Question development history for 2018 Census: Gender Identity

Overview

Summary	
Key driver for question development	• New Content
Quality priority level:	• TBD
Outcome from question development	• Cognitive testing

1 – Purpose

The purpose of this document is to capture the question development process for the Gender Identity variable, including findings from waves of testing conducted in 2015-2017.

The 2018 variable specification provided by the Customer Needs and Data (content) Team provides the background and scope of this variable.

This document is intended for use within Statistics New Zealand.

2 – Background

Definition from statistical standard:

Gender identity is an individual's internal sense of being wholly female, wholly male, or having aspects of female and/or male. Gender identity is understood to refer to each person's deeply felt internal and individual experience of gender, which may or may not correspond with the sex recorded at birth (adapted from International Commission of Jurists, 2007, p6). A person's gender identity can change over their lifetime, and can be expressed in a number of ways and forms. This expression includes outward social markers, such as name, clothing, hairstyles, mannerisms, voice, and other behaviours.

3 – Design differences between paper and internet forms

n/a

4 – Findings from testing (or review) and rationale for revision

These tables summarise in chronological order the versions of this question set that were tested (or reviewed), along with brief findings, and rationale for revision.

Reasons for variables being omitted from a sprint may include: the content need or question design is not ready, or the variable is not a focus for that sprint (eg it is not suited to the target respondents), or the sprint is not a test of content.

Summary of sprints this variable has been tested in, plus testing type and mode type:

- Sprint 4; cognitive testing and mass completions of paper forms
- Sprint 5; cognitive testing and mass completions of paper forms

Sprint 4; Cognitive testing and Mass Completions of paper forms

March 2016
Christchurch and Wellington

Aim:

General

- The primary objective of the testing is to provide recommendations to inform a Go/No Go decision on future development and testing of proposed 2018 Census content.

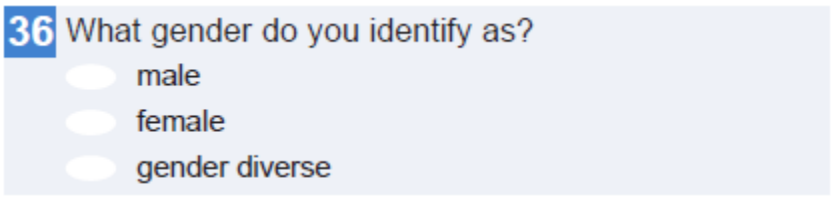
Targeted LGBTQI+ Testing

- Gain insights into acceptability, especially within a census context.

Respondents

The cognitive test participants included members of the public, students, and people with step family. The mass completion test participants included rural fire fighters, secondary students, and tertiary students (including young parents' college and ESOL students).

The question design tested in sprint 4 was new:



Sprint 5, Cognitive testing and Mass Completions of paper forms

March 2015
Wellington and Christchurch

Aim:

General

- The primary objective of the testing is to provide recommendations to inform a Go/No Go decision on future development and testing of proposed 2018 Census content.

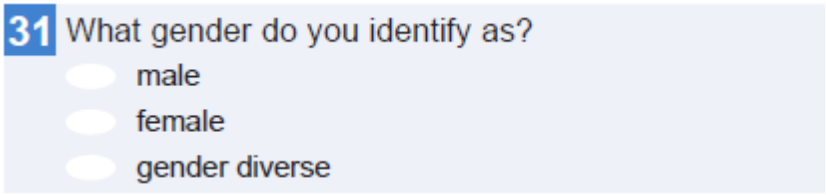
Targeted LGBTQI+ Testing

- Gain insights into acceptability, especially within a census context.

Respondents

As with the previous sprint, cognitive test participants included members of the public, students, and people with step family. The mass completion test participants included secondary and tertiary students, Age Concern, retirement village residents, and a private workplace.

The question design tested in sprint 5 was:



Sprint 5 – Cognitive testing and Mass Completions of paper forms – Findings

General Population

- Many respondents had a sense of deja vu' when they got to the gender question. In many cases the sex-based routing given in the 'babies born alive' question immediately preceding gender in Sprint 5 alerted respondents to there being two similar looking questions in the form. In Sprint 4, many more respondents didn't have a sense of Deja-vu due to the questions being spaced far apart on the form and the absence of sex based routing prior to the gender question.
- Some respondents realised there was a difference between the sex question and the gender question, but others remained confused.
- Some respondents understood the difference between sex (biological) and gender (identity), but still had difficulty understanding that the categories were not dichotomous.
- Many respondents felt that the distance between the two questions (sex and gender) was odd and wondered why the gender question was not placed together with the sex question on the form.

LGBTQI+ testing

- Respondents in the LGBTQI community liked the inclusion of this question and several felt the question was clearer than the sex question.
- Some trans respondents selected 'gender diverse', while others selected either male or female.
- Some trans respondents selected either male or female for sex and the same for gender, so it
- was not possible to identify them as trans from looking at the form.
- Some respondents queried if they could select multiple responses.

SPRINT 5 – RECOMMENDATIONS

- **Recommend inclusion of Sprint 5 question version for Volume Test with help information available.**
- While testing to date shows general acceptance of the gender question, we understand that the testing done has likely gained the perspectives of the most compliant respondents and true feelings may be masked due to the presence of the interviewer.
- Inclusion in the Volume Test will allow us to understand reaction to the gender question more fully in a larger scale and non-observed test environment. Examples may include volume of calls made to call centre, access to help information and for paper forms, multiple response and form annotations.
- Recommend development of some kind of note to explain and affirm that for many people 'sex' and 'gender identity' will be the same thing as a means to reassure respondents we are not in fact asking the same question twice.
- Question as to whether the information we can collect will meet information need expressed in topic specifications is still an issue and one we need to take advice on from Customer Needs and Data.

- **Recommend removal of the 'babies born alive' question in the context of the sex and**
- **gender questions.**
- Both 'babies born alive' and 'gender identity' are questions that must come after the routing to identify the NZ Resident Adult population as neither questions are appropriate to ask of children. Further, if asked, we recommend 'babies born alive' question precede the 'gender identity' question so 'male' respondents are routing on the basis of their sex, rather than have the complication of which basis to route. However, this increases the sense of 'deja vu' for the majority of respondents for whom biological sex and gender identity are the same.
- Regardless of the effects of asking babies born together with gender identity, it is QM&Ds view that it would be preferable to remove 'babies born' in the interests of reducing overall respondent burden, especially given the continued and consistent feedback we get from respondents regarding the insensitivity of the question.
- Recommend some advice from experts in Māori language/Māori world-view in the implications for asking in Māori language specifically and of Māori respondents in general.

5 – Data quality

<Expectations based on testing, known issues, question interactions (suggested edits). To be completed towards the end of QMD testing>

Appendix 1: testing methodology

Research objectives

The broad research objectives of testing may vary with each sprint, but generally are to:

- Understand how well individual questions and key concepts/definitions are understood by respondents
- Understand how well individual questions and the overall form design enables respondents to answer quickly and accurately
- Understand how new and changed questions may impact on other questions in the forms
- Understand respondent burden
- Understand public attitudes to new and changed questions which may influence their willingness to answer

Topics or questions may be allocated as a primary or secondary focus or not a focus of testing in a given sprint. This depends on the priority of the variable itself and how well it has tested previously.

Desktop review (paper and online)

Questions and questionnaires (paper and/or online) are reviewed before they are tested with respondents. The aim of desktop review is to:

- Check whether the forms accurately match content and design specifications;
- Identify any usability issues in the online forms (across a range of devices, operating systems and browsers);
- Identify any potential issues that should be subjected to further testing with the public.

Test participants

Testing aims to include people from a wide range of backgrounds, with a mix of age, sex, ethnicity, income, employment status, etc. However an individual sprint may target respondents with particular characteristics, for example, students, people who have children or stepfamily, Māori, or tenure (renting, home owners, etc).

Test participants have been recruited using a variety of methods. These have included flyers posted in public spaces such as libraries and YMCAs, Twitter and Facebook posts, contacting community groups eg LGBQTI+, Step Family Network and the Retirement Village Association.

Testing methods

Three testing methods have been used, each with a different focus.

Cognitive testing

This is a qualitative, observational research method that helps identify problems with questionnaire design. This methodology involves one-to-one interviews where respondents complete a questionnaire. It uses techniques such as concurrent probing, retrospective probing and think-aloud to highlight how respondents get to their answers and how they interpret certain terms.

Cognitive tests last around one hour, during which the first 30-40 minutes will involve the researcher observing the respondent completing their dwelling form and individual form. The remaining 20-30 minutes will take a semi-structured interview approach. This time will be used to probe in-depth on the focus questions described in this plan, which are relevant for respondents.

Mass completion + group interview

Mass completion tests involve asking a large group of respondents to complete a questionnaire unobserved, in a supervised environment. Mass completion is a useful diagnostic tool to confirm suspicions about a particular design or uncover unexpected reactions to questions using a larger group of respondents.

Mass completion and group interview will last about one hour. In the first half of the session, respondents will be asked to complete one or both Census forms. The remaining time will be used to probe in-depth on the focus questions described in this plan, which are relevant for respondents. The same semi-structured interview protocol can be used for cognitive testing and group interview.

Usability testing (online)

User testing involves one-to-one interviews where respondents complete a set of given tasks (e.g. complete household set up page, complete Individual/Dwelling form) on a device ie a tablet, smartphone or desktop. It is a qualitative, observational research method used to identify problems with a user interface. User testing employs think-aloud, concurrent probing, and retrospective probing techniques to understand how the design of the user interface impacts on the user experience.

Analysis

From sprint 7 onwards, findings were coded to approximately 20 codes, which were further summarised into themes:

Table: Analysis of testing findings – codes and themes used

Codes	Themes	Theme Description
Total nonresponse due to sensitivity	Sensitivity	Relates to how and why respondents perceive question content to be sensitive to themselves and other people. Sensitivity is often based on the individual person's personal experiences, worldview and personal values and can affect their willingness to respond.
Protest response		
Selection of 'object to answer' response		
Reluctant response		
Sensitivity on behalf of others		
Questioning why we ask	Value / Value +	Relates to the explicit or implicit value judgements that respondents make about a question and whether they perceive it as having value, or not. Whether respondents perceive a question to have value or not will affect both their willingness to answer and the quality of
Questioning use of data		
Willingness to answer based on value judgement		
Positive comment volunteered regarding info need		

		their response should they choose to answer.
Difficulty in recalling the requested information	Burden	Relates to the ease with which respondents are able to answer questions and the extent to which they have a positive respondent experience. There are many aspects of respondent burden which respondents may experience when answering questions. Some of these arise from ambiguous or unfamiliar terms or concepts in the questionnaire, while others may be a direct effect of the poorly designed question or form.
Difficulty in interpreting the question		
Difficulty in fitting their answer into the response formats/ categories		
Confusion or difficulties arising from interactions between questions		
Effort required to answer		
Missed routing instructions	Error	Relates to causes of respondent error that can affect data quality and reliability. Sources of error usually arise from poor question and form design, but may also include contextual factors specific to the respondent which can't be controlled for.
Instructions missed or incorrectly followed		
Subjective response		
Proxy response error		
Guesses		
Poor question construction	Defective design	Relates to respondent burden and error, specifically arising from poor question and form design. A fundamentally defective question or set of questions may negatively impact on data quality and/or the user experience.
Dissatisfaction with question/response options		
Visual design of form		

Testing collects information about people's willingness and ability to answer. Not all of these findings will result in alterations to the questionnaire, and any changes that are made may not necessarily resolve the issues found.