Pre-Testing in Survey Development: An Australian Bureau of Statistics Perspective

Disclaimer
Views expressed in this paper are those of the author(s) and do not necessarily represent those of the Statistical Clearing House. Where quoted or used, they should be attributed clearly to the author.

Introduction

1 The Australian Bureau of Statistics (ABS) generally incorporates pre-testing into the development of all its household surveys. In 2000, the ABS entered into an agreement with the Commonwealth Department of Health and Aged Care to undertake pre-testing of nine proposed modules being developed as part of the Preventable Chronic Disease and Behavioural Risk Factor Health Survey Module Manuals. The aim of this program is to aid in the development of standard modules for endorsement by the Technical Reference Group. The testing was to take place over a three year period, with the ABS providing advice about possible data quality issues and suggestions for ways to minimise identified sources of non-sample error.

2 To date, the ABS have conducted pre-testing on the first three modules, namely Demographics, Asthma, and Diabetes. This process involved an expert analysis and three rounds of cognitive interviews and produced a number of recommendations for minimising sources of non-sample error. However, there were other data quality issues identified by both the ABS and the TRG which were not suited to exploration through these methods of pre-testing.

3 Cognitive interviews are commonly used by the ABS in pre-testing surveys, although the widespread use of this technique has only taken place over the last three or so years. Given the relative rapidity with which cognitive interviewing has been accepted and adopted by statistical agencies around the world, this paper will attempt to provide an understanding of the role of cognitive interviewing in comparison to other pre-testing techniques and how it fits within the suite of pre-testing tools.

4 As the term 'pre-testing' has different meanings for different people, this paper will present an ABS perspective on pre-testing. In presenting pre-testing from an ABS perspective, it is important to note that in conducting pre-testing, the ABS operates within the boundaries of the ABS legislation and policy. The range of techniques that the ABS uses in 'pre-testing' are briefly explained, whilst some other techniques not used by the ABS are defined, along with the reasons why the ABS chooses not to use these techniques. Rather than discuss pre-testing techniques in detail, this paper aims to provide readers with a feeling for the range of techniques available to survey developers. The second part of this paper discusses the relative pros and cons of each technique, framed in terms of the criteria used by the ABS to select an optimal combination of techniques for any particular testing program.
Pre-testing Defined

5 Pre-testing refers to a range of testing techniques which are used prior to field testing techniques, such as Pilot Tests and Dress Rehearsals. Pre-testing, or pre-field testing, techniques aim to identify non-sample errors and to suggest ways to improve or minimise the occurrence of these errors. Types of non-sample errors include:-

- respondent biases which arise from interpretation of the questions and the cognitive processes undertaken in answering the questions,
- interviewer effects, arising from the interviewer's ability to consistently deliver the questions as worded,
- mode effects, caused by the design and method of delivery of the survey instrument, and
- the interaction effects between these.

Thus, whilst questionnaire pre-testing provides means to reduce errors by improving survey questions, it cannot eliminate all errors in survey data.

6 There are a range of qualitative pre-testing techniques available for survey designers to use to meet different purposes. These techniques aim to identify errors that may be introduced during the administration of the survey. Many of these techniques are based on theories from cognitive psychology which provides a framework for understanding respondents' thought processes and influences on these processes.

7 Being qualitative tools, the techniques described below which involve interviews or discussion groups all use convenience, or purposive sampling, rather than strict probability sampling. Thus, whilst pre-testing identifies issues which exist within the broader population which may affect data quality, techniques which use probability sampling are required to provide information about the magnitude of the effects these issues will have on the final data.

Pre-testing techniques used by the ABS

8 Techniques used by the ABS include:-

a. Literature review
   Particularly database and library searches plus information from other national statistical agencies and survey organisations.

b. Expert review
   A group of survey design 'experts' review a questionnaire to identify potential sources of non-sample error by understanding the respondent's task and to provide suggestions for ways to minimise potential error. Experts are individuals who are considered to be experts in the critical appraisal of survey questionnaires (Willis, Schechter & Whitaker, 1999). In practice, these are people who can apply their theoretical understanding of, and extensive experiences in, survey development in critiquing questionnaires. This technique can also incorporate subject matter 'experts' and interviewers as well.
In conducting an expert review at the ABS, experts systematically analyse the response task for each question in terms of comprehension, information retrieval, judgement and response generation.

c. Focus groups
An informal discussion on an issue or topic, led by a moderator or facilitator, with a small group of people from the survey population. Focus groups are used early in development to explore conceptual issues relevant to specific sub-populations. They can be used to:-
- determine the feasibility of conducting the survey,
- develop survey objectives or data requirements,
- determine data availability and record keeping practices,
- explore and define concepts,
- clarify reference periods,
- evaluate respondent understanding of language and terminology, and
- evaluate alternative question wording or formats and to understand respondent burden.

Through focus groups, survey developers can identify specific terminology, definitions and concepts used by respondents and can identify potential problems with data availability and intended collection methodologies. They assist survey developers to better understand the range of attitudes or understanding and the complexity of the task for respondents.

Focus groups are particularly useful because they allow a small or 'rare' segment of the population to be tested that is likely to be underrepresented in a larger field test. They are generally unsuitable for non-English speaking populations, as a translator can disrupt the flow of the conversation within the group. They are also generally unsuitable for highly sensitive or emotive topics as biases in intra-group behaviour are likely to distort information and there is a tendency for participants to give 'public' opinions.

d. Interviewer debriefing
Interviewer debriefings combine standardised interviewer debriefing questionnaires and focus group style interviews to gather information from interviewers about either a previously used survey instrument or a draft instrument. They can also be used after field tests and/or after data collection to provide information for later stages of survey development and future iterations of the survey. Whilst the ABS routinely conducts interviewer debriefings after each field test, they are less commonly used in the pre-testing stage of the development process because interviewer input is often sought in expert reviews.

e. Observational interviews
Observational interviews are commonly used at the ABS to test and evaluate self-completion forms. In an observational interview, a trained observer watches the survey process under study (eg: form completion or responses within an interview) to better understand the respondent's thought processes during survey administration.
Observational interviews aim to identify problems in wording, problems in question order, presentation or layout, and to estimate the time taken to complete the questionnaire or parts of the questionnaire. Survey designers look for behaviours that result in an error on the instrument, including the participant's behaviour (eg: reading all the questions before responding), non-verbal cues, reactions and observed cognitive processes (eg: counting on their fingers or writing calculations on the page). This technique can also use follow up probes to elicit information about why the respondent behaved as he or she did.

f. *Cognitive interviews*

A cognitive interview is an in depth one-on-one interview in which trained cognitive interviewers ask volunteer participants probing questions about the survey questions being tested. The ABS considers cognitive testing to be an iterative process, in which interviewers conduct a number of rounds of interviews, allowing for changes in the aims of testing, the questions tested and the scripted probes between each round. The ABS usually conducts between twelve and fifteen interviews per round, to ensure sufficient data is gathered in each round.

Cognitive interviews are directed at understanding the cognitive processes the respondent engages in when answering a question. Using a multi-stage model of information processing, cognitive interviewing allows survey developers to identify and classify difficulties respondents may have according to whether the source of non-sample error occurs in question comprehension, recall of information, answer formation or providing a response. As well, cognitive interviews can provide information on adverse respondent reactions to sensitive or difficult questions. Specifically, this technique is used to assess how answers are formulated by respondents, how respondents understand questions and concepts, the range of likely answers to a question and the level of knowledge needed to answer a question accurately. Thus, this technique allows both the source of, and reason for, an error in the questionnaire to be identified.

Cognitive interviewing is based on the assumption that verbal reports are a direct representation of specific cognitive processes (Ericsson & Simon, 1993). To elicit useful verbal reports, interviewers prepare scripted protocols, which contain probing questions, explanations of the respondent's task and debriefing information. Interviewers also need to be skilled in forming and asking spontaneous probing questions based on information gained through the conversation and through aural and non-verbal cues.

Cognitive interviews can be conducted concurrently or retrospectively. Concurrent probing involves asking the respondent to describe aloud his or her thought processes as he or she answers, or probing directly after each question. In an interview using retrospective probing, the interviewer administers the survey in totality and then asks specific probes about a particular question.

The ABS tends to use concurrent probing to understand detailed response processes, particularly understanding and recall issues. Probing during the interview produces context effects for subsequent questions. For example, respondents tend to think more deeply about concepts or expend more effort to recall information in subsequent
survey questions, after being required to answer probing questions. Retrospective probing is typically used by the ABS to elicit information about the questionnaire as a whole, and to identify possible context and mode effects, as this technique allows the interview to flow without being disrupted because of probing by the interviewer. The general probes that can be used in concurrent probing techniques offer more convincing evidence of errors because there is less chance of the interviewer having led the respondent. Thus, the ABS always conduct at least some cognitive interviews using concurrent probes for each topic being tested via cognitive interviews.

Cognitive interviewing can also incorporate a number of other techniques to increase the range of information that can be obtained from an interview. Three techniques commonly used by the ABS are paraphrasing, vignettes and card sort tasks.

Paraphrasing involves asking the respondent to repeat the question in his or her own words. This allows the researcher to understand how the respondent interpreted the question and whether this interpretation is consistent with the researcher's expectations. Paraphrasing can also suggest alternative and more consistently understood question wording.

Vignettes involve having the participant respond to a question, or series of questions, from the point of view of a hypothetical situation. This allows interviewers to explore participants' response processes in situations in which the participant may not have direct experience. This technique is especially useful in gathering additional information about understanding of concepts, and calculation or construction of responses.

Card sorting tasks provide interviewers with information about how respondents think about categories, group information or define particular concepts. This information is obtained by asking respondents to sort through a list of words, or concepts, according to whether they are representative or not representative of a particular concept. In particular, card sorting tasks provide interviewers with a better understanding of what respondents included or excluded when answering a survey question of the format "How many times...?"

Cognitive interviews are usually limited to about 1-1.5 hours per interview, due to both interviewer and participant fatigue. Thus, the number of questions about which detailed information can be collected is limited in each interview.

g. *Behaviour coding*
Trained coders systematically assess respondent / or and interviewer behaviour during an interview according to a predetermined list of behaviours, to identify errors. Codes can also be developed to record features of the interaction between the interviewer and the respondent. Behaviour coding can be conducted as part of field tests, as well as in the laboratory in addition to other forms of interviewing, as part of pre-testing. Behaviour coding can involve both qualitative and quantitative analyses.

Behaviour coding is based on a model whereby any deviation from the questionnaire by the interviewer, or any less than complete answer by the respondent, indicates a problem with the questionnaire (Cannell, Lawson & Hausser, 1975). It is used to
identify common problems with the administration and completion of the questionnaire. Behaviour coding indicates to the researcher that a problem may exist with the questionnaire but this technique cannot provide any information about the nature of the problem and thus possible solutions.

**Pre-testing techniques not used by the ABS**

9 The techniques used by the ABS are not definitive of pre-testing. In addition to the techniques described above, other tools that are not used by the ABS include:-

a. *Computer based tools, for example, QUAID / QUEST*
   QUAID is a software tool that identifies some potential problems respondents might have in understanding a question. It was based on a cognitive computational model. It was designed to be used collaboratively with survey designers, so the program points out potential errors and designers screen the list of errors and decide on 'fixes'. QUAID successfully critiques survey questions based on problems such as vague or imprecise relative terms, unfamiliar technical terms, vague or ambiguous noun phrases, complex syntax, or working memory overload (Graesser, Wiemer-Hastings, Kreuz, Wiemer-Hastings, 2000).

However, the ABS does not use this technique for a number of reasons. The main reason is that QUAID does not perform a complete analysis against all problem criteria. For example, it can't tell survey developers where working memory overload occurs. In addition, the program is still being developed, with work focusing on broadening the range of problems that can be identified and in improving the sensitivity of the analyses.

b. *Computational linguistics / Literature and lexicon searches*
   These techniques use computer programs to search large bodies of text to identify the generalities of language as used by different speakers and writers. They identify the co-location of a word with other words, the context in which a particular word is used and the grammatical frames in which the word occurs. These searches can suggest sources of potential comprehension problems within question wording to survey developers (Graesser, Kennedy, Wiemer-Hastings & Ottati, 1999).

The ABS does not use these tools because the available programs have been written from an American perspective in terms of language use and understanding and because the value-added by such a tool to the development process is low, given that it can only identify one type of error, which is reasonably well covered by other techniques. Computational linguistics have also been incorporated into QUAID/QUEST.

c. *Response latency*
   This technique is based on the assumption that very short or very long item response times reflect a problem with the question. Given the variation in question task complexity that is common in household surveys, establishing a baseline response time is difficult. Measures of response latency are a by-product of the cognitive
processes that occur during question answering and as such do not reveal any direct information about cognitive processes. Further, response latency studies are less useful in survey development than other techniques as they do not identify the type of error or provide any guidelines about to minimise the error.

**Other techniques**

10 The ABS also uses some techniques that are conceptually half way between pre-tests and field-tests. That is, they are either small scale field tests or qualitative components of field tests:

a. **Skirmishes**
Skirmishes test two or three narrowly defined aspects of a survey, such as the effectiveness of introductory letters or a specific field procedure. They are small field studies which typically use about 150-200 completed questionnaires.

b. **Respondent debriefings**
These are conducted after a skirmish or field test and involve a focus group style, structured discussion. They can provide information about reasons for respondent misunderstandings, as well as information about particular aspects of the survey, such as respondent's use of records to answer survey questions.

c. **Follow-up questions**
This technique, sometimes called post enumeration studies by other statistical agencies, involves asking additional questions to respondents at the time the survey is administered. The additional questions can be asked concurrently or retrospectively. The aim of these extra questions is to provide additional information for validation of data items or to probe for a range or explanation of response alternatives.

Follow up questions focus on respondents' thought processes as they completed the survey. For example, a follow up question to the question "Would you prefer to work more hours each week?" might be 'what are the reasons you would not prefer to work more hours each week?'. This information might be useful in suggesting what facets of the issue respondents are considering when responding to the question. This can yield information on context effects and whether satisficing is occurring.

**Criteria used by the ABS to select a pre-testing strategy**

11 Given the wide range of techniques available to survey developers, the difficulty lies in selecting the right combination of techniques to achieve the objectives of testing, within the available resources.

12 A number of researchers have attempted to compare the relative usefulness of different pre-testing techniques in detecting and minimising survey errors. For example, Presser & Blair (1994) found that expert reviews and behaviour coding were more reliable than cognitive interviews and interviewer debriefings and that expert panels identified more errors than other methods. However, researchers have tended to conclude that even if
suitable measures can be found by which to compare the quality outcomes of pre-testing techniques, the pre-testing techniques are not directly substitutable (Esposito, Campanelli, Rothgeb & Polivka, 1991; Willis, DeMaio, Harris-Kojetin, 1999; Willis, Schechter & Whitaker, 1999). Rather, each technique can be best used in different circumstances, with the particular strengths of each technique complementing each other at different points in the questionnaire development process. The task for survey developers is to select the combination of techniques that optimises the use of available resources, in meeting the aims of the research.

13 There are a number of factors the ABS takes into consideration when planning a pre-testing strategy for a given survey development project. These include:-

1) Resources
   • Cost
   • Labour intensity

2) Timeliness of results

3) Stage of development process

4) Aims of test
   • Range of non-sample errors identified
   • Detail of non-sample errors identified

**Resources**

14 The resources required to undertake pre-testing can be broken into:-

1. the monetary cost of testing and
2. the number and duration (intensity) of survey development staff required to actually conduct the test.

15 Cost and labour intensity are interrelated in that the ideal number of staff required to undertake a testing program will affect the cost of the program through the accumulation of salary, overhead and travel expenses. Thus, the relative costs associated with each technique may be influenced by trading off the number of staff assigned to undertake the test against the length of time the staff are required for.

16 It should also be noted that the comparison of pre-testing techniques below assumes that trained staff are available and excludes the costs of training survey development staff in these pre-testing techniques.
The cost associated with any pre-testing technique includes the human resource costs of survey development staff involved in the testing (in terms of full-time equivalent) plus costs associated with recruitment and payments to interviewers or participants. Indirectly, cost is thus dependent on the number of iterations of testing, the number of participants, the recruitment strategy etc. Generally however, some techniques require more resources than others (See table 1). For example:

a. Literature review - Low cost.
   This technique usually only requires one staff member and any associated costs are usually negligible.

b. Expert review - Low cost.
   Often the experts are not actually the staff working on the survey, so there are usually few direct costs to the project. However, there can also be a reasonable indirect cost to experts' employer in the form of the opportunity cost of experts' time, the cost of the experts' salary and any other overheads of the experts.

c. Focus groups - Medium cost.
   In addition to the human resource costs of the survey developers, focus groups can involve some travel by facilitators, payment to participants, and the cost of recruitment advertisements.

d. Interviewer debriefing - Medium cost.
   In addition to the human resource costs for the survey development staff, interviewer debriefings may involve travel by the development staff and payments to interviewers for their time.

e. Observational interviews - Comparatively high cost.
   As well as the human resource costs, observational interviews require payments to participants and sometimes interviewers and can incur costs of recruitment advertisements. Observational interviews may require more staff than cognitive interviewing if both an interviewer and observer is needed.

f. Cognitive interviews - Comparatively high cost.
   Cognitive interviews are resource intensive because they involve one-to-one interviews. Thus, they are time consuming for development staff and incur the associated salary and overhead costs. The fixed costs are higher too as setting up a cognitive laboratory requires appropriate audio-visual equipment and enough trained interviewers to ensure that interviewers do not interview more than about 2 participants a day. In addition, cognitive interviews incur costs such as payments to participants, the cost of recruitment advertisements and travel costs if using mobile laboratory equipment. Costs also vary depending on the number of rounds and number of different populations required.

g. Behaviour coding - High cost.
   Costs of behaviour coding can be high if quantitative data analysis is planned as a reasonable number of interviews will be required to yield sufficient data quality,
allowing for all skip patterns within the questionnaire. Being labour intensive, they tend to have high costs associated with salary and overheads for development staff, as well as the cost of training coders. Behaviour coding may also incur costs of recruitment, payment to respondents and development staff travel, depending on the sample selected.

**Labour intensity**

18 Pre-testing techniques differ in the number of staff required to organise, conduct and analyse data, as well as the length of time required by those staff to complete the testing. The number of staff and the time they are required for are interdependent. Both variables are also dependent on the time frame for the test, the complexity of the pre-testing objectives, the amount of information being tested, the level of detail required, the number of interviews being conducted etc. However, as a general rule, some techniques require greater labour intensity to conduct than others (see table 1).

a. Literature review - Low labour intensity.
   A literature review usually requires one staff member to complete and the ABS allows one to two weeks for this.

b. Expert review - Medium labour intensity.
   The amount of labour required depends on the complexity of the questionnaire but generally this technique requires two days to one week of work per expert, with between three and five 'experts' contributing to a review. In addition, another three to five days are required by one person to collate the results.

c. Focus groups - Medium labour intensity.
   A focus group requires about two to three development staff to organise, moderate and observe the focus groups and to analyse the data. The ABS allows about three weeks per round of focus groups, although more time may be required if staff need to travel between groups.

d. Interviewer debriefing - Low labour intensity.
   Interviewer debriefings usually require one to two staff members to organise, conduct and analyse the data. Whilst interviewers would need to be notified some weeks in advance, majority of the work involved in organising, conducting and analysing an interviewer debriefing usually takes between one and two weeks at the ABS.

e. Observational interviews - Medium labour intensity.
   In general, two to three staff are required for each round of observations conducted. The ABS allows about three to four weeks per round of observations.

f. Cognitive interviews - High labour intensity
   The ABS usually uses three to four staff over five to six weeks per round of cognitive interviews, to organise and conduct the interviews and analyse the data.

g. Behaviour coding - High labour intensity.
Behaviour coding usually requires two to four development staff to organise, conduct the interviews, code and analyse the data. The ABS allows about six to eight weeks to organise, conduct and analyse the data.

Table 1: Comparison of pre-testing techniques by resources required.

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost</td>
</tr>
<tr>
<td>Literature Review</td>
<td>*</td>
</tr>
<tr>
<td>Expert Review</td>
<td>*</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>**</td>
</tr>
<tr>
<td>Interviewer Debriefing</td>
<td>**</td>
</tr>
<tr>
<td>Observational Interviews</td>
<td>***</td>
</tr>
<tr>
<td>Cognitive Interviews</td>
<td>***</td>
</tr>
<tr>
<td>Behaviour Coding</td>
<td>***</td>
</tr>
</tbody>
</table>

key: * low ** Medium *** high

Timeliness of results

19 Timeliness of results refers to the length of time required to resolve issues identified through testing. In practice, this means the faster solutions can be incorporated into the testing process, the faster issues can be resolved. As a general rule, qualitative techniques allow for quicker resolution of issues than quantitative techniques, because the process of identifying problems and finding solutions can occur during data collection. In addition, the interview based techniques provide opportunities to implement identified solutions during the pre-testing period, making them more timely than some other qualitative techniques (see table 2).

a. Literature review - Moderate amount of time required to resolve issues. The time required is dependent on the breadth of the literature search, but some time is required to locate, read and interpret available information.

b. Expert review - Fast resolution of issues. An expert review can be the fastest method for producing results if well co-ordinated.

c. Focus groups - Fast resolution of issues. Focus groups provide broad information on a diverse range of data in a short period because much of the analysis occurs during data collection and the discussion format allows for solutions to be identified.

d. Interviewer debriefing - Fast resolution of issues. Interviewer debriefings produce quick results as some analysis occurs during data collection and solutions can be identified and discussed during the debriefing session.
e. Observational interviews - Moderate amount of time required to resolve issues. Although much of the analysis takes place during the interview, solutions can not be tested until subsequent interviews or even rounds of interviews.

f. Cognitive interviews - Fast resolution of issues. Much analysis occurs during the interview, and the iterative process allows for immediate testing of solutions to problems.

g. Behaviour coding - Slow in resolving issues. Behaviour coding is generally the slowest method for producing results as it requires a substantial amount of data to be collected before analysis can occur. The analysis process is particularly time consuming, involving data entry, qualitative analysis, statistical tests and the actual analysis. Time may also be needed to develop a model or errors.

**Stage of development process**

20 Another key difference between techniques is that they are best suited to different stages of the survey development process (see table 2). The start of the development process involves developing and defining concepts and gathering background information. The early stages include constructing the draft or questions, modules and the middle stage involves turning these questions and modules into a questionnaire. The later stage involves finalising the instrument for field testing and determining field procedures and interviewer training.

a. Literature review - This technique is used by the ABS at the start of the development process, to gather background information.

b. Expert review - Expert reviews tend to be conducted early in the development process to provide some ideas about what sources of non-sample error to focus subsequent testing on. For example, expert reviews are normally conducted after focus groups, but before cognitive interviewing. They requires a draft questionnaire or sections of questionnaire (modules) to have been specified and constructed for critique.

c. Focus groups - Focus groups can be conducted by the ABS at the start of the development process as they need only concepts or topics to have been specified. They can also be useful in gathering information about a draft questionnaire and may be used a bit later in the development process, along with cognitive interviews. As a general rule, however, they are conducted by the ABS prior to cognitive interviews.

d. Interviewer debriefing - The ABS conducts interviewer debriefings either at the start of development to explore feasibility issues, or once a draft questionnaire has been produced to better understand non-sample error related to interviewers.

e. Observational interviews - These tend to be used by the ABS in the middle of the development process because they require a draft questionnaire.
f. Cognitive interviews - This technique is used towards the middle of the development process by the ABS as a draft questionnaire or sections of a questionnaire are required to mimic the question-response process.

g. Behaviour coding - The ABS usually conducts behaviour coding later in the pre-testing process. It is also sometimes conducted as part of a field test (skirmish or pilot test). Although generally conducted after cognitive interviewing, the ABS sometimes uses behaviour coding in conjunction with later rounds of cognitive interviews, for example, to explore mode effects in a telephone interview. When conducted as part of field tests, information from behaviour coding is useful as a guide to what to focus on in interviewer debriefings.

Table 2: Comparison of pre-testing techniques by timeliness of results and stage of development.

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>Timeliness of Results</th>
<th>Stage of Development Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review</td>
<td>**</td>
<td>Start</td>
</tr>
<tr>
<td>Expert Review</td>
<td>***</td>
<td>Early</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>***</td>
<td>Start Early</td>
</tr>
<tr>
<td>Interviewer Debriefing</td>
<td>***</td>
<td>Start Early</td>
</tr>
<tr>
<td>Observational Interviews</td>
<td>**</td>
<td>Middle</td>
</tr>
<tr>
<td>Cognitive Interviews</td>
<td>***</td>
<td>Middle</td>
</tr>
<tr>
<td>Behaviour Coding</td>
<td>*</td>
<td>Later</td>
</tr>
</tbody>
</table>

key: * slower ** moderate *** faster

Aims of test
Range of non-sample errors identifiable

21 When deciding on a pre-testing technique, an important consideration in relation to the aims of pre-testing is the range of errors that can be identified by a technique (see tables 3a and 3b). For example, some techniques provide detailed information about only one area of non-sample error, whilst others provide a more limited but broader range of information.

a. Literature review - This technique can provide limited information on a broad range of sources of error.

b. Expert review- Expert reviews can identify a broad range of errors, including problems with the questionnaire layout, question wording, respondent burden, interviewer considerations. Presser & Blair (1994) found that expert reviews produced the largest and most consistent number of problems and this is consistent with ABS experience.
c. Focus groups - This technique provides a narrow range of information as it can only account for sources of respondent error.

d. Interviewer debriefing - Interviewer debriefings provide a narrow range of information, covering mainly sources of interviewer error and some types of perceived respondent error.

e. Observational interviews - Observational interviews provide a wide range of information, covering sources of respondent error, mode effects and interaction effects.

f. Cognitive interviews - These interviews provide a moderate range of information about all sources of respondent error.

g. Behaviour coding - This technique provides a broad range of information about respondent, interviewer, mode and interaction errors at varying levels of detail.

**Detail of non-sample errors identified**

22 When considering the aims of pre-testing, the level of detail in which each area of non-sampling error can be explored is also important (see tables 3a and 3b). Techniques differ in the type of non-sampling error that they can examine, with some techniques better suited to identifying areas where respondents may have difficulties, whilst others can explore interviewer errors, and still other techniques can look for mode or interaction effects.

a. Literature review - This technique provides limited information about interviewer, respondent, mode and interaction effects. At the ABS, a literature review involves gathering background information about the survey topic, administration and methodological issues. For example, research questions may include 'What similar surveys have been developed?', 'What pre-testing was conducted and what were the results?' etc. Depending on the literature available and targeted, limited information can be gained about all types of non-sampling error.

b. Expert review - These provide some information about interviewer, respondent and mode effects and limited information about interaction effects. The ABS recognises expert reviews as a useful development tool for identifying a broad range of sources of non-sample error. They can provide some information about interviewer, respondent and mode effects and limited information about interaction effects. In addition, expert reviews can provide solutions and recommendations for minimising identified sources of error.

c. Focus groups - Focus groups provide detailed information about sources of respondent errors. The ABS uses focuses groups to explore issues relating to surveys which are new or which deal with complex or ill-defined concepts or potentially sensitive topics. In sum, they are most useful in providing survey developers with information about how to word questions and structure the questionnaire in a way that minimises respondent errors.
d. Interviewer debriefing - This technique provides detailed information about sources of interviewer error and some limited information about respondent errors. Debriefings can identify potential issues with ease and consistency of administration and sensitivity to interviewers as well as provide some information about perceived respondent sensitivity. They can also provide limited information about perceived respondent burden.

e. Observational interviews - These interviews provide detailed information about interaction effects and respondent sensitivity, some information about mode effects and other types of respondent errors. The ABS has found that observational interviews are best used to explore respondent performance and conceptual problems, primarily to identify sources of error resulting from the respondent and/or the instrument. They can provide detailed information about respondent burden and the effects of interaction between the respondent and the instrument. They provide some information about mode effects, respondent sensitivity issues and response errors and limited information about sources of respondent error arising from comprehension, recall, and judgement issues.

f. Cognitive interviews - This technique provides detailed information about the sources of respondent errors and some information about mode effects. Cognitive interviewing serves to assure survey developers that respondents are answering the question survey developers think they are asking. The ABS use cognitive interviews to determine whether respondent error arises from problems of respondent comprehension, retrieval, judgement or answer formation, as well as to identify issues of respondent burden and sensitivity. Cognitive interviews using retrospective probing are also used by the ABS to provide limited information about mode effects.

g. Behaviour coding - Behaviour coding provides detailed information about interviewer consistency, response errors, respondent burden and interaction effects. Some information is available about other sources of respondent error and mode effects and limited information can be gathered about interviewer sensitivity. The ABS has found behaviour coding to be most useful for identifying errors in interviewer administration of the questionnaire and the question-asking process, as well as for identifying respondent fatigue. In ABS experience behaviour coding can provide the most detailed information from pre-testing.
Table 3a: Comparison of pre-testing techniques by sources of error.

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>Interviewer</th>
<th>Respondent</th>
<th>Consistency</th>
<th>Sensitivity</th>
<th>Comprehension</th>
<th>Recall</th>
<th>Judgement</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Expert Review</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Interviewer Debriefing</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observational Interviews</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive Interviews</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Behaviour Coding</td>
<td>***</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Key: * limited information  ** some information  *** detailed information

Table 3b: Comparison of pre-testing techniques by sources of error (continued).

<table>
<thead>
<tr>
<th>TECHNIQUE</th>
<th>Respondent</th>
<th>Mode</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Burden</td>
<td>Sensitivity</td>
<td></td>
</tr>
<tr>
<td>Literature Review</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Expert Review</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>***</td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Interviewer Debriefing</td>
<td>*</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Observational Interviews</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>Cognitive Interviews</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Behaviour Coding</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
</tbody>
</table>

Key: * limited information  ** some information  *** detailed information
Other considerations

23 As discussed above, resource issues and the objectives of testing represent the main factors taken into consideration by the ABS when designing a pre-testing strategy. In total, a pre-testing strategy should combine a number of techniques to optimise the chances of identifying and minimising as many potential sources of non-sample error as possible. In practice this means taking into account the mode of the final survey at some time in the testing process. The ABS considers simulating the final mode of administration to be particularly important when finalising the questionnaire for field tests, but less important during early stages of pre-testing, where the focus is on looking for broader comprehension and response errors.

24 Another consideration when designing a pre-testing strategy is that pre-testing is only part of the overall testing program. Thus, survey developers also need to be aware of how pre-testing fits with subsequent testing objectives and constraints.

Conclusion

25 Whilst there is a wide range of techniques available for use in pre-testing surveys, the ABS relies on a few techniques in particular. At the ABS, interviewer debriefings and cognitive interviewing are the most commonly used pre-testing techniques, with literature reviews, expert reviews and focus groups also widely used. For most household surveys, this combination of techniques provides an efficient and effective method for pre-testing surveys within the available resources. Observational interviews tend to be used when the ABS is developing self enumeration forms and behaviour coding is used only as part of larger development projects or where there are expected to be issues with interviewer-respondent interaction, such as where there are unusual field requirements or sensitive topics.

26 The ABS recognises that although cognitive interviews play an important and increasing role in its pre-testing program, this technique is not always the most suitable technique given available resources and will not necessarily achieve all the aims of testing. Thus, when selecting an optimal pre-testing strategy, cognitive interviewing is considered by the ABS as part of the available suite of techniques, rather than a tool to be used in isolation. Whatever combination of techniques is selected, however, designing a pre-testing program is always a balancing act!
References


