# Comparing education and training information in administrative data sources and census

Census Transformation

Emily Shrosbree

New Zealand Government

**Disclaimer**
This paper represents the views of the author. It does not necessarily represent the views of Statistics NZ and does not imply commitment by Statistics NZ to adopt any findings, methodologies, or recommendations. Any data analysis was carried out under the security and confidentiality provisions of the Statistics Act 1975.

**Liability**
While all care and diligence has been used in processing, analysing, and extracting data and information in this publication, Statistics New Zealand gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

**Citation**
Shrosbree, E (2015). *Comparing education and training information in administrative data sources and census.* Available from www.stats.govt.nz.

ISBN 978-0-908350-25-4 (online)

**Published in December 2015 by**
Statistics New Zealand
Tatauranga Aotearoa
Wellington, New Zealand

**Contact**
Statistics New Zealand Information Centre: info@stats.govt.nz
Phone toll-free 0508 525 525
Phone international +64 4 931 4600
www.stats.govt.nz

# Contents

# List of tables and figures

## List of tables



## List of figures

# 1 Background

## Census Transformation in New Zealand

In March 2012 the New Zealand Government agreed to a Census Transformation strategy. This strategy has two strands:

- a focus in the short-to-medium term on modernising the current census model and making it more efficient
- a longer-term focus on investigating alternative ways of producing small-area population and social and economic statistics. This includes the possibility of changing the census frequency to every 10 years, and exploring the feasibility of a census based on administrative data (Statistics New Zealand, 2014a).

The next census in 2018 will be significantly modernised, including an online completion target of 70 percent and re-use of administrative data to support collection and processing.

Continuing to meet critical information needs must underpin decisions on the future of census. Investigations into the long-term direction for census are focused on developing an understanding of future census information requirements, and the ability of administrative sources to meet those requirements.

Read more about Census Transformation in New Zealand.

## About this paper

This paper is part of the Census Transformation work. It is one of a series of investigations into identifying and exploring the potential for administrative sources to provide census-type information.

One important reason for having a census is to provide information about the characteristics of the population, such as information about education and training. Previous Census Transformation work has identified potential to provide census-type information from administrative sources such as Ministry of Education enrolment and qualification completions data.

As part of our investigation into that potential, this paper compares results from the 2013 Census with estimates produced from administrative sources for ‘highest qualification’ and ‘study participation’. Conceptual and empirical comparisons show similarities between the two sources. These comparisons provide a better understanding of the quality of the administrative data, and contribute to Statistics New Zealand’s work to transform the census model.

Overall, the administrative sources investigated show good potential for providing education and training information about those who have interacted with the New Zealand education system, and particularly for those with recent interactions. For some groups, information in administrative sources may provide higher-quality information than the census, while for others this is not the case. There is little to no information about qualifications obtained overseas and it is not possible to positively identify those with no qualifications or those who are not studying.

While the census may be more likely to be affected by respondent or processing error, it is still the only source that provides high coverage of all qualifications for the whole population.

# 2 Introduction

This paper describes our preliminary analysis of education and training information available from administrative sources. It builds on findings from O'Byrne et al (2014) in the context of future censuses.

Education and training information currently collected by the census has a range of uses important to social and economic policy. For example, the Ministry of Education uses qualifications information to determine decile rankings for schools receiving government funding and to measure the impact of educational reforms on skill levels.

Census education and training information consists of:

- qualifications gained by individuals ('highest qualification')
- identification of individuals engaged in study at the time of the census ('study participation').

These education and training variables are available for small geographic areas and can be cross-classified by information about the individuals in the household, and about the dwelling. The census is unique in its ability to provide these kinds of detailed breakdowns and information for small geographic areas across the whole country.

In a first broad look at the potential for administrative data to produce the social and economic (attribute) information currently provided by the census, O'Byrne et al (2014) assessed 'highest qualification' as unlikely to be satisfied by administrative data, and 'study participation' as likely to be satisfied. The main sources identified were educational enrolment and qualification data from Ministry of Education (MoE). This first assessment using scores against five quality measures was based on metadata and intended to be indicative only.

## Aims and scope

Our overall aim was to investigate to what extent census education and training information can be derived from existing administrative data. Three research questions guided this work:

- What is the quality of education and training information in administrative sources?
- To what extent can the data requirements for individuals' education and training attributes currently obtained from census be satisfied using available administrative data sources?
- What would be required to improve the potential for administrative sources to estimate census education and training attributes?

This paper provides reference information about the statistical concepts and administrative data sources relevant to education and training. It presents findings from analysis comparing information in the census with administrative sources.

Our investigation focused on the potential for administrative data to provide the education and training information produced by the current census, with the 2013 Census as the reference. We recognise that education and training information needs may change over time, and that the census may not currently fully meet existing needs.

The administrative sources we investigated were limited to those available in Statistics NZ's Integrated Data Infrastructure (IDI) as at May 2015. These sources include all the main government sources related to education enrolment and achievement, and allow estimates of education and training variables to be derived.

‘Highest qualification’ is the key qualification variable investigated in this paper. In the census, ‘highest secondary school qualification’ and ‘post-school qualification’ are also collected and used to derive ‘highest qualification’. We make no attempt to derive ‘highest secondary school qualification’ and ‘post-school qualification’ separately here.

We limited our analysis to national level, and to grouped qualification level in line with the New Zealand Qualifications Framework. Field of study was not investigated.

The estimates of highest qualification and study participation we derived using administrative sources are for the purpose of comparison with census information only.

# 3 Method

The method used to evaluate the potential for producing education and training information from administrative data sources, specifically the IDI, involves:

1. Describing the formal statistical concepts (section 4) relevant to education and training used in official statistics. Statistical standards and classifications provide the concepts and definitions against which both census and the administrative sources are compared.
2. Describing the data sources (section 5) used in this investigation – 2013 Census and the IDI – and developing a method for deriving estimates of 'highest qualification' and 'study participation' from the data available in the IDI.
3. Comparing the two data sources at three levels.
    i. **Comparison of concepts and definitions.** We compare the concepts and definitions used in the census and the IDI, and identify differences. These comparisons are guided by the descriptions of each data source.
    ii. **Comparison of aggregate counts and estimates.** We compare census results for education and training information with their IDI equivalents.
    iii. **Comparison of individual level records**. We compare census results for education and training with their IDI equivalents at the individual level.

The concepts of coverage error and measurement error provide a framework for assessing the accuracy of data sources (Chiang et al, 2014, and Zhang, 2012).

Coverage specifies the population from which observations for a particular topic can be drawn. For census education variables, the population of interest is people over 15 years of age in the New Zealand resident population. The aggregate level comparisons are most useful in providing insight into differences in coverage.

Measurement errors occur when the response provided differs from the real value. Such errors may be random or they may result in a systematic bias if they are not random. Measurement error may occur when administrative definitions and concepts do not align well with the statistical concept being measured. Measurement errors in the census and administrative data may also be due to errors in collection and processing systems, and may result in missing or incorrect information. The individual level comparisons can inform our understanding of measurement error.

Close agreement of responses in administrative data and the census provides strong support for good measurement in both sources. However, when we get different responses, it is harder to determine which is more likely to be correct. This will depend on a range of factors and requires a deep understanding of the mechanisms underlying the particular administrative data collection, and of how people respond to the survey questions.

Being able to integrate information with other sources by linking the same units also affects accuracy. It can result in linkage errors, which are of two types:

- links may be missed, for example if the name of a person is recorded differently on different files
- two different people may be wrongly linked, for example if their names and dates of birth are very similar.

Linkage errors may reduce the coverage of an administrative source (no information is available if links are not made when they should be), or they may introduce measurement error if the wrong people are linked together.

# 4 Statistical concepts for education and training

Statistical standards and classifications (Statistics New Zealand, ongoing) provide definitions for the key concepts in this investigation. These statistical standards and classifications are designed for use in official statistics collections and are those used in the 2013 Census.

The key concepts are ‘qualification’ and ‘study participation’. Qualifications reflect officially recognised awards achieved over an individual’s lifetime until a reference date. An individual’s highest qualification may have been achieved years before the reference date. Study participation is about whether an individual was engaged in study on a reference date.

## 4.1 Qualification – Definition

The statistical standard for qualifications defines ‘qualification’ as:

A formally recognised award for educational or training attainment, where formal recognition means that the qualification is approved by one of the following (or their predecessors):

- New Zealand Qualifications Authority (NZQA)
- Universities New Zealand – Te Pōkai Tara
- Association of Polytechnics of New Zealand
- Association of Colleges of Education in New Zealand
- approval bodies that have been recognised by NZQA
- the recognised overseas authority of a secondary school, profession, academic discipline, or trade.

A qualification is defined as requiring full-time-equivalent study of three months or more. Qualifications gained overseas, or via the Internet, are included if they are awarded by a recognised authority, and the post-school qualifications meet the minimum three-month full-time study criteria. Primary school qualifications are not included.

‘Highest qualification’ represents the highest formal educational qualification a person has received. This may be no qualification, a secondary school qualification, or higher. In the census, highest qualification is derived for usual residents aged 15 years and over.

## 4.2 Qualification level – Classification

The New Zealand Register of Quality Assured Qualifications (NZ Register or NZREG) is the classification used to record the level of the qualification. It is a flat classification with 12 categories. One category is for ‘no qualification’, one for ‘overseas secondary school qualification’ and the other 10 categories represent Levels 1–10 in the New Zealand Qualifications Framework (NZQF) (see table 1). The NZQF is the definitive source for accurate information about all quality assured qualifications, covering senior secondary school and tertiary education qualifications, and including all qualifications open to international students.

**Table 1**
**Level descriptors and qualification types on the New Zealand Qualifications Framework**

| Level | Qualification types |
|---|---|
| 10 | Doctoral degree |
| 9 | Master's degree |
| 8 | Postgraduate diplomas and certificates, bachelor honours degree |
| 7 | Bachelor's degree, graduate diplomas and certificates |
| 6 | Diplomas |
| 5 | |
| 4 | Certificates |
| 3 | |
| 2 | |
| 1 | |
| **Source:** New Zealand Qualifications Framework | |

Highest qualification is determined by the highest level assigned from the NZQF. Levels 1 to 4 of the NZQF are used to classify both secondary school and post-school qualifications. However, certificates gained at Levels 1 to 3 are mainly achieved through the NCEA system. NCEA levels 1, 2, and 3 are usually studied by secondary school pupils in years 11, 12, and 13, and are equivalent to Levels 1, 2, and 3 in the framework.

Appendix 6 of the NZQF has a summary of qualification definitions for each level.

## 4.3 Study participation – Definition and classification

Study participation measures those attending, studying, or enrolled at school or anywhere else. It is grouped into full-time study (20 hours or more a week), part-time study (less than 20 hours a week), and those not studying.

# 5 Data sources

This section describes the two data sources compared in this investigation: the New Zealand Census of Population and Dwellings, and Statistics NZ's Integrated Data Infrastructure (IDI). Specifically, we describe the information about education and training in each source.

## 5.1 New Zealand Census of Population and Dwellings

The Census of Population and Dwellings is the official count of people and dwellings in New Zealand. It provides a snapshot of New Zealand at a point in time, and measures social and economic change. Census information has a wide range of uses within and outside government. Historically, the census has been held every five years, with some exceptions. The latest census was held in March 2013.

Find more information about the census.

While the census aims to count everyone who is in New Zealand on census night, for this investigation we are only interested in New Zealand residents aged 15 and over, as education and training information is only collected for this population.

### 5.1.1 Education and training information in the census

The Statistical Standard definitions for the education and training information currently produced by the census are described in section 4. The education and training questions asked in the 2013 Census are available in appendix 1.

The education and training information collected by the census has some quality limitations. In addition to non-response, ambiguity in free-text responses sometimes means a specific qualification level cannot be assigned. The census highest qualification question was a write-in response. Some generic responses like 'diploma' or 'certificate' could not be coded easily. As it was difficult to assign a level to responses such as diploma, which could either be a Level 5 or Level 6 qualification, these levels were combined in 2013 Census outputs.

Non-response and responses that could not be classified or did not provide the type of information asked for are collectively called 'Not elsewhere included'. Table 2 presents the quality information for the education and training variables in the 2013 Census (Statistics New Zealand, 2013a and 2013b). A non-response rate over 10 percent is considered relatively high. Substitute records (records representing individuals for which no response whatsoever was received) contribute 6 percent of the non-response rate.

**Table 2**
**Quality information for the education and training variables, 2013 Census**

| Variable | Quality rating | Non-response rate (%) | Not elsewhere included rate (%) |
|---|---|---|---|
| Highest qualification | Moderate | 7.2 | 11.1 |
| Study participation | High | 10.4 | 10.4 |
| Source: Statistics New Zealand | | | |

## 5.2 The Integrated Data Infrastructure (IDI)

Statistics NZ developed the IDI as an environment in which to link multiple data sources systematically and securely. It was developed to produce official statistics outputs and to allow Statistics NZ staff and external researchers to conduct policy evaluation and research on people's transitions and outcomes. The IDI contains administrative and survey datasets, linked at the individual level. Here we have used it as a test environment for examining the potential of linked administrative data sources for producing education and training information.

The IDI continues to change as new datasets are added.
Read the latest information about the Integrated Data Infrastructure.

This section describes the structure and content of the IDI as at May 2015.

The structure of the IDI (figure 1) can be described as a central 'spine' to which a series of data collections are linked. The spine forms the conceptual centre of the IDI and all other datasets are linked to it. Broadly, the target population for the spine is all individuals who have ever been residents of New Zealand.

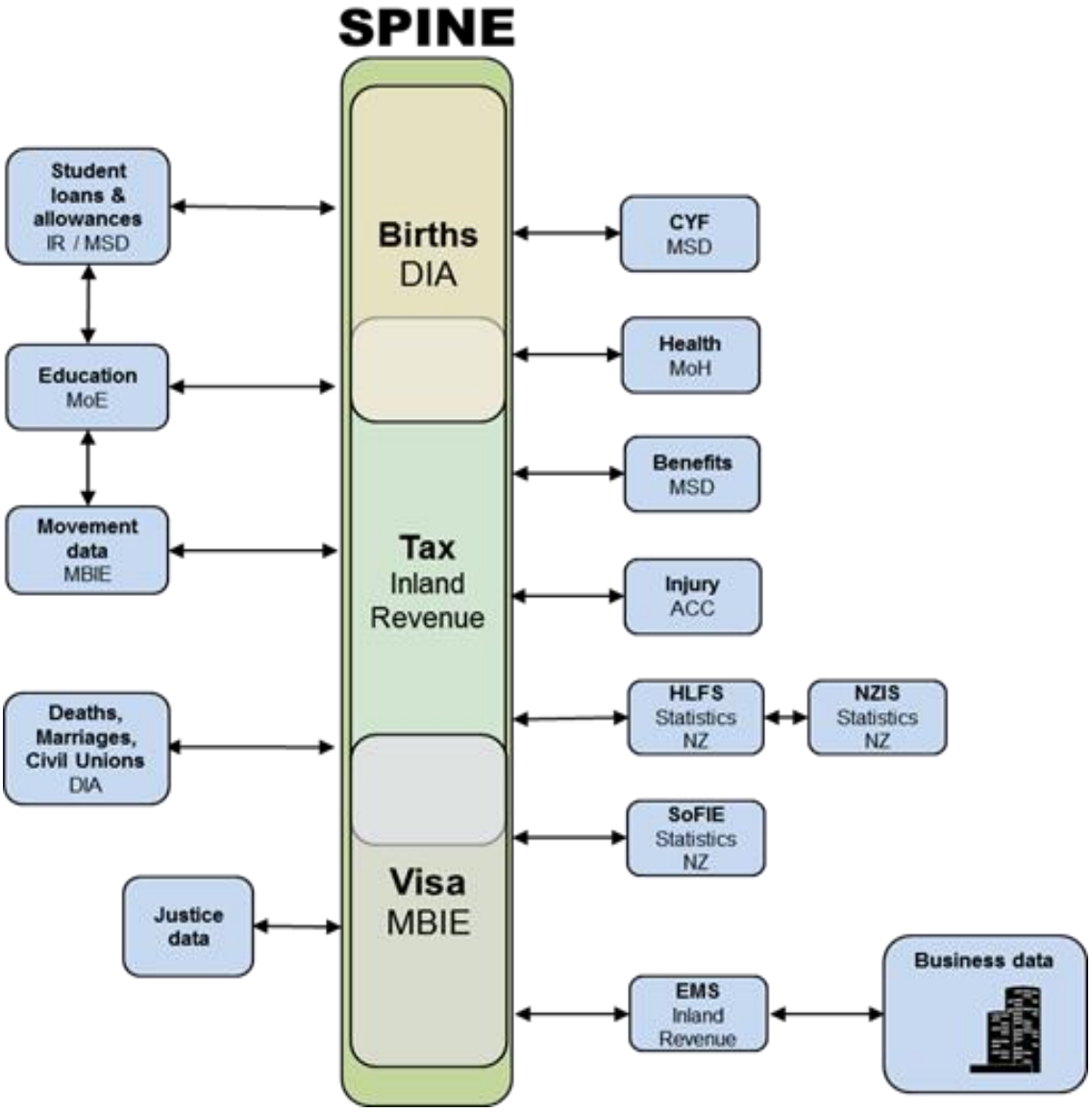
Three data sources are linked together probabilistically to create the spine:

- a list of all IRD numbers that have been issued by Inland Revenue (IR)
- a list of all births registered in New Zealand since 1920
- a list of all visas granted to migrants from 1997 (excluding visitor and transit visas).

Other datasets are linked to the IDI spine (see Statistics NZ, 2014b, for a description of the linking process). The linked datasets cover a wide range of subject areas and include:

- employer and employee job and earnings information based on IR tax data
- health information including GP enrolment and hospital visits from the Ministry of Health
- education data from the Ministry of Education
- benefit dynamics data from the Ministry of Social Development
- student loans and allowances data from several sources
- migration movements data from the Ministry of Business, Innovation and Employment
- the Household Labour Force Survey and New Zealand Income Survey data from Statistics NZ.

**Figure 1**
Structure of the Integrated Data Infrastructure in May 2015




Source: Statistics New Zealand

The IDI also contains several summary tables that provide core information about individuals (age, sex, ethnicity, and geographic information) summarised from across the available data sources.

From relevant variables within the data sources linked in the IDI, values for education and training variables can be derived. In this investigation, we used only MoE data to derive these values. The following sub-sections describe the relevant information used in this derivation process: first, the populations selected for comparison, and second, the data sources and variables selected to derive values for highest qualification and study participation.

## 5.2.1 IDI populations used in analysis

To obtain the most relevant comparison between the census and IDI information, we used different populations of IDI records for different parts of the analysis in this investigation.

### The IDI Usually Resident Population (IDI-URP)

The aggregate-level analyses in this investigation uses only individuals aged 15 and over within the IDI identified as usually resident, and present, in New Zealand as at 5 March 2013 (census day). That is, the definition we used for the administrative sources is the same as the census usually resident population count for those aged 15 and over.

These individuals form an experimental population, referred to here as IDI-URP, which aims to reflect as closely as possible those who would also be included in the definition of the census usually resident population count for the purposes of valid comparisons. The IDI-URP is derived from the spine, and using information from tax, health, ACC, education, migration, and deaths data in the IDI. More detail about the selection of a resident population from the IDI is available in appendix 2 and Gibb (in press).

Given study participation is one of the criteria for inclusion in the IDI-URP, the coverage of this group should align well with study participation measured in this investigation.

### The IDI-Census linked population

The 2013 Census was linked to the IDI spine as at May 2015 for the purposes of understanding possible future census models (Statistics New Zealand, 2014c). This linked census data was used by our Census Transformation team only and not available to other IDI researchers at this time. Records were linked probabilistically using name, date of birth, sex, usual residence, and country of birth. The individual-level analyses in this investigation use this linked IDI-Census dataset.

For this analysis, the linked population is limited to usual residents aged 15 and over (according to the census), as only these individuals were asked about their education and training in the census. This population includes all census records for usual residents aged 15 and over for which a suitable link in the IDI has been found (an IDI link was found for 93 percent of the 3,376,400 census usual residents aged 15 and over).

In some cases, a link may be made between two different individuals. This is estimated to have occurred for less than 1 percent of the links made between census and IDI. This means that linkage error could explain a small proportion of cases where education and training information is found to be different between the census and the IDI.

Furthermore, there were around 13,400 individuals in the census (0.4 percent of usual residents aged 15 and over) who indicated they were currently studying but for whom a link could not be found in the IDI.

## 5.2.2 Education and training information in the IDI – Ministry of Education data

Ministry of Education is the most extensive source of education and training information in the IDI. This source is described here, and used to derive values for both highest qualification and study participation from the IDI which are later compared to the census. Tables 7 and 8 in appendix 3 provide a summary of the information available by IDI table.

MoE provide data to the IDI from tertiary education, primary and secondary education, and industry and targeted training institutions. While there are some limitations with the historical extent of the sources, overall, MoE data provides very high coverage of educational institutions.

Individual students are identified by the National Student Number used across the education sector. This means that information about the same individual can be accurately combined from different education sources.

### Tertiary education

MoE collects information relating to tertiary education from tertiary education providers. All providers, both public and private, that receive the Equivalent Full-time Student (EFTS) based tuition subsidy or have students with Student Loans or Allowances are required to provide MoE with information on student characteristics, course and qualification enrolment details, qualification completions, course details, and student numbers (EFTS). Data are supplied by tertiary education providers three times a year (as at 30 April, 31 August, and 31 December).

This information is available in the IDI through three event-based tables: qualification enrolments, qualification completions, and course enrolments. Characteristics of the individual enrolling, or completing, are available (eg ethnicity) as well as information about the course or qualification they are enrolling in or completing (eg level and field of study). Information on short courses is included in the course enrolments data supply. Short courses include short formal qualifications less than 0.03 EFTS or less than one week duration, non-formal qualifications, and Secondary-Tertiary Alignment Resource (STAR) qualifications.

Information is only available for tertiary education institutions within New Zealand. Enrolment and completion information is available from January 1994 to December 2013 in the IDI as at May 2015.

### Primary and secondary education

MoE collates information relating to primary and secondary education. The relevant sources here are the school roll returns and the New Zealand Qualifications Authority.

#### *School roll returns*

MoE carries out annual statistical collections (roll returns) from all schools in New Zealand. The data provided through these collections are used to fund and staff schools and to support policy analysis, development, and decision-making, among other things. Schools are required to submit information about student enrolment as at 1 July each year.

Find more information about July school roll returns.

Information in the IDI from the school roll returns includes, for each enrolment since January 2007 to July 2014, start and end dates. Information on the nature of attendance, for example whether a student is full-time or part-time is available on the rolls, but is not currently available in the IDI.

#### *New Zealand Qualifications Authority (NZQA)*

One of the roles of NZQA is to administer the secondary school assessment system. NZQA provides qualification data for all individual student qualifications to MoE (via a centralised operational data store). This happens twice yearly, in February and June. This provides information on the level of qualification individuals have achieved, as well as date of achievement for all students enrolled in a New Zealand school from January 2007 until May 2014 who have completed a qualification.

### Industry and targeted training

Industry training enables individuals to gain qualifications while earning, while targeted training refers to specific government-funded training programmes that help individuals gain qualifications.

For industry training, data is collected and compiled by Industry Training Organisations (ITOs), and ultimately warehoused by MoE. ITOs cover most of New Zealand's industries: from traditional trades – like building and plumbing, the primary industries, and manufacturing and retail – through to government and community services.

In the IDI, information is available for individuals who have been or are in workplace-based training eligible for funding through the industry training and Modern Apprenticeships funds between 1 January 2003 and 31 December 2013. Information about the level and field of study of the qualification enrolled in through the industry training is available, as well as start and end dates for the period of training, and whether the training, and the qualification, was completed or not.

For targeted training, data is collected by the training providers initially by the Tertiary Education Commission, and then compiled by MoE. Information is available for individuals who have been, or are, placed in specific targeted training programmes funded by government through the Tertiary Education Commission. Information is available about the level of qualification attained through the targeted training, as well as start and end dates of the training. Some information on historical education is also available. Information is available for all enrolments from January 2001 until December 2013.

### 5.2.3 Education and training information in the IDI – other sources

Ministry of Social Development and Ministry of Business, Innovation and Employment provide, or have the potential to provide, further information about education and training to the IDI for particular population groups. These sources have not been used to derive values for education and training variables from the IDI at this stage.

Details for these sources are included in tables 7 and 8 in appendix 3.

### 5.2.4 Derivation of education and training information in the IDI

This section describes how we constructed values for highest qualification and study participation using the MoE data available in the IDI (described above).

#### Highest qualification

Highest qualification is based on qualifications completed, or on information about previous qualifications collected on enrolment in a new course, qualification, or institution. Enrolment in a qualification has not been used as indication of completion. The following steps describe how highest qualification was derived:

1. We recoded all information about level of qualification to represent the statistical standard classification of qualification levels as closely as possible. That is, we converted qualification level information in the IDI to the relevant NZQF level. As it was difficult to distinguish between Level 5 and 6 qualifications, we combined these levels.
2. We brought together information from multiple sources about all qualifications completed on or before 5 March 2013 for each individual. Where a specific completion date was not available, we included qualifications achieved in the year before census, but not those achieved in the same year as census. ‘Qualifications completed’ includes both qualifications where a formal record is available, and where one has been reported when an individual enrols in a new course, qualification, and educational institution.
3. We identified the highest qualification from all known qualifications attained by each individual, using the NZQF level indicators (eg a Level 9 qualification was identified as higher than a Level 7).

### Study participation

To derive study participation, we used information about start and end dates of study periods. This includes start and end dates of tertiary enrolments, secondary enrolments and industry and targeted training participation.

The following steps describe how we derived study participation:

1. We gave a study participation indicator to all individuals with a study period spanning 5 March 2013 (that is with a start date of before or on 5 March 2013, and if the period of study has been completed, an end date of on or after 5 March 2013) regardless of the length of study. This was based on enrolment information from all sources.
2. We assumed all those without a study participation according to step 1 to be not studying as at 5 March 2013.
3. For all individuals with a study participation indicator, we maintained any information about the full-time or part-time nature of their study from the relevant source with the individual record. We assumed all individuals enrolled in secondary school to be full-time.

# 6 Results

Results, which reflect the IDI as at May 2015, are presented in three sub-sections in line with the method:

1. comparisons of concepts and definitions between education and training information from the census and the IDI
2. comparisons of aggregate counts and estimates between the census and the IDI
3. comparisons of individual level records.

Results for 'highest qualification' are presented first, followed by those for 'study participation'.

## 6.1 Results for highest qualification

### 6.1.1 Comparison of concepts and definitions for highest qualification

The concept of qualification that both the IDI and census attempt to capture is the same as the statistical standard – that is, the achievement of a formally recognised award. However, differences in the values of highest qualification in the census and the IDI can occur because of measurement error.

- In the census we rely on the respondent's interpretation of their highest qualification, and their correct identification of its level, whereas in most cases the IDI provides the formal recognition of a qualification from a registered record of a qualification completed through NZQA. A small amount of qualification information in the IDI is from details provided by the individual about their past qualification achievements; this information relies on respondent interpretation as in the census and will only reflect an individual's achievements at the time they provided the information. Individuals providing education history information at tertiary level enrolment may not yet have officially received their school level qualifications.
- There are likely to be some differences in qualification level classifications between the IDI and census. While this investigation attempts to re-code all IDI information to the standard classifications, this has not been possible in all cases. For example, not all qualification information in the IDI uses a classification as detailed as the standard so 'National Diploma/National Certificate Levels 5–7' has been coded to Level 6 in the standard. Similarly census encountered difficulties coding Level 5 and Level 6.

Comparison of individual responses will shed light on whether these differences result in random variation or whether there are systemic differences between the two sources.

### 6.1.2 Comparison of aggregate counts and estimates for highest qualification

We have derived information about individual qualifications from MoE sources in the IDI. This section presents the results of comparisons between the IDI estimates and the census counts for highest qualification and reveals differences in coverage by age group. Sources of coverage error in the MoE data available in the IDI include:

- The administrative sources investigated only provide information about presence of qualifications. They do not provide information for individuals with no qualifications. In contrast, the census provides a respondent declaration of no qualifications, although there is some non-response.
- In the IDI, almost all information about qualifications is from individual interactions with the New Zealand education system. Very little information is available for those who achieved their highest qualification outside of New Zealand.

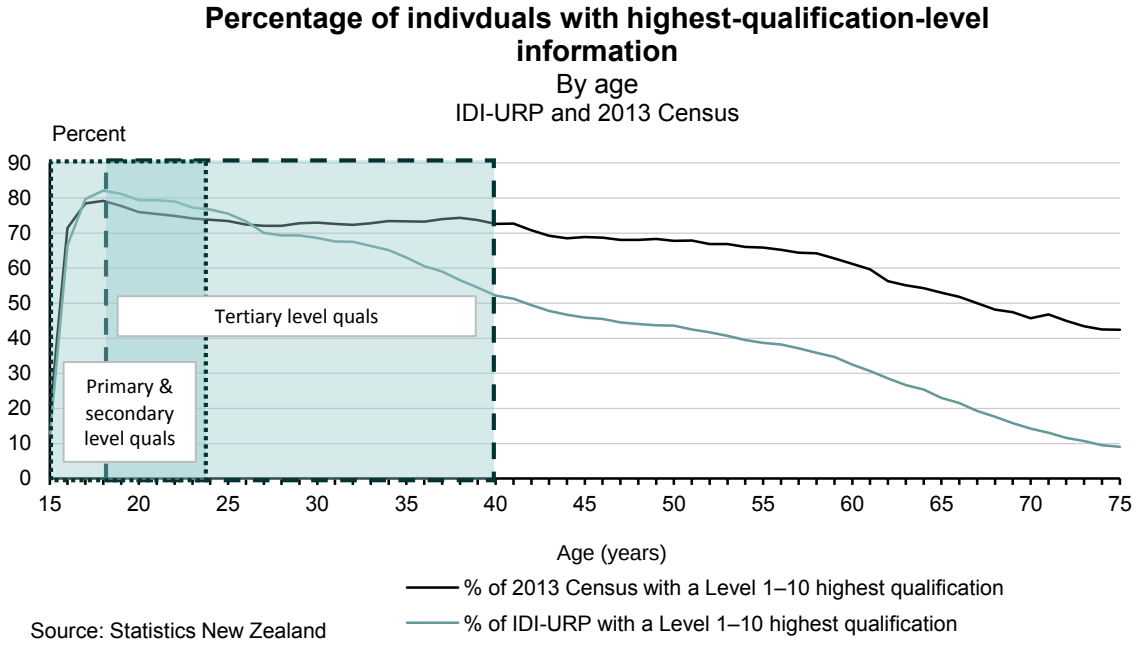
- The first year for which completions or enrolments data are available in the IDI is 1994; for many datasets it is more recent. There is no information available on individuals who achieved their highest qualification before this date, and have not enrolled in other study since.
- For some sources of qualification information, specific date of completion is not available in the IDI. For qualifications where only a year of achievement is available, some qualifications which were achieved between 1 January 2013 and Census Day 2013 will not be included.

A highest qualification between NZQF Levels 1 and 10 was derived for 47 percent of the IDI-URP aged 15 and over (of 3,519,100 aged 15 and over in the IDI-URP, 1,658,400 individuals had a highest qualification derived). This compares with 64 percent of the census usually resident population aged 15 and over who had a highest qualification at Levels 1 to 10 available.

In the IDI-URP, 233,900 individuals (6.6 percent) were identified as having no qualifications. This originates from information provided by students about their previous qualifications at enrolment for a new course or qualification. This compares with 628,400 individuals (18.6 percent) with no qualifications in the 2013 Census. The IDI estimates will be missing substantial numbers of individuals with no qualifications, because the information about highest qualification has been derived from data sources focused on collecting information relating to qualification completion and enrolments. Those who have no qualifications are less likely to have interacted with these data sources. Analysis in this section therefore focuses on those with qualifications at Levels 1 to 10. That is, it includes comparisons only of those individuals who have a highest qualification available in the IDI.

Availability of a highest qualification level varies by age (figure 2), broadly in line with the number of years that the source information is available for. For primary and secondary level qualifications, records begin in 2007. Given most qualifications at this level are achieved in school, most records are expected to relate to those aged 24 and under (dotted outline area). Similarly, tertiary level qualification completion information records begin in 1994. Those undertaking tertiary education directly after finishing school are expected to be aged 40 and under (dashed outline area).

**Figure 2**




Highest qualification in the IDI is derived based on information from several sources. Different sources contribute information about highest qualification for different groups (age groups and qualification level). While the same information about an individual's highest qualification can originate from more than one source, in most cases (80 percent) only one source contributes.

The contribution of each source varies by age. As expected, primary and secondary education sources mostly provide highest qualification information for those at younger ages (between 15 and 19). This is partly because of the length of time information has been collected for (since 2007) and partly because older people have had more years to achieve a qualification at a higher level. Most information about highest qualification comes from tertiary education for those aged 20 and above. Industry and targeted training sources only provide highest qualification level for a small number of people, with most information available for those in their twenties and thirties.

As expected, primary and secondary education sources mostly provide information for those with a highest qualification at Levels 1 to 4 and all information about qualifications at Levels 8 to 10 is sourced solely from tertiary education data.

Highest qualification estimates from the IDI are all based on information sources about qualifications gained in New Zealand, or in the case of information provided by the individual, there is a bias given all information is gathered as a result of an individual interacting with the New Zealand education system. Individuals who gained qualifications overseas cannot currently be identified in the IDI or the census, therefore the effect of overseas qualifications on these comparisons is unknown. However, given the expected availability of tertiary-level qualification completion data for those up to around age 40 who have completed their qualifications in New Zealand, it is likely that the difference between the census information and the IDI information for those aged between 26 and 40 is largely a result of qualifications gained overseas.

## 6.1.3 Comparison of individual level records for highest qualification

For the following analysis we have used a linked dataset, as described above, in which census records have where possible been linked to IDI records. We compare highest qualification information from the census and the IDI at the individual level.

There are 3,125,100 records for usually resident individuals aged 15 and over with a link between IDI and census. Of these, 1,279,000 (41 percent) have information about highest qualification at Levels 1 to 10 in both sources and of these 725,300 (57 percent) are the same in both census and IDI.

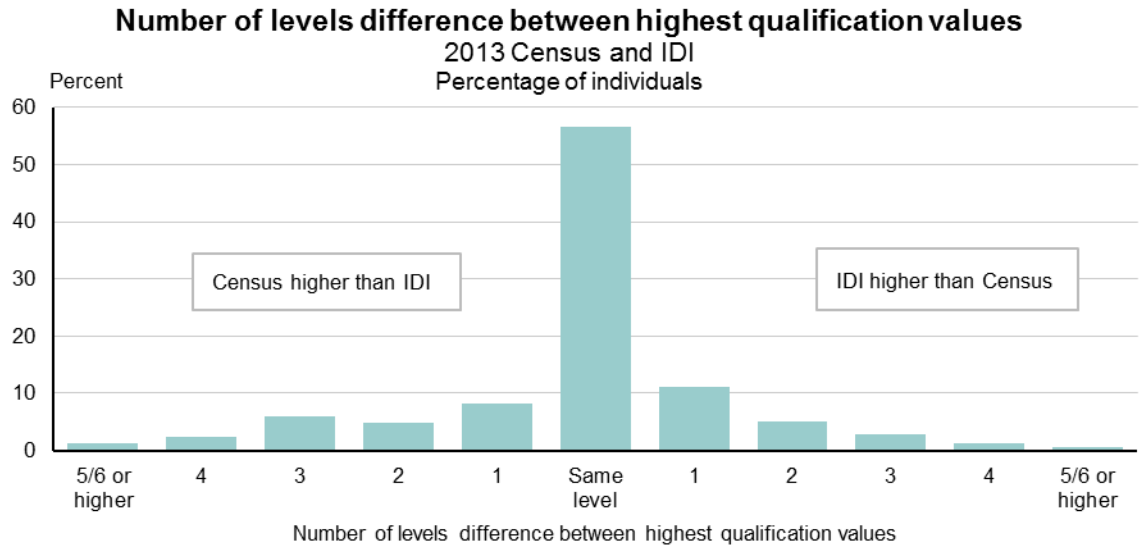
At almost all levels the largest proportion of individuals have the same level in both IDI and census. Table 3 shows the percentage of individuals at each level in the census with the same highest qualification level in the IDI. When grouped, the similarities between census and IDI reach 82 percent for secondary-school-level qualifications (Levels 1–3), and 76 percent for tertiary-level qualifications (Levels 7–10) but remain relatively low (51 percent) for the post-school, but non-tertiary-level qualifications (Levels 4–6). This may partly reflect the difficulties coding certificates and diplomas at these levels.

**Table 3**
**Percentage of individuals at each level in the census with the same highest qualification level in the IDI**

| Qualification group | Highest qualification level – 2013 Census | % with same level in IDI | % with same level in IDI (grouped) |
|---|---|---|---|
| Secondary school level qualifications | Level 1 certificate | 54 | 82 |
| | Level 2 certificate | 64 | |
| | Level 3 certificate | 64 | |
| Post-school level qualifications | Level 4 certificate | 38 | 51 |
| | Level 5/6 diploma | 42 | |
| Tertiary level qualifications | Bachelor's degree and Level 7 qualification | 67 | 76 |
| | Post-graduate and honours degrees | 56 | |
| | Master's degree | 57 | |
| | Doctorate degree | 55 | |
| Source: Statistics New Zealand | | | |

Figure 3 shows the percentage of individuals by the number of levels difference between their census and IDI qualification levels; 76 percent of the IDI-Census linked population have a highest qualification from both sources within one level of each other. A full matrix of highest qualification levels by source is available in appendix 4.

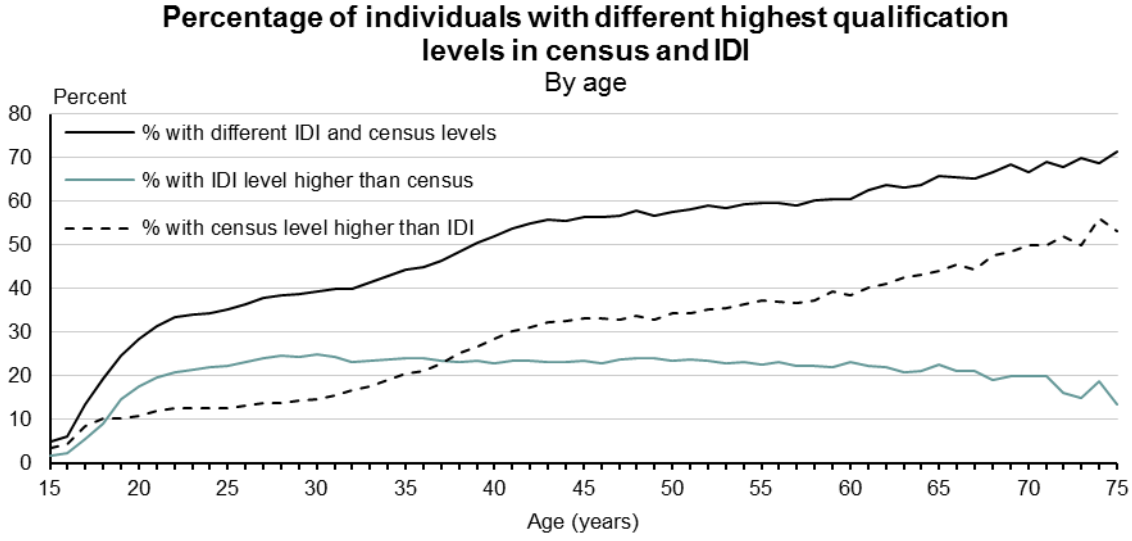
**Figure 3**




Source: Statistics New Zealand

The proportion of individuals with different levels of highest qualification in the census and the IDI increases by age (figure 4). A larger proportion of those aged 38 and over report a higher level qualification in the census than in the IDI than vice versa. At younger ages, for those educated in New Zealand, we expect an individual's highest qualification to be recorded in the MoE data. At older ages, only qualifications achieved since records began will be recorded in the MoE data. Such qualifications, while more recent, will be outside the normal school to university pathway and may not reflect an individual's highest achievement.

**Figure 4**




Source: Statistics New Zealand

Where information about highest qualification is missing or uninterpretable in the census (202,900 individuals), for most individuals (73 percent) no information about highest qualification level is available in the IDI, or information suggests no qualifications (table 4).

**Table 4**
**Distribution of highest qualification level in IDI for individuals missing highest qualification level in census**

| Highest qualification level – IDI | Census not elsewhere included | % |
|---|---|---|
| School level qualifications (levels 1–3) | 33,700 | 17 |
| Post-school level qualifications (levels 4–6) | 15,700 | 8 |
| Tertiary level qualifications (levels 7–10) | 5,200 | 3 |
| No qualification / Missing / Unknown | 148,300 | 73 |
| **Total** | **202,900** | **100** |
| Source: Statistics New Zealand | | |

# 6.2 Results for study participation

## 6.2.1 Comparison of concepts and definitions for study participation

The census attempts to capture the statistical standard concept of study participation, that is, whether an individual is attending, studying, or enrolled at school or anywhere else. In the IDI, the concept of study participation captures individuals enrolled in a qualification or recognised educational institution, but does not include informal study. Both sources use the concept of 'enrolment' as part of their measure of study participation. Differences between study participation recorded in the census and those derived from the IDI can occur because:

- In the census we rely on the respondent's interpretation of their study participation, whereas IDI provides study participation based on an administrative record of enrolment in a qualification, a course, or at school.
- The census question does not distinguish between formal or informal study. An individual may be 'attending' or 'studying' without a formal enrolment and these cases cannot be identified in the IDI. Some census respondents may identify as study participants when attending, for example, an informal community education evening class, while others may not have considered this when responding.

Comparison of individual responses will shed light on whether these differences result in random variation or whether there are systemic differences between the two sources.

## 6.2.2 Comparison of aggregate counts and estimates for study participation

We have derived information about individual study participation from MoE sources in the IDI. This section presents the results of comparisons between the IDI estimates of study participation and the census counts. It reveals differences in coverage by age group. Sources of coverage error in the MoE data available in the IDI include:

- Administrative sources only provide positive identification of study participation. Coverage of formal study in educational institutions is high, so it's likely that individuals not positively identified as studying are not studying. In contrast, the census provides a respondent declaration of non-participation, although there is some non-response.
- In the IDI, all information about study participation is from individual interactions with the New Zealand education system. Very little information is available for those

who are usually resident in New Zealand but are enrolled in a course of study overseas.

- There may be some very small differences due to home-schooling. It is unclear how home-schooled children are recorded in MoE and census data. However, there are an estimated 2,500 individuals home-schooled at the secondary level (Schools Analysis Unit, 2012), of which the proportions of 15- and 16-year-olds will only be small.

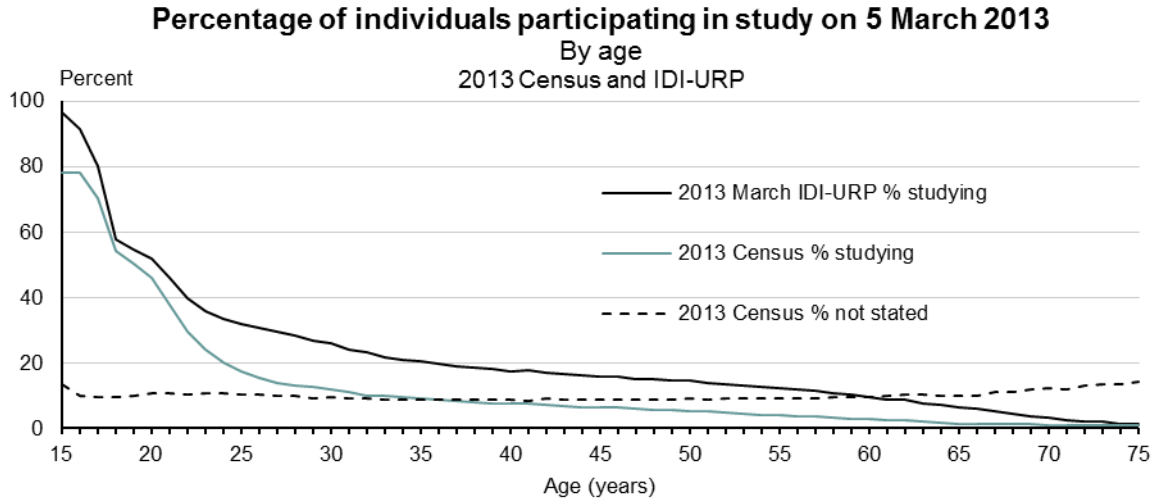
Error may also occur due to timeliness of availability of data sources. The ability to determine study participation at a particular reference date is dependent on availability of enrolment data spanning the reference date. Information about tertiary qualification enrolments to the end of the previous calendar year is currently supplied by MoE to the IDI annually in the following September. In the case of the 5 March 2013 Census date, enrolment information only became available in the IDI after September 2014.

Overall we expect participation in informal or overseas-based study to be quite small, and that the MoE data on formal study participation would have very high coverage of New Zealand residents actively participating in study. Comparison of aggregate distributions of study participation between IDI-URP and the census identify coverage differences.

A positive indication of study participation on census day was derived for 22 percent of the IDI-URP aged 15 and over (of 3,519,100 aged 15 and over in the IDI-URP, 777,000 were identified as studying). This compares with 15 percent of the 2013 Census usually resident population aged 15 and over.

Study participation varies by age (figure 5). At all ages, the IDI-URP estimates of study participation are higher than the census counts indicate. This may be partly explained by the census non-response to the study participation question – which is evident across all ages. The IDI estimates for ages 15 and 16 (97 percent and 91 percent, respectively) are more in line with expectations than the census estimates (78 percent for both 15- and 16-year-olds), given school is compulsory from ages 6 to 16.

**Figure 5**

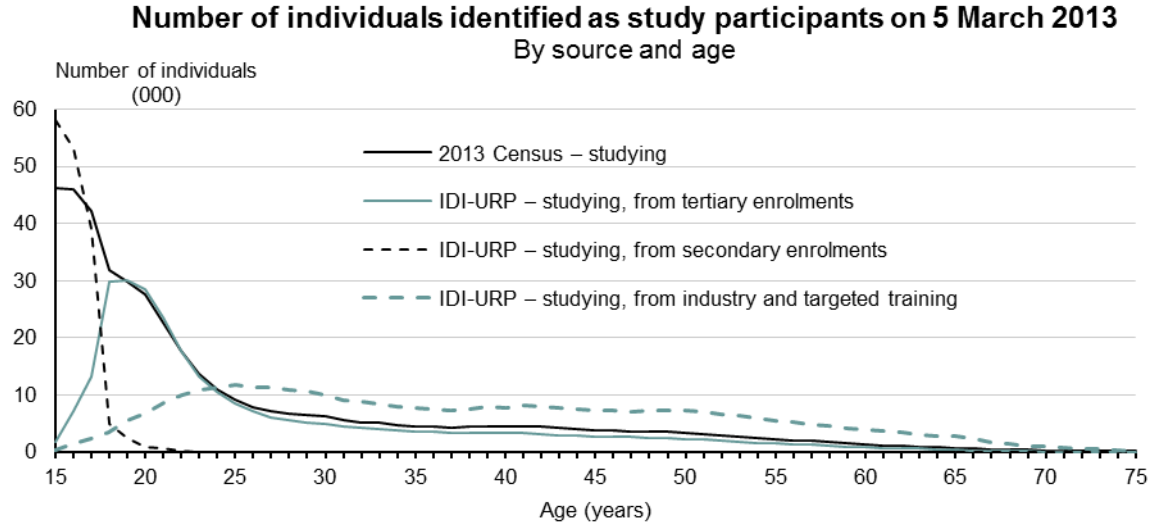



Source: Statistics New Zealand

While the census does not distinguish the type of study, the MoE estimates of study participation are derived from secondary, tertiary, and industry/targeted training data. Figure 6 illustrates which sources provide information for which age groups. As expected, study participation information for those aged 15 and 16 years comes mostly from secondary school enrolment data. Information for those aged 17 to 19 years comes from a combination of sources. Tertiary enrolments compare well with the census for those aged 20 and over. The higher number of industry and targeted training enrolments than

census study participants suggests census respondents do not consider this type of 'study' to be 'study'.

**Figure 6**




Source: Statistics New Zealand

Information about whether the study was full-time or part-time was available for 57 percent (439,300) of individuals identified as participating in study according to the IDI. Table 5 shows how the distribution of full- and part-time study compares between census and IDI-URP.

**Table 5**
**Percentage of individuals studying full-time, part-time and both, 2013 Census and IDI-URP**

| Type of study | 2013 Census | | March 2013 IDI-URP | |
|---|---|---|---|---|
| | **Number** | **%** | **Number** | **%** |
| Full-time study | 331,900 | 74 | 342,700 | 78 |
| Part-time study | 118,100 | 26 | 85,100 | 19 |
| Full-time and part-time study | 500 | 0 | 11,500 | 3 |
| **Total** | **450,500** | **100** | **439,300** | **100** |
| Source: Statistics New Zealand | | | | |

A comparison at the individual level provides further information on the differences and similarities between the two sources.

## 6.2.3 Comparison of individual level records for study participation

For the following analysis we have used a linked dataset, described above, in which census records have where possible been linked to IDI records. Study participation from the census and the IDI is compared at the individual level. There are 3,125,100 records for usually resident individuals aged 15 and over with a link between IDI and census. Of these, 94 percent have information about study participation for both records (those with no evidence for study in the IDI are assumed to be not studying).

Table 6 provides the number and percentage of individuals participating in study according to each source. Most individuals are not studying: 78 percent according to the IDI, and 80 percent according to the census, while 82 percent have the same study participation indicator in both the IDI and census. Of these 11 percent are study participants, and 71 percent are not.

The two sources disagree in some cases. Three percent are studying in census, but not in IDI. Some of these individuals may be participating in informal study where no enrolment is required. A larger group (9 percent) are studying in IDI, but not in census. Analysis of study participation by source suggests information sourced from industry training information does not align with census information. Of the 296,200 individuals in the linked IDI-Census dataset who were studying according to the IDI, but not according to the census, 256,000 (86 percent) were solely identified as study participants by their participation in industry training confirming indications from the aggregate comparisons (figure 6).

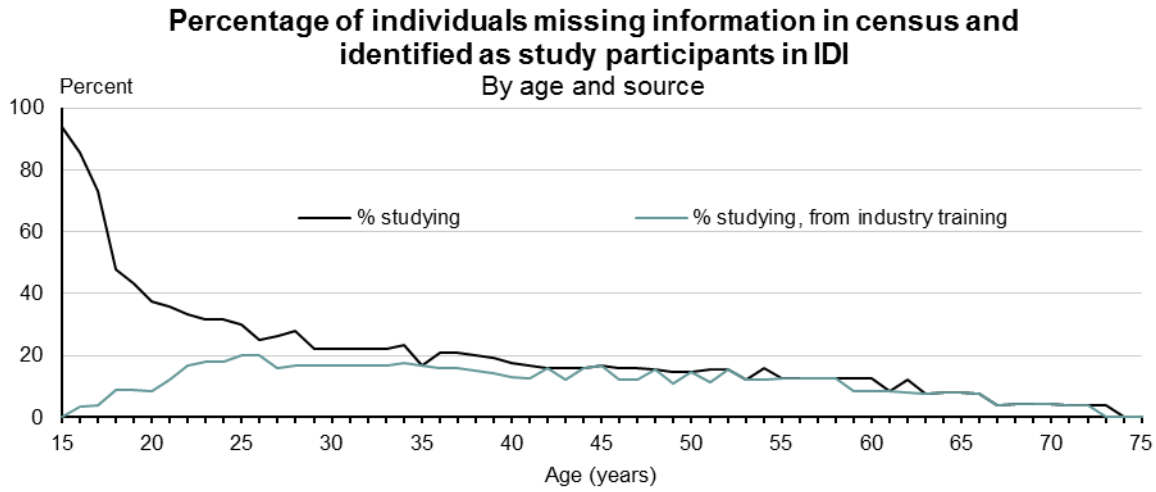
**Table 6**
**Study participation indicators for census and IDI, number and percentage**

| | | Census | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Studying | | Not studying | | Missing | | Total | |
| | | Number | % | Number | % | Number | % | Number | % |
| IDI | Studying | 352,800 | 11 | 296,200 | 9 | 31,300 | 1 | 680,400 | 22 |
| | Not studying | 84,200 | 3 | 2,212,100 | 71 | 148,400 | 5 | 2,444,700 | 78 |
| **Total** | | **437,000** | **14** | **2,508,300** | **80** | **179,700** | **6** | **3,125,100** | **100** |

Source: Statistics New Zealand

A census study participation response is missing for 179,700 individuals for which an IDI-Census link has been found (note that individuals for which no information was provided from census (substitute records) are less likely to have been linked to the IDI in the first place). While overall, 17 percent (31,300) of these individuals were identified as study participants in the IDI, the highest rates of study participation in the IDI among non-respondents in the census are for those aged 15 to 17 (figure 7). Furthermore, almost all of the census non-respondents aged around 25 and over identified as study participants were identified through their participation in industry training.

**Figure 7**




Source: Statistics New Zealand

Of the 352,800 individuals who are identified as study participants in both the census and the IDI, 82 percent have the same full-time/part-time status in both sources. A full matrix is available in appendix 4.

# 7 Discussion and conclusion

This paper presents our investigation into the potential for deriving census education and training information from administrative data currently available in the IDI. We compared 2013 Census data for the variables highest qualification and study participation with similar information derived from Ministry of Education data available in the IDI in May 2015.

Overall, the administrative sources investigated show good potential for providing education and training information for those who have had some interaction with the New Zealand education system, and particularly for those with recent interactions. For some groups, information in the IDI may provide higher quality information than the census, while for others it did not. Given the policy interest for education and training information tends to be focused on younger ages – eg youth not in employment, education, or training (NEET) – administrative sources show promise.

The qualification and study participation information recorded in the administrative sources can be taken as a formal record with minimal measurement error. The accuracy of the information does not rely on accurate respondent recall as it does in the census. There are good reasons for the Ministry of Education records to have good coverage of formal qualifications and study within New Zealand, as funding for educational institutions is linked to student participation. The enrolments and qualifications systems are well established. However, due to the relatively recent time periods for which these administrative sources are available, coverage tends to be higher for younger people, and those currently in secondary or tertiary education. Furthermore, there is little to no information about qualifications obtained overseas, and it is not possible to positively identify those with no qualifications or those who are not studying.

While the census may be more likely to be affected by respondent error or processing error, it is still the only source that provides high coverage of all qualifications for the whole population. However, administrative data on study participation does provide high coverage, and more detail about the nature of the study than census currently does.

For over half (57 percent) of the individuals compared, highest qualification level was the same in both IDI and census. For 76 percent, it was within one level above or below. Consistency between census and IDI information about highest qualification was highest for those with school-level (Levels 1–3) and tertiary (Levels 7–10) qualifications, but less good for post-school qualifications (Levels 4–6).

For study participation, similarities between individuals in the IDI and census are stronger: 82 percent had the same study participation indicator in both sources. Industry training information appears to adversely influence comparisons between census and IDI – 86 percent of those studying based on IDI, but not studying based on census, were solely identified as study participants by their participation in work-based (industry) training.

The consistency of information is affected by coverage error and measurement error in both sources, as well as by linkage error both between data sources in the IDI, and between the census and the IDI. It is not possible to say which source is ‘correct’ where they differ for individual records. However, the broad patterns identified in this paper add to our understanding of the strengths and limitations of both sources. Furthermore, the comparisons made with the IDI-URP are influenced by the accuracy of the coverage of the IDI-URP itself. While it is not possible to quantify their impact, it is important to note the presence of these underlying factors.

While the scope of this investigation was limited to level of study and a breakdown by age, it provides a basis for further and more detailed analysis. Opportunities for further

analysis include investigating field of study, as well as a breakdown by geography and variables such as ethnicity.

# Areas where potential for administrative sources could be improved

This investigation has identified two key areas in which the potential for administrative sources to estimate census education and training attributes could be improved:

1. clarifying information needs
2. increasing the usability of administrative sources for statistical re-use.

## Clarifying information needs

While our investigation focused on replicating the information currently produced by census, understanding the overall information needs is important. This investigation has highlighted several areas that require further clarification. For example:

- Whether information about informal study participation is required. While formal study participation is well recorded in administrative data, this is not the case for informal study.
- Whether industry training needs to be captured as part of study participation measurement.

## Increasing the usability of administrative sources for statistical re-use

It would be useful to explore ways of increasing the usability of education data sources. Some examples where the nature of the current collection limits usability that we have encountered during this investigation include:

- coding errors resulting from inconsistent use of NZQF Levels 1–10, especially Levels 5 and 6
- year, rather than exact achievement date, for some qualification information.

Implementing changes such as these would better align the administrative data available for statistical use.

# References


Chiang G, Reid, G, and Zabala, F (2014). *Guide to reporting the quality of administrative data*. Unpublished methodology report. Statistical Methods, Statistics NZ, Wellington.

Gibb, SJ (in press). Quality of geographic information in the Integrated Data Infrastructure. To be available from www.stats.govt.nz

O'Byrne, E, Bycroft, C, & Gibb, S (2014). An initial investigation into the potential for administrative data to provide census long-form information: Census Transformation programme. Available from www.stats.govt.nz

Schools Analysis Unit (2012) National School Roll Projections: 2011 update. Available from www.educationcounts.govt.nz

Statistics New Zealand (Ongoing). Classifications and related standards. Available from www.stats.govt.nz

Statistics New Zealand (2013a). 2013 Census Information by variable: Study participation. Available from www.stats.govt.nz

Statistics New Zealand (2013b). 2013 Census Information by variable: Qualification. Available from www.stats.govt.nz

Statistics NZ (2014a). An overview of progress on the potential use of administrative data for census information in New Zealand. Available from www.stats.govt.nz

Statistics New Zealand (2014b). Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project. Available from www.stats.govt.nz

Statistics New Zealand (2014c). Privacy impact assessment for linking census and Post-enumeration Survey data to the Integrated Data Infrastructure. Available from www.stats.govt.nz

Zhang, LC (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica, 66*(1).

# Disclaimer

The results in this paper are not official statistics. They were created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics NZ.

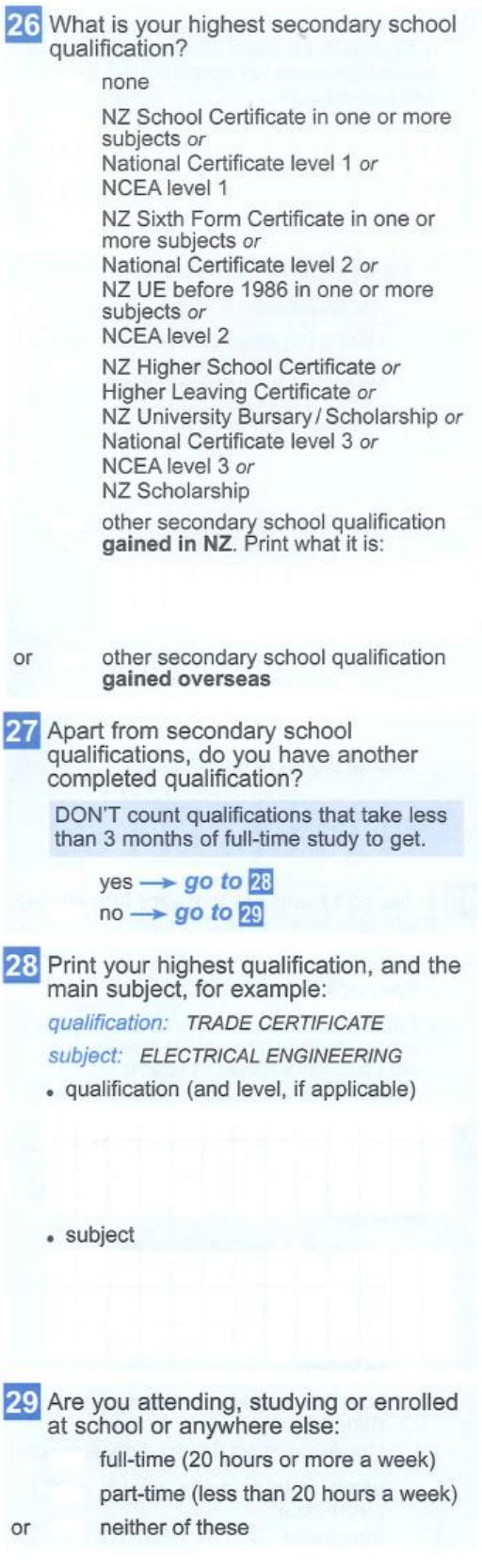
Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation. The results in this paper have been confidentialised to protect these groups from being identified.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI.

Privacy impact assessment for the Integrated Data Infrastructure (available from www.stats.govt.nz) has more information.

Note: All figures presented in this paper have been rounded to the nearest hundred to protect confidentiality.

# Appendix 1: 2013 Census questions on education and training



**26** What is your highest secondary school qualification?

- none
- NZ School Certificate in one or more subjects *or*
  National Certificate level 1 *or*
  NCEA level 1
- NZ Sixth Form Certificate in one or more subjects *or*
  National Certificate level 2 *or*
  NZ UE before 1986 in one or more subjects *or*
  NCEA level 2
- NZ Higher School Certificate *or*
  Higher Leaving Certificate *or*
  NZ University Bursary / Scholarship *or*
  National Certificate level 3 *or*
  NCEA level 3 *or*
  NZ Scholarship
- other secondary school qualification **gained in NZ**. Print what it is:

or other secondary school qualification **gained overseas**

**27** Apart from secondary school qualifications, do you have another completed qualification?

DON'T count qualifications that take less than 3 months of full-time study to get.

- yes → *go to* 28
- no → *go to* 29

**28** Print your highest qualification, and the main subject, for example:

*qualification:* *TRADE CERTIFICATE*

*subject:* *ELECTRICAL ENGINEERING*

- qualification (and level, if applicable)
- subject

**29** Are you attending, studying or enrolled at school or anywhere else:

- full-time (20 hours or more a week)
- part-time (less than 20 hours a week)

or neither of these

# Appendix 2: Constructing the IDI-URP (Census usually resident population from the IDI)

The IDI-URP is derived using information from ACC, tax, health, education, migration, and deaths data in the IDI. It consists of individuals who are present in IDI spine and had activity in ACC, tax, health, or education IDI datasets in the 12 months prior to 5 March 2013. Those not present in New Zealand on 5 March 2013 for a total of 10 months over the 12 months spanning census night are excluded using migration information available in the IDI. Individuals with a date of death prior to census night are also excluded. There are specific rules applied for different age groups.

Ages 0–4: Individuals were included in the population if they appeared in the IDI spine. For these ages there was no additional requirement of activity in the previous 12 months.

Ages 5+: Individuals were included in the population if they appeared in the IDI spine and had activity in one of the following IDI datasets in the 12 months prior to 5 March 2013:

- ACC claims
- Inland Revenue tax (employer monthly summary of tax paid at source, or annual tax return data; receipt of taxable benefit payments is included)
- Health (pharmaceutical prescriptions, GP enrolment, hospital admissions, non-admission hospital visits)
- Education (school enrolment, tertiary enrolment or attainment).

From the above populations, the following individuals were removed:

- Individuals who have a date of death prior to 5 March 2013 according to linked death records
- Individuals who according to linked migration data had moved overseas. Individuals were classified as having moved overseas if they were overseas on the reference date, and the total length of time spend overseas was at least 10 of the 12 months spanning the reference data (that is, the six months either side of the reference date).

# Appendix 3: IDI information about education and training

**Table 7**

**Information relating to determination of highest qualification level and field, and study participation available in the IDI as at May 2015**

| IDI table | Source | Highest qualification level and field | | | Study participation | |
|---|---|---|---|---|---|---|
| | | Qualification level | Qualification field of study | Date qualification achieved | Start and end of study period | Type of study (full-time or part-time) |
| **IDI tables used in this investigation** | | | | | | |
| Qualification completions | Tertiary Education Organisations via Ministry of Education | Yes<br>Some variables with same or close to statistical standard classification | Yes<br>Some variables with same or close to statistical standard classification | Yes – Year only | N/A | No |
| Qualification enrolments | Tertiary Education Organisations via Ministry of Education | N/A – But 'Highest secondary school qualification prior to enrolment' available | N/A | No – But first year in tertiary education available | Yes | Yes |
| Course enrolments | Tertiary Education Organisations via Ministry of Education | N/A – But 'Highest secondary school qualification prior to enrolment' available | N/A | No – But first year in tertiary education available | Yes | Yes |
| Student qualifications | NZQA via Ministry of Education | Yes<br>Some variables with same or close to statistical standard classification | N/A | Yes – Year only | N/A | N/A |

| | | | | | | |
|---|---|---|---|---|---|---|
| Student enrolments | School roll returns via Ministry of Education | N/A | N/A | N/A | Yes | MoE website says full-time / part-time is available on rolls, but not present in IDI. |
| Industry training | Industry Training Organisations via Ministry of Education | Yes – with completion indicator / level awarded<br>Also 'Previous qualification text' available. | Yes | No – But start of study period available | Yes | No |
| Targeting training | Training providers via Ministry of Education | Yes – but no completion indicator<br>Also 'Education history code' available. | No | No – But year of current training available | Yes | No |

| **IDI table** | **Source** | **Highest qualification level and field** | | | **Study participation** | |
|---|---|---|---|---|---|---|
| | | Qualification level | Qualification field of study | Date qualification achieved | Start and end of study period | Type of study (full-time or part-time) |
| **IDI tables not used in this investigation** | | | | | | |
| Social Welfare Number (SWN) | Work and Income via Benefit Dynamic Database via Ministry of Social Development | Yes | No | No | No | No |

| | | | | | | |
|---|---|---|---|---|---|---|
| MSD Education history | Work and Income via Benefit Dynamic Database via Ministry of Social Development | Yes | No | No | Yes | No |
| MSD Borrowing | StudyLink via Student Allowances and Loans database via Ministry of Social Development | Yes | Yes | No | Yes | Yes |
| Student qualifications | NZQA via Ministry of Education | Merged with field of study information. | Merged with level information. | Yes - Year only | N/A | N/A |
| MBIE Application points | Immigration New Zealand via Ministry of Business Innovation and Employment | Yes | No | No | No | No |
| Industry training | Industry Training Organisations via Ministry of Education | But only very broad inference from number of points claimed | Yes | No – But start of study period available | Yes | No |
| Social Welfare Number (SWN) | Work and Income via Benefit Dynamic Database via Ministry of Social Development | Yes | No | No | No | No |
| Source: Statistics New Zealand | | | | | | |

**Table 8**

**Time periods and subject populations for which information are available in the IDI as at May 2015**

| IDI table | Source | Time period | Subject population |
|---|---|---|---|
| **IDI tables used in this investigation** | | | |
| Qualification completions | Tertiary Education Organisations via Ministry of Education | Jan 1994-Dec 2013 | Individuals who have ever completed a qualification from a Tertiary Education Organisation (TEO) that receives government funding. (One record per qualification completed.) |
| Qualification enrolments | Tertiary Education Organisations via Ministry of Education | Jan 1994-Dec 2013 | Individuals who have ever enrolled in a qualification from a TEO. (One record per qualification enrolment.) |
| Course enrolments | Tertiary Education Organisations via Ministry of Education | Jan 2000 – Dec 2013 | Individuals who have enrolled in a course provided by a TEO that receives government funding. From the December 2014 Refresh, this table now also includes short courses. (One record per course enrolment.) |
| Student qualifications | NZQA via Ministry of Education | Jan 2007 – May 2014 (Some records back to 1930s included) | Individuals who have ever enrolled at a New Zealand school and have completed an NZQA approved qualification, including NCEA qualifications. |
| Student enrolments | School roll returns via Ministry of Education | Jan 2007 – Jul 2014 (Some records back to 1985) | Individuals who have ever enrolled at a New Zealand school. (One record per school enrolment.) |
| Industry training | Industry Training Organisations via Ministry of Education | Jan 2003 – Dec 2013 (Small number of records with start dates up to Jan 2016) | Individuals who have ever been or are in workplace-based training eligible for funding through the industry training and Modern Apprenticeships funds. (One record per placement) |

| | | | |
|---|---|---|---|
| Targeting training | Training providers via Ministry of Education | Jan 2001 – Dec 2013 | Individuals who have ever been placed in the following targeted training programmes, funded by government through the Tertiary Education Commission: Gateway, Skill Enhancement, Training Opportunities, Foundation Focussed Training Opportunities and Youth Training.<br>(One record per placement) |

| **IDI table** | **Source** | **Time period** | **Subject population** |
|---|---|---|---|
| **IDI tables not used in this investigation** | | | |
| Social Welfare Number (SWN) | Work and Income via Benefit Dynamic Database via Ministry of Social Development | Jan 1993 – Oct 2014 | Individuals who have ever received a main working-age social welfare benefit. Note education variable only available for individuals who have registered as a job-seeker at some time.<br>(One record per benefit recipient) |
| MSD Education history | Work and Income via Benefit Dynamic Database via Ministry of Social Development | Jan 1997 – Oct 2014 | Individuals who received a main working-age social welfare benefit. Note education variable only available for individuals who have registered as a job-seeker at some time.<br>(One record per job-seeker, per qualification level achieved) |
| MSD Borrowing | StudyLink via Student Allowances and Loans database via Ministry of Social Development | Jan 1999 - 2013 | Individuals who are eligible and who apply for and receive financial support under the student loans and allowances schemes.<br>(One record per individual per loan/allowance per year of loan/allowance) |
| MBIE Application points | Immigration New Zealand via Ministry of Business Innovation and Employment | 1997 – 31 Oct 2014 | Individuals who have applied as a principal applicant in the Skilled Migrant visa category, and have had their application approved.<br>(One record per points section of application, per application, per individual) |
| Source: Statistics New Zealand | | | |

# Appendix 4: Full matrices for similarities between census and IDI

**Table 9**

**Highest qualification level in census and IDI where a level is present for both sources, number and percentage**

| | | Census highest qualification level | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | 2 | | 3 | | 4 | | 5 / 6 | | 7 | | 8 | | 9 | | 10 | | Total |
| | | Number | % | Number | % | Number | % | Number | % | Number | % | Number | % | Number | % | Number | % | Number | % | |
| IDI highest qualification level | 1 | 100,700 | 54 | 13,800 | 7 | 5,300 | 2 | 16,600 | 11 | 8,500 | 6 | 2,000 | 1 | 400 | 1 | 300 | 1 | 100 | 1 | **147,700** |
| | 2 | 34,800 | 19 | 125,300 | 64 | 26,500 | 12 | 22,800 | 15 | 28,300 | 18 | 17,100 | 6 | 2,800 | 5 | 3,400 | 7 | 600 | 6 | **261,600** |
| | 3 | 22,400 | 12 | 26,100 | 13 | 137,700 | 64 | 22,600 | 15 | 18,800 | 12 | 26,700 | 10 | 4,200 | 8 | 4,300 | 9 | 1,100 | 12 | **263,900** |
| | 4 | 15,500 | 8 | 16,400 | 8 | 21,800 | 10 | 58,800 | 38 | 12,900 | 8 | 7,100 | 3 | 1,300 | 2 | 1,300 | 3 | 200 | 2 | **135,300** |
| | 5 / 6 | 7,200 | 4 | 8,600 | 4 | 13,000 | 6 | 22,100 | 14 | 63,600 | 42 | 12,600 | 5 | 2,400 | 4 | 1,800 | 4 | 200 | 2 | **131,500** |
| | 7 | 3,400 | 2 | 5,100 | 3 | 10,100 | 5 | 8,700 | 6 | 17,200 | 11 | 176,100 | 67 | 11,800 | 21 | 5,500 | 12 | 700 | 8 | **238,600** |
| | 8 | 100 | 0 | 300 | 0 | 900 | 0 | 300 | 0 | 2,300 | 2 | 19,100 | 7 | 31,100 | 56 | 3,400 | 7 | 700 | 8 | **58,200** |
| | 9 | 100 | 0 | 100 | 0 | 400 | 0 | 100 | 0 | 300 | 0 | 1,300 | 0 | 1,000 | 2 | 27,000 | 57 | 600 | 6 | **30,900** |
| | 10 | 600 | 0 | 600 | 0 | 800 | 0 | 800 | 1 | 1,300 | 1 | 1,400 | 1 | 300 | 1 | 300 | 1 | 5,100 | 55 | **11,200** |
| **Total** | | **184,800** | **100** | **196,300** | **100** | **216,500** | **100** | **152,800** | **100** | **153,200** | **100** | **263,400** | **100** | **55,300** | **100** | **47,300** | **100** | **9,300** | **100** | **1,278,900** |

Source: Statistics New Zealand

**Table 10**

**Study participation, including full-time/part-time status in census and IDI, number and percentage**

| | | | Census | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Studying | | | | | | Not studying | | Missing | | Total | |
| | | | Full-time | | Part-time | | Full & part-time | | | | | | | |
| | | | Number | % | Number | % | Number | % | Number | % | Number | % | Number | % |
| IDI | Studying | Full-time | 249,600 | 8 | 14,600 | 0 | 100 | 0 | 23,400 | 1 | 14,100 | 0 | **301,800** | **10** |
| | | Part-time | 13,900 | 0 | 39,500 | 1 | - | - | 12,800 | 0 | 2,100 | 0 | **68,300** | **2** |
| | | Full & part-time | 11,600 | 0 | 5,600 | 0 | - | - | 4,000 | 0 | 1,100 | 0 | **22,200** | **1** |
| | | Not available | 6,200 | 0 | 11,700 | 0 | - | - | 256,000 | 8 | 14,100 | 0 | **288,000** | **9** |
| | **Not studying** | | 40,200 | 1 | 43,800 | 1 | 300 | 0 | 2,212,100 | 71 | 148,400 | 5 | **2,444,700** | **78** |
| | **Total** | | **321,500** | **10** | **115,200** | **4** | **400** | **0** | **2,508,300** | **80** | **179,800** | **6** | **3,125,200** | **100** |

Source: Statistics New Zealand