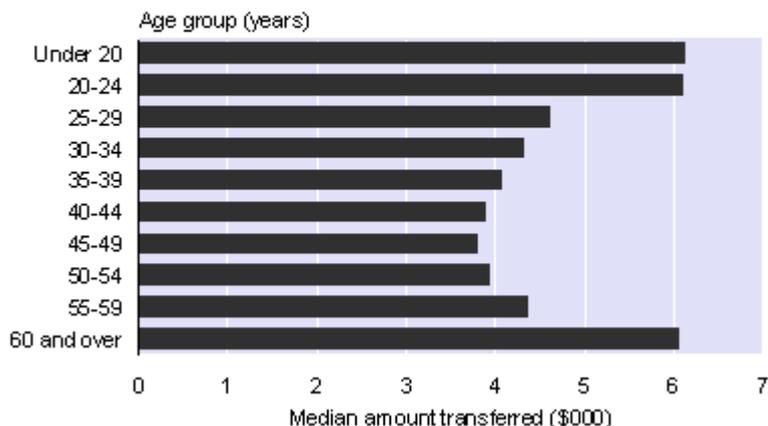# Student Loan Borrowers
# 1997–2004

## Highlights

- Borrowers under 25 years and those aged 60 years and over had the highest median amount transferred for 2004.

- The mean loan balance for borrowers overseas who had enrolled in health courses was $16,590 more than for those assumed to be residing in New Zealand at March 2005.

- Approximately one-sixth of all borrowers between 1997 and 2004 had paid back their loans in full at March 2005.

- Students at honours level or higher in their last year of borrowing had the highest rate of full repayment (21–25 percent) at March 2005.



**Median Amount Transferred for 2004**
*By age group*
For people who borrowed in 2004

Brian Pink
**Government Statistician**

There is a companion Media Release published  – *Integrated Data on Student Loan Borrowers: 1997–2004*.

# Commentary

## Student loan dataset

The statistics included in this release have been produced from an updated dataset of student loan and tertiary education data established and held by Statistics New Zealand. The main component of the dataset is an integrated longitudinal dataset on student loan borrowers, which was first created in 2002. Data has since been added, extending it to students who borrowed under the Student Loan Scheme in any of the years 1997 to 2004. Previously published data has also been updated to take advantage of improvements in the quality of the data that agencies were able to supply to Statistics NZ, leading to revisions in the 1997 data (Table 1.2).

The integrated dataset enables the education characteristics of borrowers to be linked with loans information. This supplements the administrative data available in the *Annual Report on the Student Loan Scheme*, published by the Ministry of Education, the Ministry of Social Development (StudyLink) and Inland Revenue. (See the Technical Notes of this release for further details.)

## Definitions

*Amount transferred:* the borrowings from StudyLink minus repayments to StudyLink transferred annually to Inland Revenue.

*Loan balance:* the total loan balance held by Inland Revenue.

## Borrowing in 2004

It is estimated that about 147,800 people borrowed under the Student Loan Scheme in 2004. This represents a 46 percent increase (from 101,300 in 1997) in the number of borrowers over the period 1997–2004.

In 2004, the mean amount transferred from StudyLink to Inland Revenue was $6,330, and half of all borrowers transferred less than $5,300. One percent of those who borrowed under the scheme in 2004 had repaid the full amount they borrowed before their loan was transferred. (See the Technical Notes of this release for a full definition of amount transferred.)

Close to half (46 percent) of all amounts transferred (including zero amounts) were under $5,000. The remaining 54 percent included a small proportion (5 percent) that were $12,500 or more.
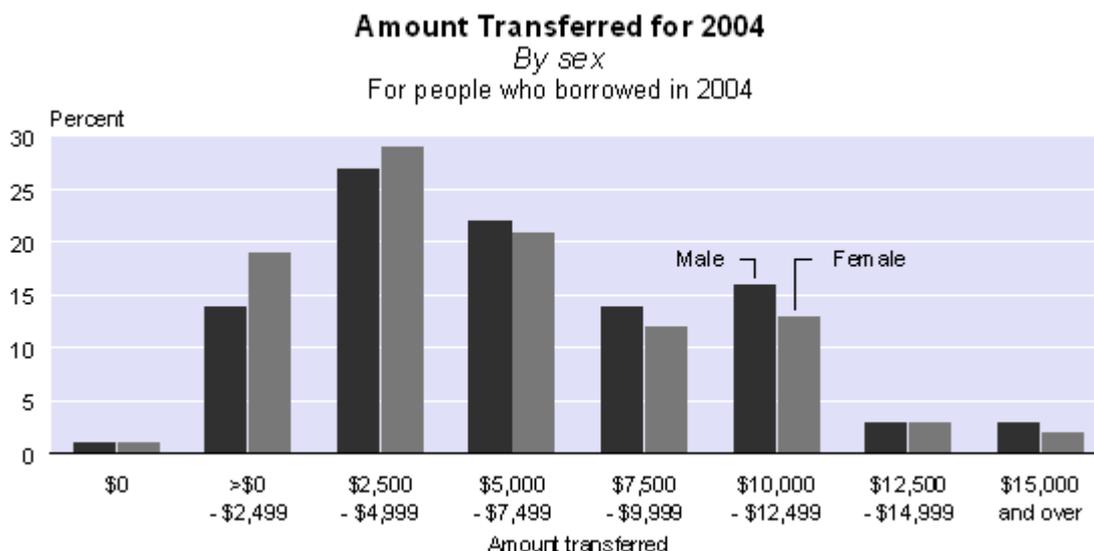
### Age group in 2004
Borrowers under 25 years or 60 years and over transferred the highest median amount (over $6,000) for 2004. Those under 25 years were 59 percent of borrowers, while only 1 percent of borrowers were aged 60 years and over.

The median amounts transferred for the age groups between 40–54 years were the lowest (all under $4,000).

The number of borrowers aged 40 years and over increased nearly two and a half times between 1997 and 2004 (from 7,305 to 17,994). The number of borrowers aged 55 years and over increased more than fivefold (from 504 in 1997 to 2,583 in 2004).

**Sex**
In 2004, 60 percent of borrowers were female. Females tended to borrow less than males. Approximately half the female borrowers had an amount transferred of under $5,000. The median amount was $5,050, compared with $5,630 for males.

**Amount Transferred for 2004**
*By sex*
For people who borrowed in 2004



**Ethnic group in 2004**
The majority of borrowers identified as European (64 percent) or Māori (22 percent).

Those who identified as Pacific peoples transferred the highest median amount ($5,470), while Māori students who borrowed in 2004 had the lowest median amount transferred ($5,130), with 85 percent of Māori borrowers having an amount transferred of under $10,000.

**Level of study in 2004**
Students enrolled in a bachelor's degree programme made up 52 percent of borrowers, and had the highest median amount transferred ($5,530) for 2004. This level of study was the only one with a median higher than the overall median.

Those enrolled in level 1–3 certificate programmes (32 percent) were the next largest group of borrowers. The number at this level more than doubled from 1997 (from 21,411 in 1997 to 46,578 in 2004).
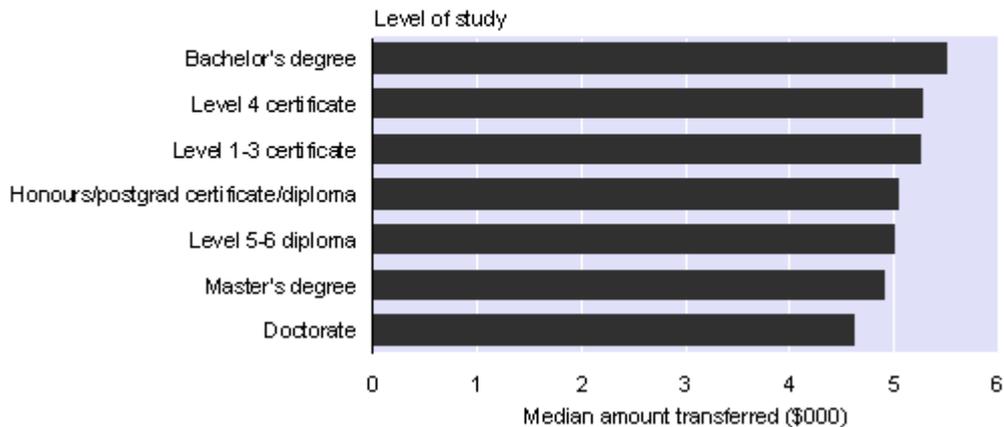
The number of borrowers in level 4 certificate programmes was six and a half times more in 2004 than in 1997 (up from 1,410 to 9,213). These borrowers had the second-highest median amount transferred ($5,300) for 2004.

Of the total number of borrowers, 5 percent had an unknown level of study.

**Median Amount Transferred for 2004**
*By level of study*
For people who borrowed in 2004

Level of study

| Level | |
|---|---|
| Bachelor's degree | |
| Level 4 certificate | |
| Level 1-3 certificate | |
| Honours/postgrad certificate/diploma | |
| Level 5-6 diploma | |
| Master's degree | |
| Doctorate | |

0    1    2    3    4    5    6
Median amount transferred ($000)

## Field of study in 2004

Students enrolled in food, hospitality and personal services courses had the highest median amount transferred ($6,550), with most of these borrowers (68 percent) having borrowed $5,000 or more.

Borrowers enrolled in education courses transferred the lowest median amount ($4,370).

A quarter of borrowers (25 percent) were enrolled in society and culture courses and 21 percent were enrolled in management and commerce courses. These were the two largest groups of borrowers by field of study.

The number of borrowers enrolled in mixed field programmes was three and a half times greater than in 1997 (up from 4,194 to 15,063).

There were 8 percent of borrowers whose field of study was unknown.

## Type of training provider in 2004

Borrowers enrolled at private training establishments transferred the highest median amount ($6,560), with 67 percent of them borrowing $5,000 or more.

University students, half of all student loan borrowers, had the next highest median amount transferred ($5,550).

There were four times as many borrowers enrolled at private training establishments in 2004 as in 1997 (up from 6,678 to 28,017). This could reflect the increase in enrolment in certificate programmes at this type of provider. The increase could also be due to enrolment data for students at private training establishments only being fully available from the Ministry of Education from 2000.

The number of borrowers enrolled at wānanga was 14 times greater in 2004 than in 1997 (up from 804 to 11,115), reflecting the sharp increase in enrolments in certificate programmes at wānanga.

The smallest number of borrowers was at other tertiary education providers (1 percent). Borrowers at a college of education had the lowest median amount transferred ($4,040).

# Loan balances in March 2005

It is estimated that about 490,600 people borrowed under the Student Loan Scheme between 1997 and 2004. The data in Table 2 gives an indication of the sizes of the student loans in the integrated dataset, at 31 March 2005.

The overall mean loan balance of these borrowers was $13,120, while the median loan balance was $8,400.

By 31 March 2005, approximately one-sixth (16 percent) of student loan borrowers had paid back their loans in full. (See the Technical Notes for definitions.) Of all borrowers, 39 percent had loan balances of less than $10,000, 22 percent had balances between $10,000 and $19,999, and 23 percent owed $20,000 or more.

### Sex
Males and females had the same rate of full repayment (16 percent). Due to the short duration of the student loan integrated dataset, this pattern of full repayment for male and female borrowers may vary when more data become available for analysis.

The great majority of all loans were less than $20,000. Male borrowers had a higher median loan balance ($8,750) than female borrowers ($8,140), and 1.2 percent of male borrowers owed $70,000 and over compared with 0.6 percent of female borrowers.

Over half (56 percent) of borrowers were female and 56 percent of this group owed less than $10,000. The high proportion of women in the dataset reflects enrolment trends; women were more than half of all enrolments in each year covered by the dataset.



Student Loan Balance at 31 March 2005
By sex
For borrowing in any year, 1997 to 2004
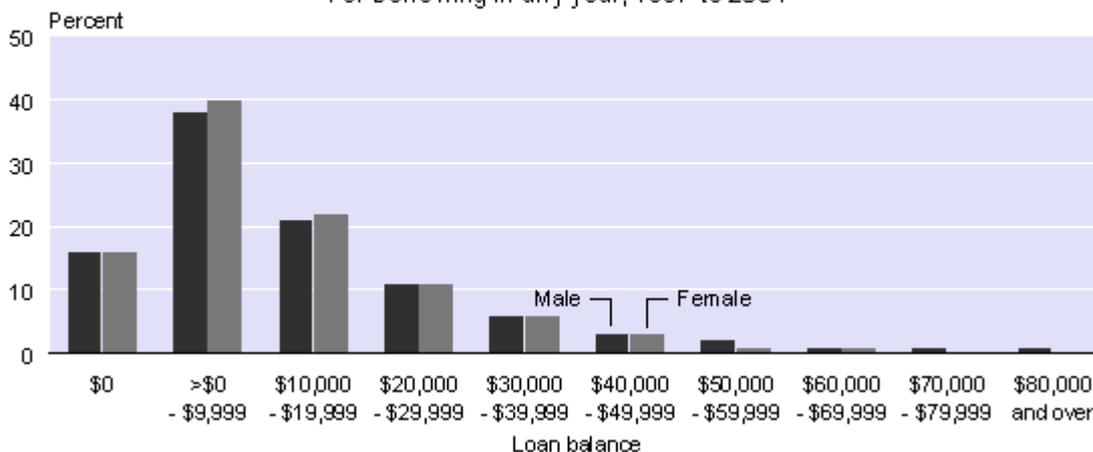
## Age group in 2004
The rates of full repayment for borrowers between 25 and 59 years were higher than average (16 percent), ranging from 20–25 percent.

Those least likely to have repaid their loans in full were borrowers in the youngest age group (under 20 years). Only 3 percent of these borrowers had fully repaid their loans. It is likely that most were still studying and had not started repayments.

Students aged 20–24 years were 30 percent of borrowers. They had the highest median loan balances ($11,150) and the second–lowest rate of full repayment (10 percent). It is likely that many were still studying and had not started repayments, while others in the early years of their careers had made little progress in repaying their loans.

The next two highest median loan balances were for those aged 25–29 years ($10,720) and those 60 years and over ($9,010).
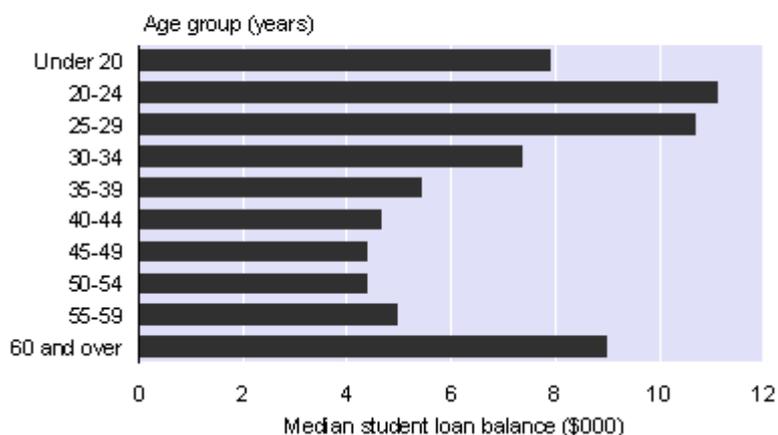
Borrowers aged 40–59 years had lower median loan balances (under $5,000) than the overall median ($8,400).

There was a high concentration of borrowers in the greater than $0 to $9,999 loan balance band across all age groups.



**Median Student Loan Balance at 31 March 2005**
*By age group*
For borrowing in any year, 1997 to 2004

## Ethnic group in 2004
Compared with other ethnic groups, European students had the highest rate of full repayment (18 percent).
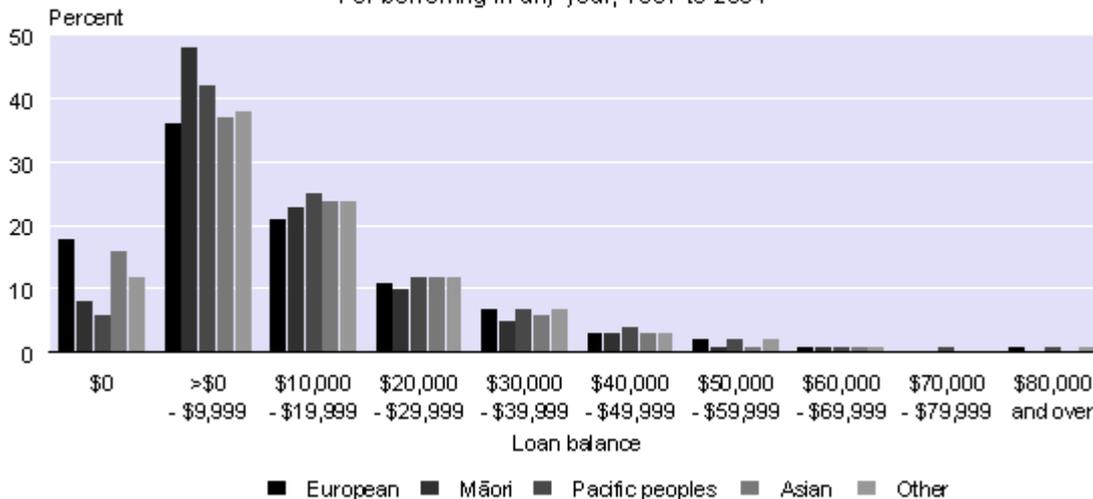
Pacific peoples and Māori borrowers had the lowest rates of full repayment, 6 percent and 8 percent, respectively. Māori students also had the lowest median loan balance ($8,480).

Students who identified as Pacific peoples had the highest median loan balance ($10,300).

**Student Loan Balance at 31 March 2005**
*By ethnic group*
For borrowing in any year, 1997 to 2004



## Level of study in last year of borrowing

Those who had studied at honours level or higher in their last year of borrowing had higher rates of full repayment (21–25 percent) than average (16 percent). This could be attributed to higher levels of qualification leading to higher incomes (see *Income of Student Loan Scheme Borrowers*, published by the Ministry of Education, March 2005). This leads to a faster rate of repayment (see *Paying Off a Student Loan*, published by the Ministry of Education, 2005).

The median loan balances for those enrolled in bachelors degrees or higher were much higher than the overall median loan balance ($8,400): specifically, doctorate ($17,700), master's ($12,990), honours or post-graduate diploma or certificate ($12,350) and bachelor's ($12,210). This is indicative of higher fees and longer periods of study.

Nearly two-fifths of borrowers (39 percent) were enrolled in a bachelor's degree. Those enrolled in level 1–3 certificate programmes formed the next largest number of borrowers (35 percent), but they had the lowest median loan balance ($6,640).

**Median Student Loan Balance at 31 March 2005**
*By level of study*
For borrowing in any year, 1997 to 2004
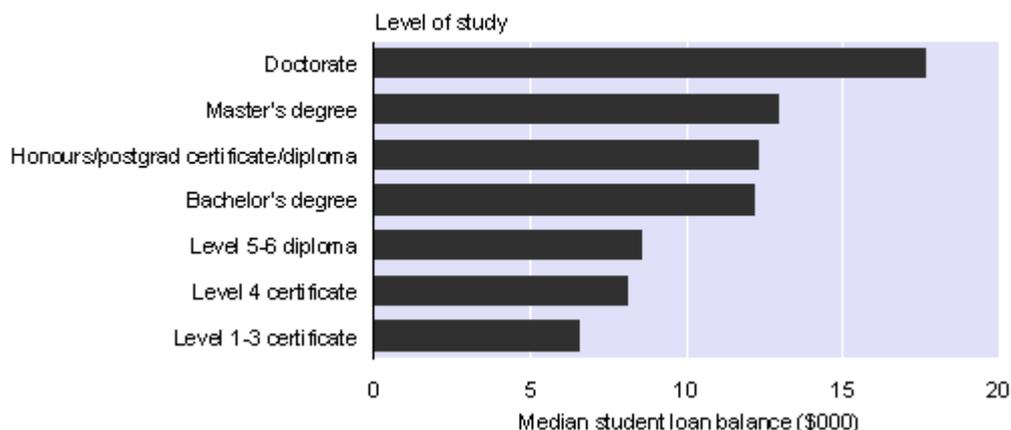
**Field of study in last year of borrowing**
The rates of full repayment were highest for engineering and related technologies courses (20 percent) and agriculture, environmental and related studies courses (19 percent).
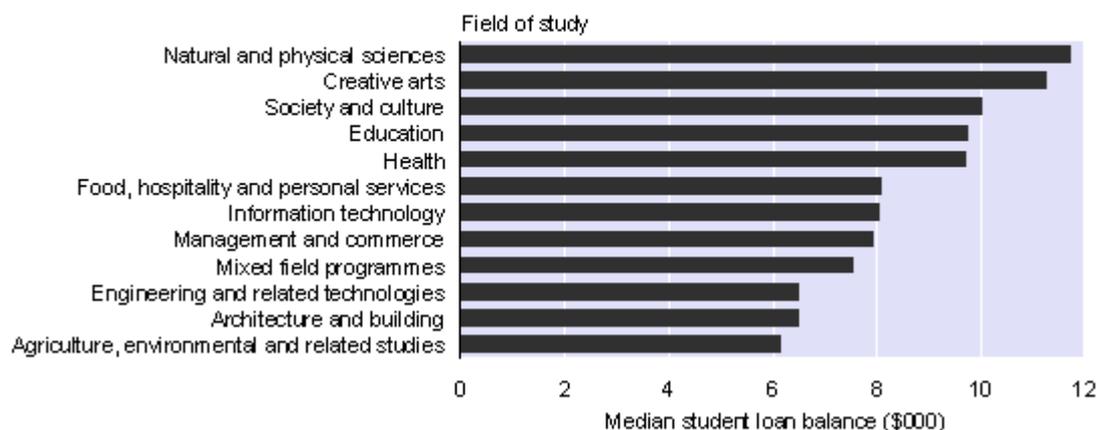
Students enrolled in natural and physical sciences courses had the highest median loan balance ($11,770) while those in agriculture, environmental and related studies courses incurred the lowest median loan balance ($6,200).

The largest number of borrowers by field of study were those enrolled in society and culture courses and management and commerce courses (21 percent each).

**Median Student Loan Balance at 31 March 2005**
*By field of study*
For borrowing in any year, 1997 to 2004

Field of study

| Field of study | Median student loan balance ($000) |
|---|---|
| Natural and physical sciences | ~11.8 |
| Creative arts | ~11.3 |
| Society and culture | ~10 |
| Education | ~9.7 |
| Health | ~9.7 |
| Food, hospitality and personal services | ~8 |
| Information technology | ~8 |
| Management and commerce | ~7.8 |
| Mixed field programmes | ~7.4 |
| Engineering and related technologies | ~6.3 |
| Architecture and building | ~6.3 |
| Agriculture, environmental and related studies | ~6.1 |

**Type of training provider in last year of borrowing**
The borrowers at different provider types had different rates of full repayment. The rate was highest for students enrolled in a college of education (21 percent) and lowest for students enrolled in wānanga (4 percent).

Students enrolled in other tertiary education providers formed the smallest number of borrowers (1 percent), but they incurred the highest median loan balance ($11,790).

Two-fifths of student loan borrowers (40 percent) were studying at university and these students incurred the second highest median loan balance ($11,760).

Institute of technology/polytechnic students formed the second largest number of borrowers (37 percent), but these borrowers had the lowest median loan balance ($6,170).

# Student loan holders overseas in 2004

Information on student loan holders who were overseas relates only to individuals who notified Inland Revenue that they were leaving or had left New Zealand. This means that the number of student loan holders residing overseas is underestimated by an unknown quantity, because not everyone advises Inland Revenue of their residency status.

At March 2005, the mean loan balance of people who borrowed in any year between 1997 to 2003 but did not borrow in 2004, and who were overseas, was much higher ($22,640) than those assumed to be residing in New Zealand ($10,850).

## Field of study during final year of borrowing

The mean loan balance for borrowers overseas who had enrolled in health courses during their last year of study was $16,590 more than for those assumed to be residing in New Zealand as at March 2005.

The highest mean loan balances for borrowers overseas were held by those enrolled (during their last year of study) in health courses ($28,350) and the lowest by those in food, hospitality and personal services courses ($14,190).

The highest mean loan balance for borrowers assumed to be residing in New Zealand ($13,920) was for those who had enrolled (during their final year of borrowing) in creative arts courses and the lowest ($8,330) was for those in agriculture, environmental and related studies courses.

## Sex

Male borrowers had a higher mean loan balance than female borrowers for both the overseas and assumed to be residing in New Zealand categories. However, the difference between male and female borrowers was greater for those overseas ($2,550) than for those assumed to be residing in New Zealand ($1,100).

## Age group in 2004

Borrowers overseas in the age groups from 25 to 34 years incurred mean loan balances higher than the overall mean of $22,640. The mean loan balance was $26,190 for those aged 25–29 years and $27,040 for those aged 30–34 years.

The difference in mean loan balance between student loan borrowers overseas and those assumed to be residing in New Zealand was greatest for borrowers in the 25–29 and 30–34 age groups. The difference was over $11,000 for both age groups.

For technical information, contact:
Mei-Fen Sundgren or Guido Stark
Wellington 04 931 4600
**Email:** info@stats.govt.nz

# Technical notes

## Introduction

The statistics in this release have been produced using a database of student loan and tertiary education data held by Statistics New Zealand. One of the important components of the database is an integrated dataset on student loan borrowers. This dataset was created by linking administrative records from a number of government agencies:

- individual students' tertiary enrolment data from the Ministry of Education (MoE)
- individual students' borrowing data from the now-defunct Student Loan Account Manager (SLAM) provided by Inland Revenue and the MoE
- individual students' borrowing data (from 2000 onwards) from StudyLink, a service of the Ministry of Social Development (MSD)
- individuals' repayment and income data from Inland Revenue.

The integrated dataset contains data on the loans of students who borrowed under the Student Loan Scheme in any of the years 1997 to 2004. It has recently been updated to include data for 2004, and this Hot Off The Press provides the first release of results following the update. The statistics focus on the characteristics of student loan borrowers and how these relate to borrowing levels and cumulative loan debt in the latest year, rather than on comparisons with previous years.

The process of updating the data involved linking the new 2004 data, and backfilling information for previous years to take advantage of improvements in the quality of data that agencies were able to supply to Statistics New Zealand. This means previously published 1997 data (Table 1.2) has thus been revised.

The integrated dataset is the only data source that links information on borrowers' incomes, loan balances and repayments with their enrolment details and other characteristics. The inclusion of information on borrowings under the Student Loan Scheme, and the ability to link records over time, further enhances its usefulness as a single integrated data source from which to provide statistics for strategic policy and financial analysis.

Some statistics presented in this report are not unique to the integrated dataset. The agencies that supply data on student loan borrowers also publish a range of statistics in the Student Loan Scheme Annual Report. In some cases these may differ from the statistics in this report because they have a different effective date, are derived from a different administrative database, or use a different definition of the target population. For example, in this report the figures on borrowers include only those whose loans had been transferred to Inland Revenue at the date the data were extracted, and borrowers include those who had fully repaid their loans. (Further details are provided below.)

# Background

The Student Loan Scheme began in 1992. In June 2000, the Auditor-General released a report entitled *Student Loan Scheme – Publicly Available Accountability Information*.The report proposed that Statistics New Zealand integrate selected datasets relating to the Student Loan Scheme with a view to providing statistics for strategic policy, financial risk management, financial reporting and forecasting. As a result of work undertaken to follow up the Auditor-General's report, Statistics New Zealand was directed to lead an investigation into the privacy, logistical and data issues around data integration. Other agencies involved in the exercise were the MoE, Inland Revenue, The Treasury, the then Department of Work and Income, and the then Ministry of Social Policy. The Privacy Commissioner was consulted over the privacy issues involved in integrating personal data from different sources.

That scoping exercise led to Statistics New Zealand being directed to undertake further investigations. In particular, work was required to address some issues raised by the Privacy Commissioner. A trial integration exercise was also undertaken to determine the types of matching required to maximise the representativeness of the integrated dataset (ie to minimise the differences between students in the integrated dataset and the overall group of students with loans). That work was successfully completed and reported on to Cabinet.

In April 2002, Cabinet funded Statistics New Zealand to construct an integrated dataset on student loan borrowers dating back to 1997. Annual updates were also funded, with statistics to be released by Statistics New Zealand each year. Cabinet's approval for the work was given in the knowledge that the Privacy Commissioner was comfortable with the revised methodology proposed by Statistics New Zealand. During 2002, Statistics New Zealand established the first integrated dataset on student loan borrowers, containing data on students who borrowed in any of the years 1997 to 2000. The first release of data from this new integrated dataset occurred in December 2002. A second release was issued in November 2004, when the dataset was updated to include data on students who borrowed in 2001 and 2002. A third release was issued in November 2005, when the dataset was updated to include data on students who borrowed in 2003.The current document is the fourth such release and includes data on students who borrowed in any of the years 1997 to 2004.

## Differences between data published in 2002, 2004, 2005 and 2006

As well as the integrated dataset expanding to include data on students who borrowed in 2001 and 2002, some other substantial changes meant that the statistics released in 2004 were not comparable with those released in 2002. As familiarisation with and understanding of student loan data increased, ways were identified to improve the integrated dataset and the quality of the statistics resulting from it. This led to changes in the way some variables are specified. For example, the date behind the structure of Inland Revenue data became the 'return period date' (ie the financial year to which a transaction applies), rather than the date on which a particular transaction had an effect on an account.

Additionally, to provide a more complete picture of student loans, $0 was included as a category in tables. In Table 1, this reflects people who borrowed during a year but paid off the full amount before the loan was transferred to Inland Revenue, and in Table 2, people who had paid off their loan in full by March 2005. Compared with the data published in 2002, the inclusion of these zero values will have reduced the mean and median dollar figures.

From 2004, Statistics New Zealand also made changes to the integration process. The decision was taken to re-link the student/borrower records for the years 1997 to 2000 which made up the original integrated dataset. This is because:

- varying definitions of 'year' (calendar, study and tax year) were found to have introduced inconsistencies into the linked data
- the data model (used as the basis for storing data on the database) was revised and simplified, meaning the previous link file could not be used.

As part of the reintegration, borrowing data was matched in two separate tasks: one involving older SLAM data from 1997 to 2000, and the other involving MSD data from 2000 to 2002. Loans data was merged with Inland Revenue data through the use of tax file numbers, and this combined data was then probabilistically matched to MoE data (see below).

In 2005, the addition of data for 2003 became available, and the student/borrower records for the years from 1997 to 2002 were also re-linked as improved data became available for these years.

Compared with previously published data in 2002 and 2004, there was a significant reduction in the number of people included in the $0 category, which was previously overstated. A respecification of the target population for the output tables enabled Statistics New Zealand to more accurately define this category within the dataset. In consequence, the number of people in this category were not comparable with that shown in the previous release. This reduction in zero values also tended to increase the mean and median dollar figures.

### Current changes as reflected in this Hot Off The Press
The addition of data for the latest available year, 2004, has increased the total number of borrowers included in Table 2 by approximately 12 percent to 490,560. Increases in the number of borrowers covered will continue to be a feature of each subsequent release of the student loans dataset, and will reflect the addition of new borrowers to the scheme.

The categorisation of fields of study is now more accurately derived from the New Zealand Standard Classification for Education (NZSCED), due to enhancements in the data supplied by the MoE for all years 1997–2004. Field of study figures produced from current MoE data should therefore not be compared with those previously published.

### Future changes
Users should be aware that another exercise being undertaken at the MoE this year will again review the consistency and quality of 'NZSCED' in MoE enrolments data, which may affect future data releases.

## Matching methodology

The integrated dataset was created in two stages. The first stage merged the records of Inland Revenue and MSD (or SLAM) using tax file numbers. The second stage used a probabilistic matching methodology to match the MoE data to the already linked 'loans' dataset created in the first stage. Variables used in this second stage were: student identification number; tertiary institution number; year of study/borrowing; sex; day, month and year of birth; surname; first initial of first name; and ethnicity. Probabilistic matching allows records to be linked when some of the matching variables are not unique, have incorrect values, or are missing.

The overall link rate for the integrated dataset, after both exact and probabilistic matching was completed, was 93 percent. There was considerable variation in the link rate over the years, due in most part to the lack of education data for students at private training establishments (PTEs) prior to 2000. For 1997 to 1999, the average link rate was 87 percent, and for 2000 to 2004, the average link rate was 96 percent.

It is important to note that when Statistics New Zealand is satisfied that all checking has been completed, unique identifiers are not carried through to the integrated dataset. Annual updating of the integrated dataset is achieved through the use of a non-reversible algorithm. Statistics New Zealand runs the algorithm against the unique identifiers to produce a new linking identifier for each unit record. It is this new variable that enables each year's data to be added to the integrated dataset.

Further information on the matching methodology can be found in two methodological reports published on Statistics New Zealand's website (see www.stats.govt.nz/research-reports/student-loan-integration).

## Differences between all student loan borrowers and those in this release

Users should be aware that official statistics on student loans are also published in the *Student Loan Scheme Annual Report*, jointly published each year by the MoE, Inland Revenue and the MSD. The official statistics on student loan borrowers included in this Hot Off The Press may differ from statistics published in the annual report. This is because the source administrative data was provided to Statistics New Zealand at a specified cut-off date which differs from the one used for the compilation of the annual report. In addition, the tables in this Hot Off The Press use different populations from those in the annual report.

For example, because all financial data in the tables comes from Inland Revenue, information is restricted to borrowers whose loans had been transferred from StudyLink to Inland Revenue by the date Inland Revenue extracted the data for Statistics New Zealand (October 2005). The figures show the situation as at the end of the tax year, on 31 March 2005. This means that in this Hot Off The Press the population of students who borrowed in 2004 is defined as those who borrowed during the 2004 academic year, whose study end date had occurred by 31 December 2004, and whose borrowing record for that year had been transferred from StudyLink to Inland Revenue by October 2005.

The use of probabilistic matching significantly increased the percentage of student loan borrowers whose information was included in the integrated dataset. However, it should be noted that some types of student loan borrowers were still under represented, even after both stages of matching were completed. Due to the absence of data from the MoE's dataset on students attending PTEs before 2000, the link rate of PTE students prior to 2000 was around 10 percent. From 2000 onwards, the link rate rose to around 91 percent.

# Key variables in the Hot Off The Press tables

## Demographic and socio-economic characteristics
- *Age group:* the person's age at 1 July in the reference year in the table.
- *Ethnic group:* data for the years prior to 2000 are sourced from the MoE, and assign each individual to one ethnic group only. For those who identified with more than one ethnic group this allocation was based on the following order of priority: Māori, Pacific peoples, Asian, Other, European. For example, if a student loan borrower identified as both Samoan and European, then they were assigned to the Pacific peoples ethnic group. Subsequently, unprioritised ethnicity data began to become available, from StudyLink (from 2000 onwards) and the MoE (from 2001 onwards), and where available this unprioritised ethnicity data has been used. The data from both sources may include up to three ethnicities for each individual, meaning that figures relating to 2000 may reflect identification with as many as three different ethnic groups per student, and those from 2001 as many as six different ethnic groups per student. From 2001 onwards, unprioritised ethnicity data are available for nearly all borrowers. In a small number of cases, where unprioritised ethnicity data are not available, the older prioritised MoE ethnicity data have been used. The change from prioritised to unprioritised ethnicity data will affect ethnic group comparisons between the tables relating to borrowers in 1997 and 2004. Tables spanning the 1997 to 2004 period will reflect a combination of prioritised and unprioritised ethnicity data.
- *Declared overseas:* includes people who advised Inland Revenue they were living overseas or were departing to live overseas during the reference year. There will be additional holders of student loans who were overseas and had not advised Inland Revenue, but this figure cannot be accurately quantified.

## Study-related variables
- *Level of study:* the qualification(s) for which the student was enrolled (completed or not). A student can be enrolled in more than one level of study.
- *Field of study:* the New Zealand Standard Classification of Education of the field(s) of study or subject of a programme of study in which the student was enrolled. A student can be enrolled in more than one field of study.
- *Provider type:* the type of provider at which the student was enrolled (completed or not). A student can be enrolled at more than one provider type.

## Student loan variables
- *Amount transferred:* the amount transferred from StudyLink to Inland Revenue (ie drawdowns minus repayments that were made before the loan was transferred). StudyLink manages individual loan accounts on a yearly cycle, with the loan balance being transferred each year to Inland Revenue for collection. The transfer takes place on 28 February for the majority of loan accounts in which the course of study was completed by 31 December the previous year.
- *Loan balance:* the total loan balance owing, including the administration fee and interest accrued (net of any write-offs), less any repayments received. An individual's total current student loan balance assessed and held by Inland Revenue.
- *Full repayment:* in this report refers to the repayment of all assessed student loan balance so that the loan balance held by Inland Revenue is zero.

**Year**

Note that enrolment and borrowing data generally relate to the calendar year. Inland Revenue data relate to the tax year (1 April to 31 March). Student loan balances are shown as at 31 March, while students' borrowing and summary characteristics generally relate to a calendar year as indicated in the tables.

## Student loan and tertiary education data held by Statistics New Zealand

Statistics New Zealand's student loan and tertiary education database has several components:
- The integrated dataset (ie matched data for students who had loans in any of the years 1997–2004)
- Inland Revenue information for 1997–2004 for student loan borrowers whose records were not matched
- 1992–1996 information from Inland Revenue on student loan borrowing
- Some income information from 1996 onward for any person who took out a student loan before 1997
- MoE enrolment data for all formal students from 1997-2004, regardless of whether they were student loan borrowers. A formal student is one who is enrolled in a formal programme of study at a tertiary education provider for more than one full-time week
- SLAM information for students who borrowed in or before 1999. SLAM data was provided by both the MoE and Inland Revenue.

Several of these components are updated each year. Data on other tertiary education issues may be added to the database in the future.

The statistical tables included with this release are just examples of data that can be produced from the database. Customised requests can be run by Statistics New Zealand and, on application, researchers may be able to access unidentified data in Statistics New Zealand's Data Laboratory.

## Reliability of the data

Statistics New Zealand has made every attempt to minimise errors in the student loan and tertiary education database, but two types of error will have occurred: errors in source data and errors due to record linking. Statistics New Zealand validated source data as it was received from each agency. If errors were detected, agencies were asked to rerun their data and provide it again. There will still, however, be some errors in the source data supplied. Omissions in the original collection of information and errors in data entry and processing will have occurred. These cannot be quantified. In terms of errors in record linking, Statistics New Zealand used sound probabilistic matching methodology, but errors will still be present. For example, over all years an average of 7 percent of loan records were not linked to a MoE record, with considerable improvement from 2000 onwards. There will be some cases where an Inland Revenue record has been linked to the wrong MoE record, but this is estimated as affecting fewer than 1 percent of the links.

# Appendix 1: Example from the 2001 Census of the degree to which ethnic identification may be understated by the use of prioritised ethnicity data

The 2001 Census recorded ethnicity using a 'total response' approach, where an individual could identify with more than one ethnic group. Consequently, the total of all ethnicities collected is greater than the total population because up to six ethnicities were recorded for each individual.

Total response ethnicity data is sometimes referred to as 'unprioritised' ethnicity data, in contradistinction to 'prioritised' ethnicity data, which assigns each individual to one ethnic group only. Prior to 2000, data on student loan borrowers sourced from the MoE assigned students to one ethnic group only. Where a student identified with more than one ethnic group, this allocation was based on the following order of priority: Māori, Pacific peoples, Asian, Other, European. For example, if a person identified as both Samoan and European, then they were assigned to the Pacific peoples ethnic group.

The integrated dataset on student loan borrowers is unusual in combining both prioritised and unprioritised ethnicity data, as described in these Technical Notes. This is done for practical purposes, because prior to 2000 unprioritised ethnicity data was not available from the existing administrative data sources. Without using prioritised ethnicity data a large proportion of students who had borrowed in at least one year between 1997 and 2004, the target population in Table 2 for example, would have been recorded with unknown ethnicity.

The use of prioritised ethnicity data will tend to undercount the number of people in certain ethnic groups. The table below gives an example from the 2001 Census of the degree to which ethnicity is understated by prioritised ethnicity data. It shows the percentage by which the number of people identifying with each ethnic group is understated when individuals are assigned to one ethnic group only using the same order of priority applied in the pre-2000 student loan data.

## Percentage of ethnicities understated by prioritised ethnicity data

| Order of priority | Ethnic group | Under 15 years | 15–19 years | 20–24 years | 25–29 years | 30–34 years | 35–39 years | 40–44 years | 45 years and over | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NZ Māori | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Asian | 9 | 4 | 5 | 5 | 4 | 3 | 2 | 2 | 5 |
| 4 | Other | 13 | 7 | 7 | 5 | 3 | 4 | 4 | 4 | 7 |
| 5 | European | 19 | 15 | 13 | 11 | 8 | 7 | 6 | 3 | 9 |
| 2 | Pacific | 23 | 16 | 13 | 8 | 6 | 6 | 6 | 3 | 14 |

The number of Māori is not understated at all, because Māori is the first ethnic category in order of priority. Yet Pacific peoples, the second in order of priority, is most understated (14 percent overall).

The table also shows that the number of younger people, particularly those under 25 years, tends to be understated more because they are more likely to identify with multiple ethnic groups.

This suggests that while results for Māori will be unaffected by the use of prioritised ethnicity data in the student loan dataset, in the earlier years (pre-2000) the number of people in other ethnic groups will be understated – particularly European and Pacific peoples in the younger age groups. Tables which span the 1997 to 2004 period and include a combination of prioritised and unprioritised ethnicity data (eg Table 2) are likely to be affected to a lesser extent.

## Copyright

Information obtained from Statistics New Zealand may be freely used, reproduced, or quoted unless otherwise specified. In all cases Statistics New Zealand must be acknowledged as the source.

## Liability

While care has been used in processing, analysing and extracting information, Statistics New Zealand gives no warranty that the information supplied is free from error. Statistics New Zealand shall not be liable for any loss suffered through the use, directly or indirectly, of any information, product or service.

## Timing

Timed statistical releases are delivered using postal and electronic services provided by third parties. Delivery of these releases may be delayed by circumstances outside the control of Statistics New Zealand. Statistics New Zealand accepts no responsibility for any such delays.

**Next release ...**

*Integrated Data on Student Loan Borrowers: 1997–2005* will be released in 2007.

Statistics New Zealand: The first source of independent information for your key decisions.

# Tables

The following tables can be downloaded from the Statistics New Zealand website in Excel 97 format.  If you do not have access to Excel 97 or higher, you may use the Excel file viewer to view, print and export the contents of the file.

## List of tables