Towards Best Practice for Design of Electronic Data Capture Instruments

by

Rob Burnside

Phone: (02) 6252 7816

Fax: (02) 6252 8015

e-mail: rob.burnside@abs.gov.au

written for the Methodology Advisory Committee,

November 2000 Meeting

Executive Summary

1. Little definitive is known about standards, guidelines and the methodological impact of electronic data capture (EDC) instrument design and functionality choices. This is because of the relatively recent emergence of the Internet as a serious data capture route. Even less is known about the statistical and design issues specific to business collections.
2. This paper argues that EDC standards for business forms are best arrived at through starting from existing applied statistical, form design and, in particular, computer interface and related design standards and theory. These should be used to develop test instruments for identifying issues that have the potential to make a significant statistical impact. This will involve evaluating them through building, testing and deploying instruments in pilot tests in an empirical and iterative evaluation cycle.
3. The first two main sections cover the main areas of interest and concern for the design of electronic instruments for data capture from businesses. The first contrasts business collections with household collections, where important differences include those related to questionnaire design and the expertise of individuals involved in the data capture process. The second contrasts differences between paper and with electronic collection methods, where differences flowing from the medium itself may contribute to modal effects.
4. The third section outlines the development process we are following to investigate EDC issues, identifying some of the main issues and what we have learned to date. The fourth section briefly considers possible directions for future work, including evaluation processes and using instrument performance metrics to further assist design and performance assessment.
5. We would like advice and comments from the MAC on:
    - work in Australia on electronic data capture instrument design and modal bias and multi-modal collection issues;
    - thoughts on the design process and work so far; and
    - appropriate areas to focus on and methods of comparison and potential bias measurement.
6. The present research and testing work will lead to initial conclusions about EDC effects and directions for further work by mid-2001.

Introduction

7. The prospect of significant amounts of data being collected from business electronically with EDC instruments is increasing for a number of reasons. These include Commonwealth Government policy to encourage electronic interaction with governments, perceived cost, timeliness and data quality advantages, and provider expectations. At present putting suitable security in place is a major brake on live trials, but this barrier is expected to be at least partially removed shortly as providers become more familiar with secure on-line interaction (for instance through their dealings with the Australian Taxation Office

(ATO)), and ABS infrastructure is enhanced.

8. An important assumption in current work is that design and modal effects may have an impact on published statistics, so they need monitoring, understanding and controlling where necessary or possible.

9. The immediate aim of the current work is to develop methodologically defensible and practical interim standards and guidelines for electronic data capture instruments for national statistical collections from businesses. The intention is that these will be used in a similar way to, and in conjunction with, existing ABS paper form design standards and guidelines. An important contributor to this aim is understanding respondent behaviour in an electronic data capture environment.

10. One of the problems with electronic form design is that the area is relatively new. While there is a considerable body of knowledge on paper form design, user interface and systems design, particularly for the use of internal users in controlled environment, there is as yet little research or experience with electronic self completed questionnaires in an external environment.

11. Reflecting this, the statistical impact aspects of electronic data capture instrument design have received relatively little attention in Australia or overseas. Other statistical agencies are no further advanced than we are with respect to the quantitative methodological or design aspects of EDC instruments, though in most cases they have more experience with live pilot collections.

12. ABS is aware of developments in the Australian Taxation Office (ATO), The Australian Prudential and Regulatory Authority (APRA), the US Bureau of the Census (USBC), Statistics Canada, The UK Office of National Statistics (ONS), and Statistics New Zealand (SNZ), as well as other national statistical agencies. Of these only the USBC is believed to have undertaken any formal testing of self completed electronic instruments, though several others recognise that the methodological aspects of EDC should be considered.

13. Electronic data capture instruments or electronic forms are defined for the purpose of this paper as self completed screen-based questionnaires or interfaces that the respondent enters data into manually. Specifically excluded are structured or *ad hoc* data files, whether from business providers or as administrative by-product data, spreadsheets, and other information collected electronically, through telephone interviewing or in free text e-mail.

14. This paper uses the term 'EDC instrument' rather than the more common or colloquial 'e-form' or 'Web-form' because it is not necessarily the case that the instruments will be on the World Wide Web, and the former carries connotations of an on-screen reproduction of a paper form. While some early EDC developments used such 'e-forms' (which, along with fill and print' images, remain (basic) models), the design features, limitations and functionality of the instruments we envisage make them as much specialised computer interfaces as 'forms'. 'EDC instrument' is therefore used to emphasise these differences and to avoid confusion.

Background

15. Individual subject matter areas of the ABS have, over time, undertaken a range of EDC projects with varying degrees of success (see Attachment 1). Apart from a rapid advance in technological capabilities that is leading to expectations of the widespread use of EDC, current work differs in that it approaches the problem from a corporate or central (rather than local) direction with methodological considerations to the fore, rather than using a technology which is available and seeing what can be done with it.

16. ABS paper forms for business collections are subject to a formal approval process after being constructed and tested by subject matter areas in line with form development standards and guidelines that specifies most elements of the form design and recommend testing and evaluation procedures. They use standardised layout templates and include mandatory wording for national accounts data items,

17. The proposed EDC standards will cover:

    - design and functionality, including explanatory and help components, use of graphics, layout, and edit and error handling;
    - development and evaluation guidelines; and
    - the suitability of different approaches for different surveys and/or data items.

18. We anticipate that the formal approval process incorporate standards will be extended to EDC instruments to:

    - maximise the statistical and operational benefits of EDC through implementing methodologically defensible best practice widely;
    - present data providers with broadly consistent interfaces and behaviour to minimise their effort in using new instruments;
    - allow the development of coherent multi-modal collection strategies that may mix paper, electronic and telephone data collection - ie provide consistency when multiple collection modes are used inside a collection, either to reach different groups of providers initially or as part of response follow-up; and

- more generally, facilitate instrument development in the ABS' decentralised business survey instrument development model and avoid duplication of effort.

19. This last point follows partially from the history of ABS paper form standards, where at one time a number of different styles, approaches and systems were used in various collection areas, and organisational disruption and friction resulted from standardisation. While a decentralised model will still be followed for developing production EDC business collection instruments, we intend to avoid divergence by putting an agreed framework in place before EDC becomes widespread and in good time to consider and refine standards.

20. It is also worth noting that, while standards and guidelines are theoretically independent of the technology employed, in practice a major influence on best practice is the functionality, and defaults in the IT tools and skills available. However we anticipate that one result of the exercise will be identification of essential functionality or design capability that will have an influence on future tool selection.

I. Business versus household and internal instruments

21. A number of differences between business and household collections and instruments impact on EDC instrument design. Unfortunately much of the mode effect and electronic data capture survey research relates to household collections and forms, and a significant part of 'e-form' work to date has focused on the advantages of putting internal enterprise administrative and workflow tasks and forms on-line, for instance see chapter three in Gates (1999).

22. The first important difference between business and household collections is that the latter are usually tightly sequenced, and often delivered through Computer Assisted Personal Interviews (CAPI) or Computer Assisted Telephone Interviews (CATI). These ensure that the respondent is 'driven through' the instrument, usually in a single session by sequencing or routing based on answers to previous questions. Not only is there a tight logical dependence between responses, but the instruments contain more questions, to cater for different response sequences, of which only a subset are answered by most respondents.

23. This contrasts with business forms, which are generally self-completed instruments that are dealt with as part of the individual's main and often unrelated job. Business forms generally include a smaller number of questions than a household form, with a higher proportion of these questions being applicable to all respondents. Data are assembled over a period of time from different data sources and areas of the business. This occurs either by sending all or part of the form around the organisation or by an individual coordinating and assembling the required data.

24. One important design consequence for EDC is that it will be necessary for instruments to be capable of non-sequential or random access completion over time - ie they should be savable in a partially completed form, and have minimal or no routing, mandatory questions or edits to prevent or delay moving through the instrument to sections that can be completed at any given time.

25. The second difference arises because household CAPI and CATI instruments involve a guiding, trained interviewer. Similarly internal administrative forms are used in a controlled and supported environment with 'respondents' who are familiar with the subject and the system. These situations contrast with business forms, where most respondents are better thought of as inexperienced data providers, either because their organisation has not participated in the survey before or for very long, or because of staff changes and turnover in some or all of the business' internal data assembly and reporting chain - ie business form respondents are on their own.

26. At the same time the nature and sources of data are different for business surveys, being mostly based on quantitative, well defined accounting items or recognised industry-specific terminology, and the information is generally available from enterprise statutory and tax accounts or internal management information systems. Household data is often subjective, qualitative, and self-reported and so vulnerable to cognitive bias, recall error and social desirability effects.

27. An EDC design implication, supported by recent ABS visits to local government providers reporting data on spreadsheets, is that where there is less scope for provider confusion or subjectivity, instrument design is less critical, particularly with respect to the availability of help and definitions.

28. For EDC business collection instruments, the 'permanently inexperienced respondent' problem will be amplified by a 'permanent novice user' effect, where instrument designers cannot assume that there will be repeated use of the instrument or any significant learning effect over time. This points to the need for relatively simple and intuitive instruments, perhaps at the expense of theoretically desirable and technically desirable features such as complex editing.

29. Lack of control over the provider's IT environment (especially computing power, screen sizes, colours, resolutions, windowing and security arrangements), and the desirability of catering to as wide a range of potential respondents as possible also point to the need to develop relatively simple instruments.

II. Electronic Data Capture (EDC) versus Paper

Mode Effects

30. A major early concern with screen-based instruments relates to potential modal effects or bias, where there is a departure from a notional 'true' value due entirely to collection instrument and associated operational factors. While some bias is likely in any collection instrument, different bias in different instruments (ie EDC and paper) and its effect on data and subsequent statistics, particularly as the mix of instrument use changes, is an issue. This concern stems from:

    - experience in household collections in Australia and overseas where conversion from paper-based and/or interviewer collection to telephone interviewing or computer assisted interviewing has resulted in collection mode effects;

- situations where paper collection forms have been changed significantly and data effects have resulted; and
- from interface design experience and theory, which raises several human-computer interaction issues that may impact on statistical reporting using electronic instruments.

31. These modal effects arise from several factors or the interaction of factors such as the nature of the data sought and the reporting instrument. Well known examples are:

    (i) reporting of sensitive personal health or lifestyle information in face to face interviews tend towards the perceived socially desirable norms, whereas self-completed questionnaires more effectively capture less desirable or more extreme behaviour, especially if the respondent gives credibility to promises of confidentiality;

    (ii) list or order effects and response option visibility in questions where multiple answers are possible ('how many of the following do you own ...'). List effects manifest themselves because self completion paper form respondents (and respondents given showcards in face to face interviews) view a list of response options and satisfice by selecting a disproportionate number of activities in the top part of lists so they don't have to read all of them. Alternatively, when the list is just read out to respondents by the interviewer, such as in telephone surveys, they disproportionately select items from the end of the list that they remember better. A well-designed form or electronic system could minimise this by asking a 'yes/no' for each possible response;

    (iii) the visibility of grouped related questions tending to lead to greater consistency between responses; and

    (iv) on paper forms, the proximity and visibility of related questions gives context and additional information that assists accurate completion and retrieval of information, particularly when meaning or definitions are in doubt.

32. This point was noted in an EDC instrument test where navigation was screen-by-screen and in which the respondent specifically complained that he had to move between several screens to identify exactly what data was sought through comparison with previous responses, when it was easier to just look back on the paper form.

Paper versus screen

33. It is generally accepted (for instance Dillon 1992) that, for screen versus paper work:

    - reading speed is lower;
    - screens are particularly poor for reading large amounts of text;
    - comprehension of what is read is probably lower, possibly by up to 30%;
    - the onset of fatigue is faster and, possibly, user performance falls faster;
    - while eye movements are similar, skimming is more likely on screen;

- handling and learning the medium is more difficult - the application as well as the computer;
- manipulating information and data is more difficult; and
- navigation is more difficult.

34. Jakob Nielsen puts it a little more succinctly in a column (1997) on 'How Users Read on the Web' - 'They don't'.

35. This belief forms the basis for the new field of 'writing for the Web', which essentially argues for short phrases, dot points, lists and small chunks of text because users skim text on-screen. This is supported by the results of some of our own testing, where providers faced with a definitional or interface problem did not see or ignored what we thought were clearly visible instructions or navigation features, even after repeated prompting.

Starting Principles

36. While it is not appropriate or even possible to produce a paper form and its associated features on screen because of inevitable differences (such as using keys or pointing devices to navigate and enter data, screen sizes and resolutions), many of the principles for designing paper forms apply to EDC instruments. In particular, changes to instruments presented on screen that are driven purely by technical considerations or ease of coding should be avoided.

37. There are two main reasons for this, following from expectations about printed material and collection forms. These flow from human computer interaction (HCI) theory which emphasises (in part) that people rely on previously formulated cognitive models to interpret and interact with systems they encounter. ('Systems' being used very generally, and including anything from a book to a workplace or a society through to a computer game or a microwave oven.)

38. These models are, initially at least, based on knowledge of and experience with the conventions of similar systems or environments. The longer their exposure to what they perceive to be similar systems, the stronger their intuitive expectations are and the more likely a learnt 'natural' behaviour will be attempted, based on the model they bring to the new system. The efficiency of their interaction therefore depends initially to a large degree on the accuracy of their internal model and the implicit and explicit expectations it offers them as compared to the behaviour or design of the 'real' model embedded in the system.

39. Printed material, data collection forms in general and ABS statistical returns in particular are examples of a 'system' with which providers have some familiarity, so many general print design principles can be reasonably assumed to apply on screen.

40. The first 'model' respondents have is for printed material where common behaviours and expectations are:

    - reading or scanning from left to right across the information area then moving down and repeating for each line or block of information;

- larger or heavier typefaces having more 'importance' and being used to draw attention to a feature;
- certain colours and symbols have cultural associations (red/green; tick/cross);
- intelligibility - through using plain language and avoiding over-long lines of text.

41. Secondly, providers are familiar with the conventions used in ABS and other paper forms - these include such features as:
    - numbered questions which have some relationship to nearby questions;
    - sequential progress through a questionnaire;
    - help near questions;
    - question text to the left of or above answer boxes;
    - list order and category name considerations, and conventional visibility of available response options, including offering familiar 'other' and 'please specify' structures where appropriate.
42. Dillman (2000) agrees that existing paper form design principles are a useful starting point, and in particular notes that many paper instrument design principles apply equally to electronic instruments. These include using a sensible visual layout and grouping related questions, clear writing, unambiguous question structure and retaining the same ranges where applicable. He also offers the guiding principles that if something is a problem or works on paper, it is sensible to assume that, in the absence of firm evidence to the contrary, the same problems or advantages will apply on screen, so there should be a good reasons for explicitly departing from paper form conventions.
43. The EDC medium offers both advantages and disadvantages compared to paper. Likely positive effects from EDC instruments are more consistent data through in-line editing, fewer errors through access to fuller help and instructions, management of list effects through easy randomisation, improved response rates and better respondent cooperation through the availability of reporting vehicles which they find convenient. Contributors to detrimental modal effects include screen size limitations and the effect on reading, comprehension and the amount and style of on-screen information presentation, ease or otherwise of navigation, the possibly intrusive or irritating impact of interactive edits and error messages, and possibly a greater that usual reluctance to use instructions or help.

44. Some computer form devices are attractive, such as radio buttons, drop-down list boxes, allowing multiple selections in list boxes, spin boxes, using colour, other graphics, and colourful icons, and there will always be a temptation to 'use the new medium'. Of these only radio buttons with their automatic, mutually exclusive choices appear to be of outstanding clear benefit, and even then the ability of the respondent to change a previous selection automatically without realising it raises questions.

III. The Development process

45. There are three sources for our developing and evolving detailed standards:

(i) Existing Standards and Research

46. As well as contact with other Australian agencies with EDC projects, especially the ATO, a considerable amount of research has been done to identify other standards, guidelines and relevant theory. The theory is based on cognitive psychological and human computer interaction research and its application inside a user centred design approach.

47. Several comprehensive computer interface standards and guidelines sources are available, including Microsoft Corp (1999), Galitz (1997), and international ISO standards (1998) which draw on these sources too. There has been a potential conflict between these more established interface principles and rapidly developing but essentially *de facto* World Wide Web guidelines, as articulated in particular by a handful of on-line design and useability pundits and researchers as well as advice and guidelines on IT vendor Web-sites - Spool (User Interface Engineering - www.world.std.com/uieweb/index.html) and Nielsen (www.useit.com) are two of the best-known individual commentators, while many IT vendors, such as IBM, Sun Microsystems and Microsoft, all have large design standards sections.

(ii) Applied statisticians' work

48. The relative novelty of online, and particularly Internet, data capture means that there is little work available yet from applied and Official statisticians that specifically addresses instrument design and modal effects for electronic survey instruments. While this is starting to change, notably with Dillman's 'Mail and Internet Surveys' (2000), most of the emphasis so far has been on the response rate and self-selection nature of electronic surveys, with much of the work addressing simple e-mail questionnaires.

(iii) ABS testing

49. The third major source of ABS standards is the lessons learned from our initial field testing of prototype 'e-forms' for in the ABS Public Finance and Internet Service Providers surveys (twenty five respondent visits in all) and the development and deployment of two voluntary online Web forms to gather users' views on output for the next Population Census (approximately 220 live responses). Field testing of and a third survey - Business Expectations - is due to start early in November.

50. The visits to local government agencies and Internet service providers combined aspects of useability testing, observational studies and background user visits. Where possible respondents opened the instruments and worked through them to stimulate discussion and allow us to observe their interaction with them.

51. The impressions and design decisions coming out of this testing have been largely qualitative and based on a relatively small numbers of visits, though it is reasonable to draw conclusions form limited numbers of users - discussing Web-site testing, Nielsen (1999, from Nielsen and Landauer 1993) suggests that fifteen tests will be necessary to discover all useability problems, but go on to argue that because five tests can reveal 85% of errors, doing a small number of tests then testing changes is a better use of resources.

Census 'User Views' - Web-form

52. Both versions of the Census User Views form were targeted at individual users of Census data, both inside and outside organisations. They consisted of around nineteen questions with yes/no, text and multiple choice list components, not all of which were applicable to all respondents. The first version consisted of several screens, and had a particularly dense second page which appeared to result in a large number of respondents who started the form abandoning it. The second version was in the from of one scrolling page. It collected similar information, and much more attention was paid to layout and spacing in the second version; it also contained more complex edits and sequencing.

53. Both forms allowed respondents to add comments about the e-form itself. These were generally favourable to the design and the electronic reporting medium - though that could be expected from respondents who chose to use the 'e-form' in the first place. Points made included allowing text boxes to wrap, being able to print responses in full, being able to add notes or qualifications to yes/no answers, and comments about the sequencing.

54. The Web-forms were not savable - ie they had to be completed in a single session. We included an explicit 'time taken' question, and while this arguably did not make clear whether background reading time was included or not (as some respondents noted), the server also captured the time the form was open, so enabling a rather interesting comparison between reported and on-line times.

55. There were 48 records where respondents provided feedback. Eleven were removed because their reported time was zero and on-line time was very small - these are assumed to just have been looking through the form.

56. Of the remaining 37, mean reported time spent on the form was just over 19 minutes, whereas online time was 27 minutes suggesting that perceived 'provider load' for the instrument was significantly less than actual load, and implying that EDC instruments may lessen perceived reporting burden for on-screen instruments.

Local Government - Spreadsheets

57. The seventeen local government authorities in Victoria and New South Wales we visited were already using an EDC instrument that had been constructed in a spreadsheet (Excel). These used multiple worksheets for different groups of data items, the NSW instrument having more data items (350 against 70)

because it is an annual collection that obtains a fuller dissection of the items sought - however many items are not applicable to most providers, particularly in the larger annual collection.

58. The results of the visits were remarkably consistent:

    - the form is usually completed by a single person with a good knowledge of the data - typically a qualified public sector accountant;
    - respondents had no significant definition or help needs due to the ABS data items being closely aligned with Grants Commission, Local Government and applicable accounting standards and terminology;
    - there were no significant graphics/navigation/layout issues as all respondents were very familiar with Excel;
    - the main source of data were internal management reports/systems, and these contained all the data necessary to complete the return; and
    - data was usually transcribed from other reports to the ABS spreadsheet, except for some the councils in NSW who use a piece of report generating software, in which case their system can produce the ABS requirements directly from their source data.

59. While not directly related to immediate EDC design issues, but of interest in the context of streamlining ABS reporting where possible, respondents were generally opposed to the collection of other (ie non-financial) ABS information on the same form.

Internet Activity Survey (IAS) - dedicated electronic form tool (FormFlow99)

60. The IAS instrument has 20 questions, several of which ask for breakdowns by diffenrt tyoeps of activity. It was by far the most complex instrument tested, and included a range of in-line (immediate) and on-completion edits, extensive help and definitions, and elaborate combined navigation and completion status features. One result was that system compatibility and installation problems were a major issue because the instrument only worked in Microsoft Windows 95/98 and Internet Explorer 4/5 browsers. This prevented us doing several visits because providers who did not support this configuration were not interested even though we offered other alternatives such as using an ABS laptop or bringing them into an ABS office.

61. Testing was in three groups (of one, four, then three visits) with significant changes being made after each of the first two groups. Given the nature of the subject business, there was a split between technically orientated programmers and managers or entrepreneurs, with the programmers (perhaps not surprisingly) offering a lot of useful advice on additional functionality they would like to see.

62. Heavily requested features were:

    - help - a keyword search facility, 'tooltips' for all buttons and more use of hyperlinks;
    - a 'come back later' button and flag/icon for skipped questions;
    - a default 'ok' for accepting address and contact details; and

- a fully printable form and/or printable summary of questions.

63. Navigation remained an issue through the tests, with quite divergent views on the best approach - buttons versus our eventual 'explorer-style' tree structure, which tended to be either misunderstood completely or accepted and used as intended and without apparent thought.

64. Respondent initiated action sequences for editing, verification, saving and data submission and their associated buttons, dialogue boxes and descriptive wording proved particularly difficult to get right, both in terms of the result the users expected and the wording of individual text messages.

Testing Summary

65. Testing to date has confirmed that basic and accepted screen design and interface principles also hold for EDC instruments. For visible design areas these include such aspects as text and background colours, fonts and font sizes, user dialogue boxes and wording, grouping and spacing.

66. At the level of instrument behaviour they include discoverability (allowing the respondent to go through the instrument experimentally to explore it's functionality without making any irrevocable decisions), predictability (where the instrument behaves as expected when actions are performed - eg entering data, saving, making errors, and moving to different screens), and forgiveness (where actions are reversible and do not delete data or close down future interaction or reporting options, for instance by permanently closing off sections of an instrument based on what may be an incorrect response).

67. Significant themes and definite support for issues identified by prior research are that:

    - users want to print and save the instrument both before and after completion;
    - the presentation and implementation of help and navigation are real problems;
    - wording of pop-up dialogue windows respondents interact with for edits, validation, saving etc require care and testing;
    - users often have clear expectations about what the instruments should look like and how they should behave, but these can vary between users;
    - respondents appear to be even more reluctant to read questions, notes and other on-screen data-specific instructions than they are for paper instruments, but want access to help or definitions to be convenient and immediate; and
    - systems issues, including software, security and screen settings are limiting and complicating factors.

68. It appears increasingly likely that different instrument standards and functionality will be better suited to certain data and respondent types. These could range from Web-based single session or non-resumable (non-savable) forms, through spreadsheets to purpose built (programmed) instruments or those constructed

using dedicated electronic form design tools.

69. For instance, for quantitative economic or accounts sourced data, data providers need the ability to complete their returns over time and non-sequentially as they assemble the data sought. As a consequence they need to be able to save partially completed returns without submitting them to the ABS.

70. Three important procedural results have emerged from the visits. The first is that it has sometimes proved quite difficult to get respondents to focus exclusively on the instrument, rather than the survey content, in a controlled way. The second is that testers are very easily distracted by trivial errors, bugs and wording they dislike, and expect all instrument functionality to be present. This is in seeming contradiction to much conventional useability testing and instrument development wisdom, which suggests using outlines, prototypes and partially developed forms or applications to get early 'look and feel' style feedback.

71. The third procedural issues is that we are having considerable difficulty recruiting sufficient testers, even amongst Internet service providers who were initially seen as a very sympathetic and supportive respondent group.

Further work

Live Tests

72. Two series of field tests for the Business Expectations Survey (BES) will be undertaken before the Christmas break, after which live pilots of the Internet Activity Survey and BES are planned for the first quarter of 2001 - as soon as technical security and confidentiality problems are overcome. These will be followed up by post enumeration studies to visit comparable electronic and paper data respondents, with further design iterations expected to result.

73. The process to date has been aimed at producing viable instruments that we can use for controlled live pilot testing, ideally with around thirty respondents for each instrument type. We intend that the live trials in the two surveys mentioned above will approach enough respondents to start to undertake some limited quantitative evaluation, probably through simple ANOVA and t-tests on reported data, response rates and timeliness of responses.

74. Consideration will then be given to broadening research by developing and pilot testing instruments in an ABS 'Major Economic Indicator' survey with a different respondent- and data-type profile - ie mainstream private sector respondents providing mainly or entirely quantitative economic data that feeds into the national accounts.

Editing

75. The question of an appropriate EDC editing philosophy or strategy arises because one of the most forcefully propounded and widely accepted assumptions about EDC is that data quality will improve through pushing editing and validation down to the respondent so that they can resolve problems before data is returned.

76. The important aspect is the degree of detail to which editing and validation is taken in the instrument, including logical dependencies, tolerances and completion of mandatory fields, and whether any, or which types of, edits should be 'fatal' in the sense that failure to resolve them prevents submission or further progress through an instrument.

77. A second aspect is where in the data capture sequence various edits should be executed - for instance in a data entry field, on exiting a question, a page, a section, or at the end of the form. While some edits of the 'alpha required, numeric entered' type clearly belong at the point of data entry, the balance between making an instrument so laborious to complete that take-up is uneconomically low and getting worthwhile benefits through a lighter validation touch is not apparent.

78. In both cases we anticipate that extended live pilots tests and comparisons with comparable paper sourced data will be necessary to gain an understanding of the appropriate balance.

Instrument Evaluation and metrics

79. In the longer term, EDC instruments, particularly Web-based ones, offer potential for incorporating instrument metrics that will give us 'real' performance and evaluation measures that could prove to be of tremendous value (Nichols and Sedivi 1998).

80. Variations of the established keystroke or 'click-stream' tracking can be used to capture measures such as time taken, feature or instrument usage and navigation patterns, the use of help and definitions, incidence of edit and validation failures and tracking of queries. Such 'real' instrument performance indicators could be used for quantitative continuous testing, comparisons, experimentation and fine tuning in the routine live environment rather than an artificial test one, though privacy and monitoring concerns will necessitate careful presentation of such tools.

Second Order EDC Effects

81. While the present focus is on developing instrument standards and building up expertise, other EDC effects on non-response estimation and provider management are likely. These will flow from the gradual introduction of EDC if there are significant modal effects related to self-selection, switching of reporting mode, and changed and changing response rates and timing.

Conclusion

82. Overall, economic electronic data instruments are harder to design, build and test than their paper equivalents, and there are fewer established guidelines.

83. Controlled live tests of reasonable size are necessary to compare and quantify design issues and effects.

Points for MAC discussion

Are there any major issues we have not canvassed?

Does MAC have any additional comments on the availability and applicability of existing research and standards?

What does MAC think about the idea of EDC instrument 'horses for courses' - ie different functionality and design for different data types and provider groups?

What are MAC's thoughts on:
- comparisons between EDC and paper instruments when it comes to attempting to quantify differences in reporting, particularly relating to possible bias and data quality, and avoiding confounding in comparisons and trials?

- multi-modal approaches and data consistency, where different instruments are used to collect conceptually identical data from the same population.

References

Dillon A
Reading from Paper versus screens: a critical review of the empirical literature
Ergonomics 1192 Vol 35 No 10 p1297-1326

Dillman D
Mail and Internet Surveys: The Tailored Design Method
Wiley Second Edition (2000)

Galitz WO
The Essential Guide to User Interface design - An Introduction to GUI Design Principles and Techniques.
Wiley New York 1996

Gates W III
Business @ The Speed of Thought
Penguin London 1999

International Organisation for Standardisation
Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 17: Form Filling Dialogues

Microsoft Corporation
Microsoft Windows User Experience
Microsoft Press Redmond Washington 1999

Nichols and Sedivi (US Bureau of the Census)
Economic Data Collection via the Web
Proceedings of the ASA meeting in Dallas Tx 1998

Nielsen J
How Users Read on the Web
www.useit.com
Alertbox for October 1, 1997:

Nielsen J
Why you only need to test with 5 Users
www.useit.com A
Alertbox for March 19 2000:

Nielsen and Landauer
A mathematical model of the finding of useability problems
Proceedings of the ACM INTERCHI'93 Conference Amsterdam 24-29 April 1993 pp 206-213