SELF-COMPLETION QUESTIONNAIRES: THE IMPORTANCE OF TESTING (BASED ON EXPERIENCE FROM THE BRITISH NRS)

Hilary Birt, Ipsos-RSL

Introduction

There are many design factors which can influence the results of self-completion readership questionnaires: the length and nature of written instructions and examples; the question order; the layout of the response grid; typescript *vs* masthead representation of titles; and so on. Apparently subtle differences can have an impact. Hence the need for rigorous testing before going live.

This paper will present the methods and results of cognitive testing of alternative self-completion questionnaire designs for the Great Britain National Readership Survey (NRS). The proposal was that a self-completion questionnaire could be used in circumstances where it would prove difficult to obtain response by the standard face-to-face (CAPI) method.

Two versions were proposed for testing – one emulating the question order and filtering of the standard face-to-face NRS questionnaire, the other a shorter version with a different question order and approach.

The criteria for assessing the self-completion questionnaire designs were ease of completion, likelihood of generating 'correct' results (in terms of reflecting the reality of respondents' reading behaviour) and comparability of results with the standard CAPI NRS data.

The tests described in this paper were qualitative in approach and scale and designed to address the first two criteria. This test would then be followed by quantitative field testing, to assess comparability with the standard NRS data, before the self-completion questionnaire would be adopted on the NRS.

Background and objectives

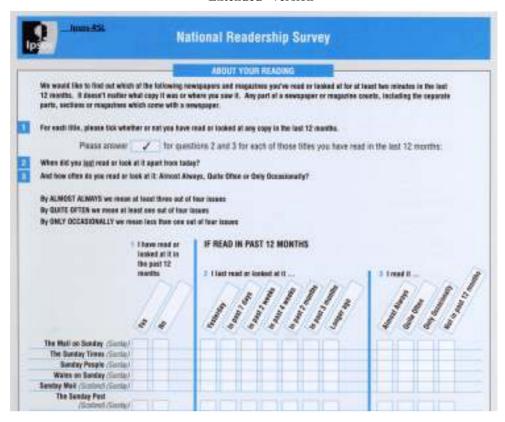
The NRS employs a face-to-face methodology using CAPI. For several years now there has been increasing concern over falling response rates on the survey, especially in the London area where working patterns and accommodation access problems (e.g. entry phones) make the interviewer's task very difficult. In response to this concern, NRS Ltd, through the NRS Development Advisory Group (DAG), commissioned Ipsos-RSL to conduct a Mixed Methodology test in November 1999. The aim of this test was to explore whether recourse to different interview methods, as well as different field management techniques, could improve response rates in certain situations. The findings of this test indicated that the use of a self-completion questionnaire, after attempts at obtaining a personal interview have failed, had the potential to improve response rates. The NRS DAG, with Ipsos-RSL, therefore embarked on a programme of development work to design and test possible self-completion questionnaire versions for the survey.

There were three key issues which we considered when designing a self-completion questionnaire for the NRS:

- The self-completion questionnaire must be clear, straightforward to complete and not open to misunderstandings, as no interviewer would be present to clarify any queries. An over-complicated or visually daunting questionnaire would also not encourage the respondent to complete the task.
- The method by which the respondent answers the questions must be considered as this could have a bearing on the recording of readership claims in a variety of different ways.
- The self-completion questionnaire should, as far as possible, be compatible with the CAPI questionnaire as the data
 from both would need to be combined when finally analysed. It was recognised, however, that the self-completion
 method would, in any case, generate different results from the personal interview method (as with any methodological
 change).

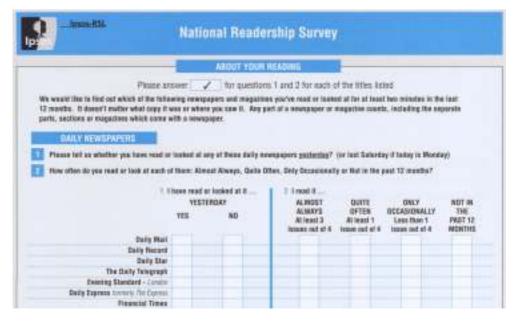
As a result of discussion of these issues, it was decided to test qualitatively two alternative questionnaire versions. The first, which we term the "extended" version, incorporated a more detailed in-depth form of the readership questions, more akin to the present NRS questionnaire (incorporating an initial read past year filter and with a full recency scale common to all publication frequency groups):

"Extended" version



The second, which we term the "short" version, used a simplified version of the readership questions, similar to that already used on established self-completion surveys amongst business audiences conducted by Ipsos-RSL. Whilst this approach was well established for these surveys, it was important to bear in mind that an appropriate solution for business audiences might not be appropriate for an all adults sample like the NRS. Key features of the "short" version were that there was no read past year filter question and the recency question had a simple yes/no response for each publication frequency group to generate a measure of Average Issue Readership (AIR), that is yesterday for dailies, past 7 days for weeklies and so on:

"Short" Version



A further difference between the two versions was the grouping of titles: by 'generic' groups of titles on the "extended version" (eg. gardening titles), following the NRS pattern; and strictly by frequency on the "short" version.

Methodology

The qualitative study consisted of 30 interviews conducted at a central location over two days in September 2000. Respondents were pre-recruited to set quotas in order to give a demographic spread in terms of sex, age, social grade and working status.

On arrival, the respondent was briefed on the task by an Ipsos-RSL interviewer and then asked to complete one of the questionnaire versions (these were alternated such that each one was completed first by roughly half the sample) on their own. On completion they were given the other questionnaire version to complete. Each respondent was observed and timed on the completion of each version, with regards to both the readership and marketing/demographic sections. (The marketing/demographic questions were only completed once, for the first version).

A debrief was then conducted by an Ipsos-RSL executive focusing on two key themes:

- the (perceived) ease with which the different versions were completed (respondent viewpoint).
- an evaluation of how the different versions worked in terms of the question mechanics and whether they reflected actual reading habits

Ease of completion

Both versions of the questionnaire were generally acceptable to respondents: there were very few queries during completion, which indicated that the task presented to them was relatively clear; and on further questioning, the reaction was that both versions were fairly straightforward. There were, however, a few specific negative comments.

Common to both versions was the observation that the instructions were rather long and wordy. Most respondents claimed to read them, but it is important to remember that this was a somewhat artificial situation in which respondents would be likely to feel obliged to fulfil the task "correctly": the question is whether, in a normal in-home situation, respondents would take the time to read the instructions so thoroughly. With both versions there was a tendency for respondents not to refer back to the instructions after first reading: this puts a lot of weight, therefore, on the column headings to convey all important information, to enable respondents to complete the grid correctly.

A small number of respondents were confused as to whether the questions related to publications bought or just read. This tendency to think that what is required is information on "primary reading" is not unusual, but is important to recognise in terms of clarifying the instructions, as far as possible.

The existence of the 'no' column on Q1 was queried by some (RPY -"extended", AIR - "short"), particularly on the "short" version: respondents felt that if they made a positive 'yes' claim, that, by definition, all the others would be 'no', without needing to fill it in.

Respondents divided fairly evenly between those preferring the "extended" and those preferring the "short" version. Sometimes it was a function of the order in which they had completed the two versions (the second version to be completed was sometimes found to be more straightforward). However there were also clear reasons behind people's preferences, which are explained below.

Key benefits of the "extended" version, from the respondent viewpoint were as follows:

- The common recency scale for all publications meant that respondents did not have to change their frame of reference in the course of completing the questionnaire.
- The detailed recency scale, although regarded by some as more complicated compared to the "short" version, acted as a better memory prompt to work out when it was that a given publication had last been read; furthermore, the detailed scale was felt to provide a better reflection of their reading patterns.
- Some felt that the fact that the boxes were closer together made completion easier as it was easier to read across for a given title.
- The angled column headings, although disliked by some, were liked by others because they stood out, particularly
 against the white background.

On the negative side for the "extended" version...

- The frequency question was perceived by some to be almost a repeat of the recency question, highlighting confusion in respondents' minds over "how recently" and "how often".
- The 'not in past 12 months' column on Q3 was felt to be redundant given respondents had already indicated 'read past year' at Q1; the presence of this code, therefore, caused some confusion.
- Whilst some found the closeness of the boxes a bonus, others felt it made the questions look more complicated and generally unappealing.
- Again, whilst some liked the angled column headings, others disliked them as being more difficult to read.

Plus points for the "short" version, according to some respondents, were as follows:

- The grouping by frequency aided completion, enabling the respondent to focus on, for example, weeklies all together.
- The smaller number of boxes, compared to the "extended" version, was preferred by some as it looked lighter in appearance and easier on the eye.
- The fact that the numerical scale was included in the frequency column headings was a benefit as respondents did not have to refer back to the instructions for an explanation of the scale.

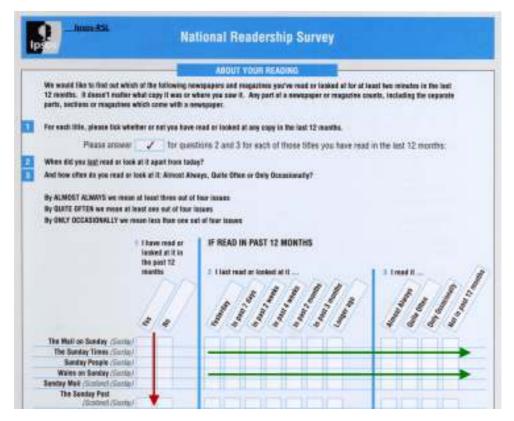
On the negative side for the "short" version...

- The changing recency period by publication group was felt, by some respondents, to be confusing.
- There was some uncertainty as to whether to fill in the 'no' column at Q1.

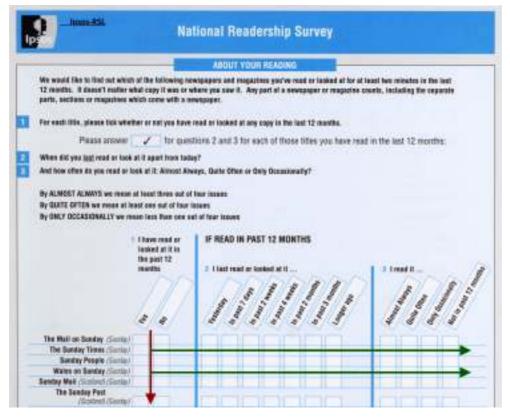
How were they completed?

It was important to observe and assess how the two questionnaire versions were completed. This is important as the mode of completion might have a significant bearing on the results.

The most common way in which the "extended" version was completed was vertically for Q1 then horizontally for Q2 and 3 together for each 'yes' title at Q1:



Some respondents filled in Q2 and 3 for each 'yes' title as they came to it at Q1, as opposed to going all the way down the page and then returning to the 'yes' titles:

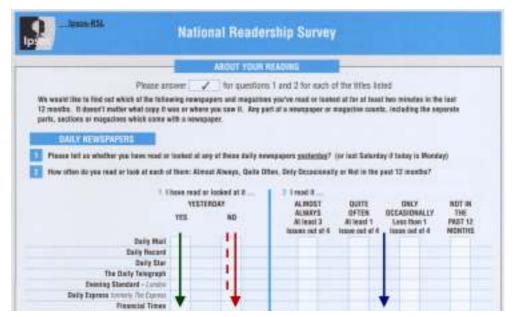


In some cases the 'no' column at Q1 was missed altogether: there was a tendency, already mentioned, to think that the lack of a 'yes' claim, by definition, meant 'no', without having to fill it in religiously.

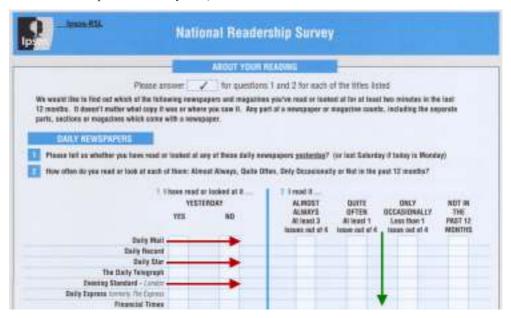
Some respondents completed the questions page by page, some by group of publications at a time.

A small minority completed Q1-3 horizontally by title.

The majority completed the readership questions on the "short" version vertically. Some went down the 'yes' column, then the 'no' column (although this was missed by a fair number of people), then Q2:



Some respondents completed Q1 'yes/no' together by title, then back to Q2 and down the page again (the balance between these variations on vertical completion was fairly even):



There was a tendency for respondents to treat Q1 recency as a filter for Q2 frequency. The effect of this is discussed later in this paper.

A very small minority of respondents completed the questions horizontally, title by title.

Generally the questions were completed by frequency group.

Evaluation of Questionnaires

Evaluating the findings from a research viewpoint, taking into account both the findings of the executive debrief interviews and the analysis of the questionnaires, more problems were observed with the "short" version, compared with the "extended" version which, generally speaking, appeared to be less prone to respondent error.

The "extended" version had a number of research benefits over the "short" version (it should be noted that these are similar to, but <u>not</u> the same as respondents' own perceptions of the benefits):

- The key benefit of the "extended" version, and one not particularly highlighted by respondents themselves, was the fact that Q1 'read past year' acted as a clear and effective filter to Q2 and 3. The fact of starting with a very wide, effectively 'ever read', filter tended to help respondents in prompting their memory of even very distant reading occasions; this was then narrowed down by Q2.
- The detailed recency scale acted as an effective memory prompt, compared to the single recency period "short" version, by which you were either 'in 'or 'out' for a given title.
- The common recency scale for all titles meant there was a consistent frame of reference throughout the questionnaire. It was observed that this would make the "extended" version less prone to error, as it did not rely on respondents noting the changing period as the "short" version did.
- And finally, the proximity of the boxes had the effect of bringing the questions together, thus encouraging horizontal completion at least for Q2 and 3.

The only observed downside of the "extended" version was the fact that the numeric frequency scale was not consistently referred to as it was not included in the column headings.

There were, however, more problems observed with the "short" version:

- There was a tendency for Q1 to be treated as a filter to Q2. In a number of cases our observation was that respondents tended to overclaim AIR in order to say that they had read the title, and to 'justify' continuing to complete Q2. The alternative scenario was that respondents were claiming AIR correctly but not filling in Q2 for non-AIR titles, leading to under claiming of RPY and cumulative readership. A broad analysis of the data patterns confirms our hypothesis that the trend was towards the former: over claiming AIR in order to fulfil the perceived Q1 filter.
- The Q1 AIR 'no' column was often left blank: respondents tended to regard it as unnecessary, believing the absence of a 'yes' must inevitably mean 'no' (this effect may have been exacerbated in the "short" version by the relatively wider spacing of the boxes compared with the "extended" version).
- The single recency period did not always act as an effective memory prompt: as opposed to structuring the memory process as with the "extended" version scale, it was a question of either being 'in' or 'out' which, it was observed, could lead to over claiming in order to record positive reading claims.
- The changing recency period by publication interval confused some respondents: this is an important observation as respondents are not necessarily au fait with the publishing intervals of the different publications they read, even those they read regularly.
- A final observation was that the spaced out layout seemed to encourage vertical completion (which would tend to be more prone to error, as respondents might miss publications when going back up the page).

Issues for Further Research

This test raised a number of key issues which had to be considered carefully before proceeding to quantitative testing:

- The length and layout of the instructions to maximise clarity.
- The layout of the column headings for clarity and to include all important information, so the respondent does not
 have to refer back to the instructions.
- Incorporating clearer routing instructions between the readership questions.
- The publication groupings and layout: should they be organised by frequency or by generic type?

A further issue for the NRS DAG was whether mini mastheads should be incorporated on the questionnaire to aid recall.

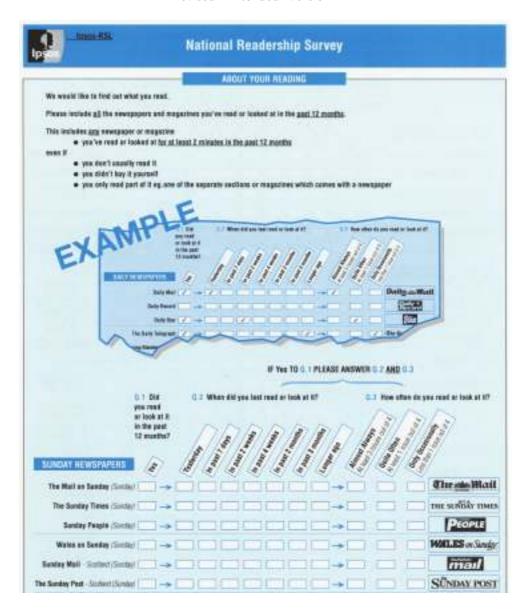
Whilst both versions were generally acceptable to respondents, and recorded timings were very similar, the extended "version" appeared to be less prone to error. However, it was agreed at the NRS DAG that the short version, improved with substantial design changes, should still be included in further testing.

As a result, it was agreed to test qualitatively two further versions: one version which incorporated minor amendments to the "extended" version; the other a considerably altered "short" version.

In both versions the instructions were substantially simplified and reduced, focussing on the key points. In addition, a completion example was included in both versions.

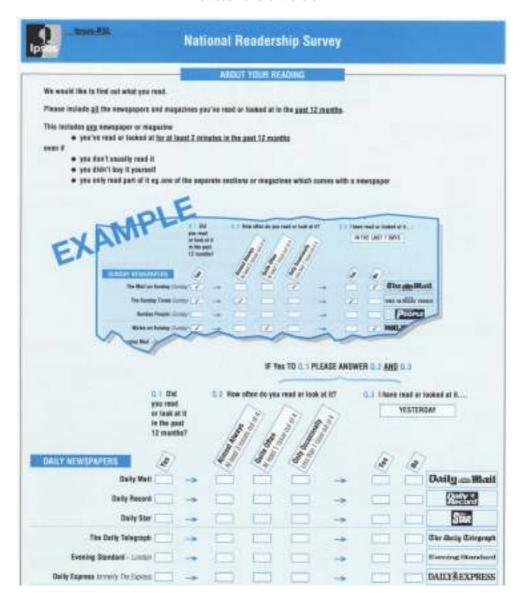
The key changes to the "extended" version were: a simple 'yes' rather than 'yes/no' read past year filter, plus the inclusion of arrows in the grid, to guide respondents to complete Q1-3 horizontally; the inclusion of both verbal and numeric frequency scale elements in the actual column headings (Q3). A final change was the incorporation of mini mastheads on the right hand side of the grid, typescript remaining as the 'primary' prompt on the left.

Revised "Extended" version



The new "short" version differed substantially from that tested previously: a read past year filter was added (positive 'yes' only); and the order of the recency (AIR) and frequency questions was switched so that the former was not treated as a filter to the latter.

Revised "short" version



Further testing

The two revised versions were again tested qualitatively with a sample of 22 respondents.

The design adaptations to both versions were observed to overcome most of the problems identified in the earlier test, notably: the simplified instructions; the comprehensive information included in the column headings; the read past year filter on the short version; and the simple 'yes' read past year response option which was found to trigger horizontal completion.

Interestingly, whilst the simplification of the instructions was found to be positive, the inclusion of the example was concluded to be redundant, perhaps on account of the improved instructions and 'sign-posting' in other ways and also on account of it 'getting in the way' somewhat between the instructions and the main response grid.

An interesting outcome of the positive only read past year response was the fact that blank pages were not unusual. This led to some feelings of guilt with respondents querying whether they had completed the task 'correctly'. It was felt that a tick box on each page for "not read any publication on this page in the past 12 months" would help.

Mastheads

The findings in relation to the use of mastheads were of great interest as this element had not been included in the earlier testing.

Respondents were fairly evenly split between those using and those not using the mastheads. Where used, they tended to be used in a secondary function, as a check when the respondent was not sure about a given title.

A minority used the mastheads as the primary stimulus which, it was observed, could lead to problems as they tend to prompt completion from right to left: on the 'short' version this could lead to confusion of the average issue readership with the read past year question and a tendency to omit completely to fill in Q1 and Q2.

A minority actively disliked the presence of mastheads as superfluous and distracting.

On the basis of these observations it was concluded that, on balance, the presence of mastheads was positive but that it was important that typescript representation remained on the left hand side as the primary stimulus.

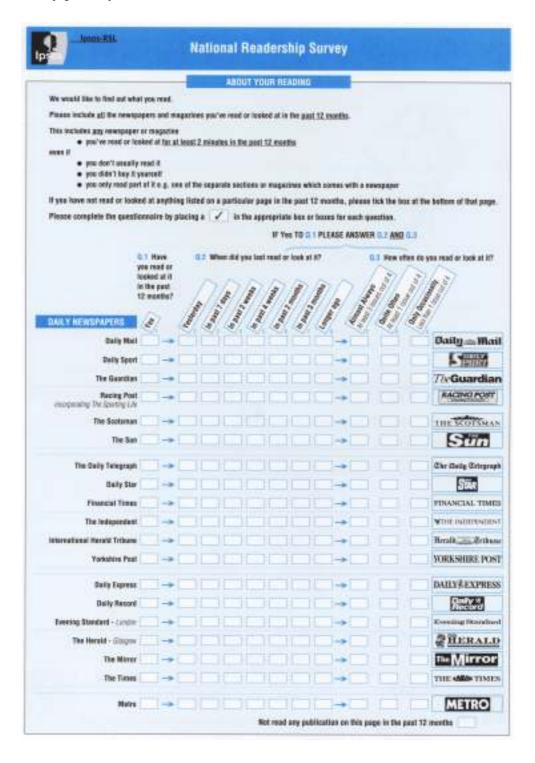
Comparative evaluation of revised versions

Whilst both versions showed improvements from the earlier test work, on account of the design changes, the balance still came out in favour of the "extended" version, for the following key reasons:

- the extended recency scale offered more options which made respondents think and give a more specific response, and gave respondents the feeling of giving more 'complete' information about their reading.
- the question order was felt to be more logical with the 'precise' recency question leading to the more 'qualitative' frequency question.
- and the grouping of titles in generic groups rather than by frequency was felt to be more likely to reduce the problem of title confusion.

Outcome of testing

As a result of these two stages of testing, the NRS DAG agreed to proceed to quantitative testing of the "extended" version with a small number of further design changes, notably: deleting the example; and adding a tick box for 'not read any publication on this page in the past 12 months'.



Lessons to be learnt

The most important lesson to be learnt from this test work was that apparently subtle design aspects can have a real impact on how a questionnaire is completed: the design of the response grid is key in determining how respondents complete it (horizontally or vertically) and the potential impact of this on the results; question order is also a major factor; and clarity of the instructions, both introductory remarks and the sign-posting in the grid, is essential.

As researchers, it was of great interest to observe that it was not necessarily the 'short' approach to questions and question responses which proved more respondent-friendly: as described earlier, the perceived logic of the question order and having a range of response options to choose from made respondents feel their reading was being better reflected.

Qualitative observational testing provided invaluable input into the design of a self-completion questionnaire for the NRS. This small scale testing highlighted important aspects affecting ease of completion and mode of completion which could have an important bearing both on response and on the actual readership data. However, it remains the job of the quantitative field test (in progress at the time of writing) to determine the outcome in terms of the readership results and how comparable the results of the chosen design are to the standard CAPI NRS data.