

2 GENERAL FEATURES

Intelligent Character Recognition (ICR)

2001 Census examples

4 When were you born?

day (eg 15)	month (eg 7)	year you were born (eg 1966)
<input type="text"/>	<input type="text"/>	<input type="text"/>

8 On the night of Tuesday 6 March, which address are you at?

at the address you gave in question **5**

at another address. Print the address you are at including, if possible, all of these:

- flat number (if it is a flat)
- street number and name
- suburb or rural locality
- city, town or district

Background

The addition of Intelligent Character Recognition (ICR) boxes for all write-in questions appearing on the census forms was identified as a key development objective for 2006. ICR boxes have the potential to reduce costs by making the processing of completed forms faster and more efficient. These boxes were included for selected questions in 2001 with good results.

ICR boxes were introduced to test forms at the start of the development for 2006 and monitored throughout the testing programme to evaluate how well respondents were able to use the boxes, follow instructions and provide answers that would be readable by scanning equipment.

Development objectives/plan

The objectives identified for the 2006 redevelopment were to:

- maintain data quality (ensuring boxes did not obscure or divert attention from available response options)
- assess how well respondents would be able to use these write in boxes

- determine the optimum size of boxes to maximise the number to be included on the forms while maintaining the legibility of answers
- detect any data quality effects associated with the fields, such as deterioration of responses, truncation of responses or any other impact on data
- determine whether the fields contribute to respondent fatigue.

To meet these objectives, a comprehensive programme of testing was conducted which included cognitive testing with respondents in both English and Māori, analysis of data from field tests and consultation with key stakeholders.

Cognitive testing

The cognitive testing programme was conducted between April 2003 and May 2005 with respondents of varied age, gender and ethnicity. Interviewers observed respondent form-filling behaviour and asked a series of follow-up questions.

The ICR boxes were redeveloped and their use was monitored during all four rounds of the cognitive testing programme.

Cognitive testing results

Round one: April and May 2003

In round one ICR boxes were added to all write-in questions. To maximise the number of boxes able to fit on the form, boxes extended from the left to right margins of each column and were 5.5mm high x 4.5mm wide.

In cognitive testing respondents tended to write legibly in the boxes provided. However, many respondents failed to write in upper case and often drew characters which extended beyond the borders of the box. Occasionally respondents would ignore the boxes altogether and write freely across the space provided, squeezing several characters into the same box. Answers commonly exceeded the number of characters allowed, and truncated answers were common. There was a very strong tendency for people to write individual words on new lines, rather than wrap answers.

Boxes generally took up a large amount of room and the amount of white space on the page tended to dominate the form and draw the respondent's eye, so that they missed seeing other options available to them. Sometimes respondents would incorrectly use the boxes to write in answers for alternative options. Occasionally respondents commented on the 'daunting' appearance of the form.

Rounds two, three and four: June to December 2003, March to July 2004 and November 2004 to May 2005

In later rounds of cognitive testing the form was increased in size from an A4 sheet to the A3 size used in 2001. This allowed a greater number of boxes to be included on the forms and meant that designers could indent boxes from the left margin so that they did not break the main 'navigation path' through the form or obscure response options appearing underneath the boxes.

Although this appeared to reduce the number of times response options were missed, problems with the quality of write-in answers remained.

Towards the end of testing box size was reduced to 5mm x 4mm and respondents' answers were evaluated for readability. There appeared to be no noticeable reduction in the number of legible answers provided and a greater number of boxes were included within the space.

Field test results

Field test: March 2003

Forms were not scanned in March. However a sample of test forms – selected from a single Auckland meshblock – was examined. The following problems were identified:

- respondents writing in lower case
- respondents writing outside the scannable area, eg underneath, alongside and outside the boxes
- respondents not using one letter for each box, ie cramming letters into the space allowable
- poor handwriting (badly formed letters or letters of obscure shape).

Some of these problems occurred because there were too few boxes for respondents to complete their answers.

Field test: November 2003

Forms were not scanned in November. However, a sample of test forms was again examined. Results suggested that respondents were generally using the boxes very well. In contrast to March, most responses were contained within the box space. For most questions there appeared to be enough boxes for respondents to complete their answers, although in some cases respondents wrote outside the scanning area to finish writing in their answer. A large number of respondents failed to wrap as instructed.

Field test: March 2004

To investigate the impact of reducing the ICR box sizes, the forms development team produced two forms for the March test. Each of these forms (version A and version B) was produced using a smaller-sized form. However, version A used a larger sized box (5.5mm x 5mm) and version B used a smaller sized box (5mm x 4mm). Analysis from this test indicated that there was a reduction in recognition rates using the smaller box size.

Dress rehearsal: 8 March 2005

ICR boxes were scanned for the dress rehearsal and a sample of forms was also examined. In the vast majority of cases, respondents were using the constrained boxes as instructed. While recognition rates were not as good as expected, answers were able to be corrected during processing.

The clear majority of answers provided in the dress rehearsal were able to fit within the number of boxes provided on the forms, although there were still cases of respondents using lower case, writing across boxes and failing to wrap.

Other testing

Testing of the ICR boxes also involved some qualitative testing using eye-tracking software to determine how well respondents followed the questionnaire path. The number of tests conducted was small and therefore results were indicative only. However, a comparison of formats using both smaller and larger fields suggested that respondents were more likely to miss options when the boxes were positioned hard left and less likely to miss options when the boxes were indented.

