Australian Government
Department of Education, Skills and Employment

Factors Affecting Higher Education Completions

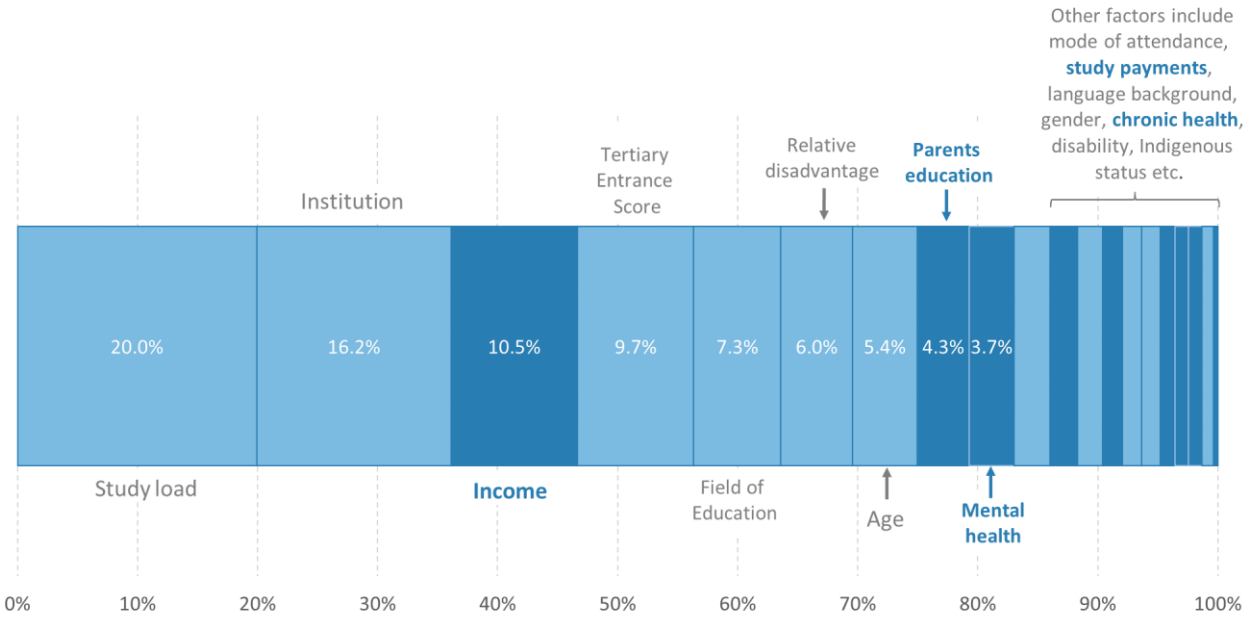
What predicts higher education student completion?

- *Combining health, tax, welfare and demographic data with student data improves our understanding of what drives higher education completions.*
- *The analysis predicts student university completions with 81 per cent accuracy.*

Predicting university completion rates

The completion of a university bachelor's degree has significant benefits for individuals, the economy and society in general.[i] Many circumstances and events in a student's life can influence the completion of a bachelor's degree. Existing quantitative research has been limited to what can be known about a student's situation based on self-reported survey or administrative information to universities upon enrolment, and what is reported by institutions over the period of a bachelor's degree student's study.[1,2] By linking a student, health, tax and welfare data through the Multi-Agency Data Integration Project (MADIP), we were able to generate new insights into what factors drive student completion (Figure 1).

Figure 1. Relative importance of factors predicting bachelor's degree completion, 2011-2016



Source: Multi-Agency Data Integration Project (MADIP), custom analytical extract, 2009-2016.

[1] Cherastidtham I, Norton A, and Mackey W (2018) *University attrition: What helps and what hinders university completion*. Grattan Institute.
[2] Marks, G (2007) Completing university: characteristics and outcomes of completing and non-completing students. *LSAY Research Reports, 55*.

Notes: The relative importance is how much a variable contributes to the accuracy of the random forest compared to the other variables. Random forest model balanced by SMOTE procedure. Test accuracy of the model was ~80.9 per cent. N= 118,956. Data are for those students who commenced a higher education course on a Commonwealth Supported Place in 2011. Factors in dark purple include a number of new demographics, health status and support payment variables sourced through MADIP. Factor definitions are in the Data and Methodology section.

We followed a cohort of students enrolled in a bachelor's degree for the first time in 2011 over five years in over 100 higher education institutions across Australia, we were able to capture the relative importance of 19 different factors on a student's ability to complete their study within the six-year window (see Data and Methodology). Figure 1 shows the relative importance of these 19 student-level and institutional factors in predicting student completion. We use the predictive model as a descriptive tool to complement the logistic regression. It returns a variable importance list, which is robust to collinearity and focuses on how well each variable predicts completions (as opposed to estimating effect sizes). The model produces an accuracy of 80.9 per cent. The number of courses per semester (study load), the institutions they attended, and their gross income while studying are the three most important factors predicting six-year completion rates of bachelor's degree students.

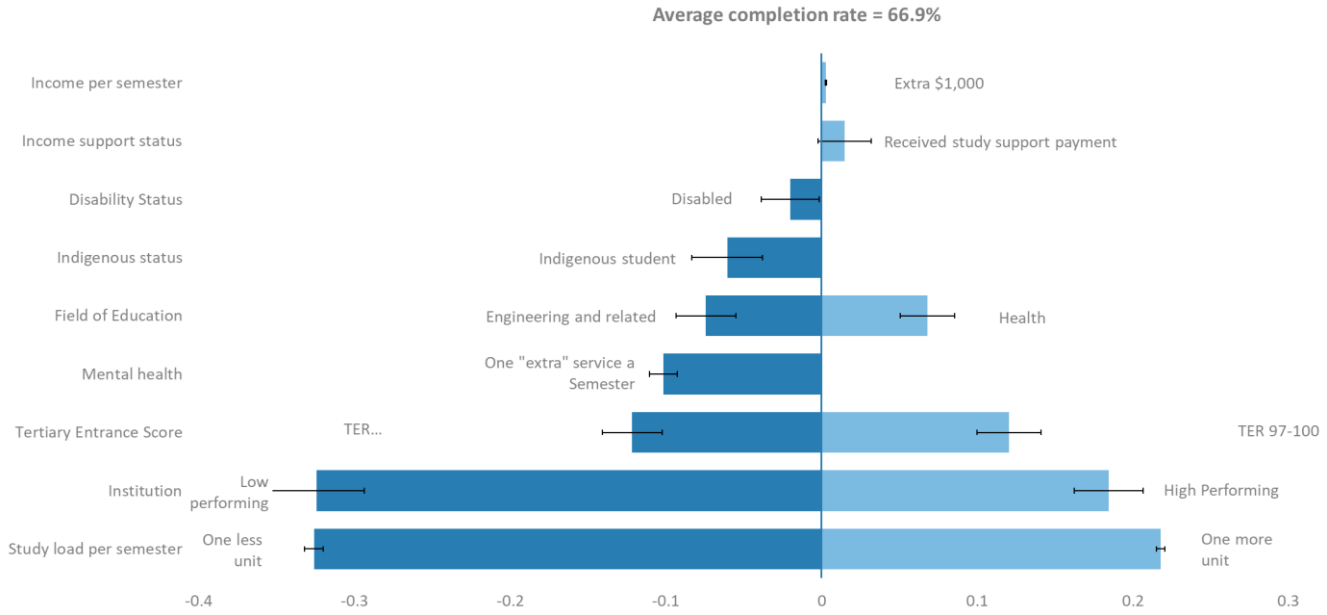
What factors affect university completion rates?

Our analysis also identified a number of socio-economic and background factors (such as parents' education, disability and accessing mental health services) as important predictors of completion of a bachelor's degree (see *Disability* and *Mental Health* factsheets). Figure 2 shows the relative size and direction of select variables with respect to the average completion rate for bachelor's degree university students, whilst holding everything else constant. Other factors such as taking a gap year or receiving study assistance are not as important (see *Study assistance* and *Gap year* factsheets).[3,4]

To read this chart, imagine that you enter university with the average (67 per cent) likelihood of completing your degree in six years. This likelihood to complete increases or decreases, depending on each factor as indicated in Figure 2. For example if you identify being an Aboriginal or Torres Strait Islander the likelihood of completing your study within six years will drop from 67 to 61 per cent; that is six percentage points as shown in Figure 2 (see *Indigenous* factsheet).[5,6] Equally, if you receive income support payments whilst studying your likelihood of completing your degree will increase by three percentage points from 67 per cent to 70 per cent.

[3] https://www.education.gov.au/transition-study-assistance
[4] https://www.education.gov.au/transition-gap-year
[5] Due to the model type being used to create these percentages, these two percentages are not additive, so an Indigenous student (minus 6 percentage points) who receives income support (plus 3 percentage points), your likelihood of completing, will not be three percentage points less.
[6] https://www.education.gov.au/transitions-indigenous

Figure 2. Marginal contribution of selected variables to the average higher education student completion rate, 2011-2016



Source: Multi-Agency Data Integration Project, custom analytical extract, 2009-2016.

Notes: The multinomial logistic regression (χ^2=31,734; p<0.0001; N= 118,956) had an eta-squared of 24.3 per cent. The eta-squared is analogous to R-squared and explains the amount of variance the model explains. The deviations from the average completion rate are calculated from the logistic regression model mentioned above. For categorical binary variables, the least square means are used, at the average completion rate and for continuous variables the odds ratios were applied at the average completion rate to measure the impact of a single unit change in those variables. Least Square Means levels for the continuous variables were set at, 20.4 years old at enrolment, 0.028754 chronic health services per semester, $5,693 dollars of income per semester, 3.828 units per semester and no mental health services per semester. Error bars are 95 per cent confidence intervals.

Consistent with Figure 1, study load and the institution attended are the two most important factors explaining the variance in the completion of bachelor's degrees (Figure 2). Previous research has also stressed the importance of these factors explaining students' attrition rates.[7,8,9] Other key results from the analysis show that:

- Not surprisingly, average study load per semester had the biggest impact on student six-year completion rates (Figure 2). Students taking on an extra unit a semester are 22 percentage points more likely to complete their studies within six years. Taking on one less unit a semester, equivalent to part time student status, is associated with being 32 percentage points less likely to complete (Figure 2 and Table 1).
- Tertiary entrance rankings (TER), which historically has been the primary factor in attaining acceptance into university is only the fourth most important factor (Figure 1) in predicting completion. After controlling for a number of factors, the data show that the isolated effect of TER is reduced. The effect of TER on completions is further investigated in the *Multiple Equity Groups* factsheet.[10]
- While income per semester contributes a lot to predictive accuracy (Figure 1), incremental changes to average income per semester had a much smaller relative impact on completion rates. For example, according to the model, the positive effect of taking one extra unit is equivalent to the impact of earning an extra $36,000 per semester. However, this does not take into account the impact of the relationship between the time needed to

[7] Cherastidtham I, Norton A, and Mackey W (2018) *University attrition: What helps and what hinders university completion*, Grattan Institute.
[8] Marks, G (2007) Completing university: characteristics and outcomes of completing and non-completing students. *LSAY Research Reports, 55*.
[9] Higher Education Standards Panel (2017) *Improving retention, completion and success in higher education: Discussion Paper*, Australian Government Department of Education and Training. ISBN: 978-1-76051-090-0.
[10] https://www.education.gov.au/transitions-equity

earn that money and how that affects the ability of the student to maintain a full-time study load. This relationship is further investigated in the *Work and Study* factsheet.[11]

Data and Methodology

This paper uses linked records from combining Higher Education Information Management System into the Multi-Agency Data Integration Project (MADIP). The MADIP data contains records from Census 2016, Social Security, Medicare Benefits Scheme, Personal Income Tax and the Pharmaceutical Benefits Scheme. The records have been de-identified and are accessed via that ABS DataLab, a secure server, run by the ABS who maintain the integrity of the data held on the DataLab.

This factsheet examines all students who commenced a Bachelor Graduate Entry and Bachelors Honours award courses on a Commonwealth Supported Place in 2011. There were 118,956 students in this cohort, of which 67 per cent completed their bachelor's studies within the 6-year time period.

We use a random forest model to determine with high accuracy the factors predicting completion rates. The random forest model creates a number of decision trees, which create a series of decision points (nodes) on each variable. The amount of accuracy gained at each node, is calculated for each tree in the model, using the Gini importance metric. The relative importance indicates how much a variable contributes to the accuracy of the random forest compared to the other variables.

The totals across all models are summed and converted to a relative percentage, which is presented in Figure 1. This metric is slightly biased towards variables with high cardinality, as the nodes in the trees have more options for splitting the variables.

The sample fed into the random forest model was imbalanced, so the Synthetic Minority Over-Sampling Technique (SMOTE) was used to balance the sample and lead to better prediction outcomes. The final random forest model had an accuracy in classification between completing and not completing of 80.9 per cent.

A multinomial logistic regression model was used to estimate the factors explaining six-year completion rates. The full 118,956 students in the cohort were used in this model. A total of 19 input variables were included and are detailed in Table 1, alongside their values of variation completion ($\eta 2$) from the multiple logistic regression.

More detail can be found in the *Methodology* factsheet.[12]

[11] https://www.education.gov.au/transitions-work-study
[12] https://www.education.gov.au/transitions-methodology

Table 1. Factors explaining the variation in completion of university bachelor's degrees

		Gap-year		First year in study (Gap-year)		First year in study (No gap-year)	
IRSAD (decile)	Likelihood of gap-year taking (%)	Likelihood of earning income from personal exertion (%)	Median income from personal exertion ($)	Likelihood of earning income from personal exertion (%)	Median income from personal exertion ($)	Likelihood of earning income from personal exertion (%)	Median income from personal exertion ($)
1 (lowest)	11 ± 1	83 ± 2	13,830	89 ± 1	13,760	66 ± 1	6,811
2	12 ± 0	85 ± 1	14,184	92 ± 1	13,971	71 ± 1	7,280
3	13 ± 0	90 ± 1	14,241	93 ± 1	13,612	75 ± 1	7,733
4	13 ± 0	89 ± 1	14,095	93 ± 1	13,982	78 ± 1	7,895
5	14 ± 0	88 ± 1	13,770	94 ± 1	13,841	79 ± 0	7,959
6	13 ± 0	90 ± 1	13,378	94 ± 1	13,314	79 ± 0	8,009
7	12 ± 0	90 ± 1	12,732	94 ± 1	13,435	80 ± 0	7,911
8	12 ± 0	88 ± 1	11,857	93 ± 1	13,302	80 ± 0	7,847
9	11 ± 0	86 ± 1	10,580	92 ± 1	12,533	81 ± 0	7,530
10 (highest)	14 ± 0	86 ± 1	8,517	93 ± 1	11,037	83 ± 0	6,838