement or replace visually presented questions and answers with audio ones that may be heard through personal headphones. However, the respondent may interrupt the audio recording whenever they choose, enter an answer and move on to the next field. By comparing the audited time spent by the respondent in each field with the length of the sound file for each question, an assessment of the extent to which questions were, or were not, listened to could be made.

Implementing the audit trail

The process of implementing an audit trail is quite straightforward. First, the appropriate audit trail dynamic link library (DLL) must be selected or developed. The audit DLL is an external file which enables the audit trail to be written. The Blaise system provides two: audit DLL and auditkey DLL. The former writes all audit information into one file, the latter into separate files for each form according that form's primary key. The auditkey DLL thus provides distinct and identifiable files for each session in each form and was the DLL chosen for the trial. Although not all the information produced by this DLL may be needed it provided a starting point. Customising the DLL to the specific requirements of the survey could be considered at a later stage.

This DLL was installed on interviewer laptops along with the survey instrument. The audit trail was invoked by adjusting the audit trail section of the mode library editor. The 'make audit trail' box was ticked and the audit trail DLL name and path specified. The instrument was then compiled using the new mode library.

The audit output information from a DEP session is stored in an .adt text file, its content and appearance determined by the audit DLL. For each .adt file the standard Blaise DLL produces a chronological series of records based on events. Each record may contain up to five different fields. The first two fields are common to all records: a date timestamp field recording the date and time that a particular event occurred, and, an action field containing a record of any form or field event in the Blaise database due to menu selection or certain keystrokes. Most commonly, the actions recorded here will be 'leave field' or 'enter field' together with a specified field name, but a range of other actions can also be recorded. In addition, metafile information may be stored in this field, for example, the metafile name. Any other record details, including values, remarks and field status, are contained in up to three further fields in the .adt output. In the event of a 'leave field' action, details of the 'cause' will always be recorded: for example moving on to the next field by answering a question or navigating through the form by arrow or function keys.

In the trial, the initial installation and testing of the audit trail DLL on office computers and laptops was successful. However the audit trail did not appear to function properly when used on interviewers' laptops. The data transmitted back from the field included an .adt file for each interview, recording the timing of the beginning and the end of each session, but no information about interactions of the session itself.

Initially, attention centred on difficulties with memory as a possible cause of the problem and the interaction between the data entry programme, the questionnaire and the audit DLL. The problem might relate to the capabilities of the laptop computers used by interviewers and the large size of the questionnaire, which contained numerous sound files. This highlights one of the potential difficulties of using audit trails. A procedure which records every movement through a questionnaire from one field to another, and records values together with dates and times for each movement, can quickly generate an extremely large amount of information which can either affect performance or overwhelm the disk space available. However, reports from interviewers in the field seemed to indicate that laptop performance had not been noticeably affected.

Eventually the problem was diagnosed, not as a memory issue, but relating to the auditkey.dll source code and its interactions with the ONS case management system CaseBook. In summary, while the .adt file containing information about the session interaction was generated, it was written to an unexpected location on the interviewer laptop. This meant the .adt file was not transmitted back to the office along with the interview data.

While these problems were being overcome, the trial continued to the next stage of constructing processing systems and building an analytic model.

Processing the audit trail data

The detail and sheer volume of data produced by the audit trail presents the immediate problem of how to effectively conduct analysis. In the first instance, it is of course possible to examine each audit output by hand. Where the number of cases is relatively small, or where interest is limited to a few specific instances in the instrument, this may be a viable approach. In this example, the A-CASI pilot data was expected to consist of small number of cases (n=50). On the other hand, while the general area of interest was confined to audit output relating to the CASI sections of the questionnaire; this in itself consisted of over 60 questions and produced hundreds of audit records for each form. When targeting a particular question or question type, for example the handful of instances where an open text answer was required, the 'by hand' technique may be the most effective way of evaluating the audit data. However, in most instances the need to summarise the data in some form was plain.

Processing the audit trail file can be done via Manipula. Two standard Blaise programs are provided for this purpose one of which, auditsummary.man, was useful in this context. It summarises the contents of the data by keeping only the last value of any field and deleting all other lines. Further basic improvisations can readily be made to the Manipula programs to streamline the presentation of data and select particular areas of interest. Another tool for summarising data was kindly made available to us by colleagues at the Bureau of the Cenus in the United States, who had developed an audit trail report generator utility[2]. The purpose of this utility was to convert the contents of raw .adt output files to a simplified format, summarising the time a user spent within given sections of the questionnaire. A combination of all these methods was employed in preparing an analytic model for studying the LAC data.

Analysing the audit trail data

The first element of the model was concerned with timestamp data. The Blaise Audio-CASI method permits the respondent to interrupt an audio file as it is being played. The respondent may therefor listen only to some of the question, or if they listen to the full question, only some of the answer categories. This timestamp audit data can be used to reveal the extent to which interrupted listening has occurred. In developing a strategy for this analysis previous work in this field by Caspar and Couper was particularly useful (Caspar and Couper 1997).

A series of questions were selected for analysis from each of the following sequential sections of the questionnaire: moods and feelings, troublesome behaviour, alcohol and illegal drug use, sexual activity and exclusion from school. Initially, only those questions asked of all respondents were included to simplify the analysis. In the interview the question and answer options were played via headphones, while only the answer options were displayed on-screen. For each field the length of the sound file for the question and the answer categories was calculated. This could be compared with the actual time spent in each field by the respondent, and used to calculate two 'listening scores'.

The first listening score relates to the question only, and reflects the proportion of the question sound file the child had listened to. A score of one would indicate that the child had listened to the full question, a score below one, that the question sound file had been interrupted. The second listening score is a combination of both the question and answer categories, and can be expressed as a ratio. For example, if a child spent 20 seconds in a field where the combined time of question and answer categories was 10 seconds, the listening ratio would be 2. The differences between 'recorded time' and 'actual time' could then be used to produce an aggregate ratio for each section, and overall patterns of behaviour studied.

The implications of a low listening score for respondent understanding vary according to the wording of particular questions and answer options. A simple question may be adequately comprehended before the answer options have been heard or seen, or even before the full question text has been heard. On the other hand, uninterrupted listening may be essential for questions which end with important qualifying phrases. The respondent may need more time to consider questions where essential elements appear in the response categories.

The audit data can be used at the individual question level to explore these effects. At the aggregate level, listening scores can be assessed and compared across sections. We might expect scores to fall in general as the user progresses through the questionnaire and becomes more familiar with the format, and the audit data could be used to examine this theory.

A second element of the analysis was concerned with use of specific keys. In particular, a repeat function was provided to respondents so that the recording of a question could be replayed. Monitoring the use of this function would enable an assessment of how useful respondents found it to be, but could also highlight any

[2] We gratefully acknowledge the help of Ellen Soper from the US Bureau of the Census

problems in understanding particularly complex or difficult questions, or where the sound quality of a question had been insufficient.

Use of the keyboard outside the range of specified keys could also be examined. For most questions, respondents were instructed to use only the ‘enter’ and ‘repeat’ keys (labelled blue and white) together with the numeric keypad. Use of other keys could indicate difficulties in understanding the instructions, using the keyboard, or problems with concentration.

The final element of the analysis focussed on the children’s use of the limited number of open text boxes that the self-completion section of the instrument contained. Because the audit trail is not a keystroke file the extent to which behaviour in these boxes can be monitored is limited. Only the value of the field before and after entry is recorded. Indeed, it is not possible to capture ‘within-field’ deliberation of this type in any question response, though the number and length of ‘revisits’ to a question field could be studied.

However, the audit data does permit consideration of another aspect of respondent reaction. Open text questions only arise given certain responses to previous questions. Using the navigational data contained in the audit records it is possible to see whether respondents have ‘backed-up’ after reaching such a question and altered a previous answer to change their path through the questionnaire.

Conclusions and further work

In conclusion, the Blaise audit trail can provide useful data about respondent use of CAI instruments which it would not otherwise be possible to capture. In relation to A-CASI, the provision of ‘listening scores’ of the type described could be particularly valuable. The processing and analysis of audit data is not straightforward. However, it was possible to formulate a strategy for dealing with audit data which was not unduly cumbersome or time consuming, given the proviso of a relatively small survey dataset. The lack of a key-stroke capture facility in the Blaise audit trail did limit some of the options for analysis, but was not a serious impediment to the overall methodology proposed here.

The problems in fully operationalising the audit trail in the field were, in the short term, a more serious concern. Having been overcome, the next stage of the program could be to further develop and automate a processing system so that greater quantities of audit data might be efficiently analysed. This could include using other existing audit trail utilities such as WesAudit as well developing in-house systems. Another important step could be to customise the standard DLL used in this trial to tailor audit output. In these ways the utility of the audit tool might be further enhanced: while the application of the Blaise audit trail may in practice be best employed in testing relatively small quantities of data and experimental designs it is nevertheless a valuable tool.

References

Bradburn, N., Frankel, M., Hunt, E., Ingels, and Pergamit, M., (1991) “A Comparison of Computer-Assisted Personal Interviews with Personal Interviews in the National Longitudinal Survey of Labour Market Behaviour -—Youth Cohort", *Proceedings of the US Bureau of the Census Annual Research Conference*, Washington, DC: US Bureau of the Census, pp. 389 - 397

Caspar, R., and Couper M., (1997) “Using Keystroke Files to Assess Respondent Difficulties with an Audio-Casi Instrument”, *Survey Research Methods - Proceedings of the American Statistical Association*, pp239-244

de Leeuw, E., and Nicholls, W., (1996) “Technological Innovations in Data Collection”, *Sociological Research Online,* 1:4 <http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>

Gatward, R., (2001) "Audio-CASI with challenging respondents", paper in this volume, 7th International Blaise Users Conference

O'Reilly, J., Hubbard, M., Lessler, J., Biemer, P., and Turner C., (1994), "Audio and Video Computer Assisted Self Interviewing", *Journal of Official Statistics*, 10, 2, pp. 198-214

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