



The likelihood of completing  
a VET qualification:  
A model-based approach

KEVIN MARK  
TOM KARMEL  
NATIONAL CENTRE FOR  
VOCATIONAL EDUCATION RESEARCH

The views and opinions expressed in this document are those of the authors  
and do not necessarily reflect the views of the Australian Government or state  
and territory governments.

**FOR CLARITY, PLEASE PRINT IN COLOUR**

© National Centre for Vocational Education Research, 2010

This work has been produced by the National Centre for Vocational Education Research (NCVER).  
Apart from any use permitted under the *Copyright Act 1968*, no part of this publication may be reproduced by any process without written permission. Requests should be made to NCVER.

The views and opinions expressed in this document are those of the author and do not necessarily reflect the views of NCVER.

ISBN 978 1 921809 03 3 web edition  
TD/TNC 100.14

Published by NCVER  
ABN 87 007 967 311

Level 11, 33 King William Street, Adelaide, SA 5000  
PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

ph +61 8 8230 8400 fax +61 8 8212 3436  
email [ncver@ncver.edu.au](mailto:ncver@ncver.edu.au)  
<<http://www.ncver.edu.au>>  
<<http://www.ncver.edu.au/publications/2272.html>>

## *The likelihood of completing a VET qualification: A model-based approach*

Kevin Mark and Tom Karmel, NCVER

This paper estimates vocational education and training (VET) course-completion rates, in order to fill a gap in performance measures for the VET sector.

The technique we use is to track all VET course enrolments within a three-year window, centred on the year of interest. Then, using an absorbing Markov chain model for a VET course enrolment, we estimate the proportion of VET course enrolments commencing in the year of interest that will eventually be completed. This approach allows us to estimate the completion rate without having to longitudinally track course enrolments over a long period of time.

## Key messages

- ✧ The national estimated completion rate of VET course enrolments at certificate I level or above commencing in 2005 is 27.1%. For full-time VET students aged 25 years and under in 2005, this rate is 34.7%.
- ✧ When cut by fields of education, the completion rates in 2005 range from 13.3% (for course enrolments in Mixed field programs) to 48.3% (for course enrolments in Education).
- ✧ Course enrolments at certificate III level had the highest rate of completion at 33.5% compared with other qualification levels.

One of the distinctive characteristics of the VET sector is that many students wish to learn specific skills and have no intention of completing a full qualification. For these students a more sensible measure of success is the proportion of modules passed (the load pass rate). Overall, the load pass rate of 2005 enrolments was 79.1%. NCVER is conducting a survey of student intentions in 2010–11, with a view to establishing the groups of students for which the qualification completion rate is an appropriate performance indicator.

Tom Karmel  
Managing Director, NCVER



# Contents

---

Tables and figures	6
Introduction	7
Methodology	9
The state space of the Markov chain	9
Definitions and assumptions	10
Findings	12
State and territory course-completion rates	12
Completion rates by broad fields of education	13
Completion rates by AQF levels of qualification	14
Final remarks	15
References	16
Appendices	
A: Completion rate of new VET students	17
B: Flow diagram	19
C: Completions formula	21
D: Sample sizes	23

# Tables and figures

---

## Tables

1	Calculated completion rates (and load pass rates) for 2005 commencing cohort of VET course enrolments, by state and territory, for the whole population and for full-time VET students aged 25 years and under with no prior VET qualification	12
2	Calculated completion rates (and load pass rates) for 2005 commencing cohort of VET course enrolments, by broad fields of education (FOE), for the whole population and for full-time VET students aged 25 years and under with no prior VET qualification	13
3	Calculated completion rates (and load pass rates) for 2005 commencing cohort of VET course enrolments, by AQF levels of qualification, for the whole population and for full-time VET students aged 25 years and under with no prior VET qualification	14
D1	Number of VET course enrolments commencing or continuing in 2005, by state and territory	23
D2	Number of VET course enrolments by a full-time student aged 25 years and below with no post-school qualification commencing or continuing in 2005, by state and territory	23
D3	Number of VET course enrolments commencing or continuing in 2005, by broad fields of education	24
D4	Number of VET course enrolments by a full-time student aged 25 years and below with no post-school qualification commencing or continuing in 2005, by broad fields of education	24
D5	Number of VET course enrolments commencing or continuing in 2005, by levels of qualification	24
D6	Number of VET course enrolments by a full-time student aged 25 years and below with no post-school qualification commencing or continuing in 2005, by levels of qualification	25

## Figures

B1	Defining the state of a VET course enrolment in year $n$	19
B2	Defining the state of a new VET student in year $n$	20

# Introduction

---

One of the challenges for the vocational education and training (VET) system is to come up with indicators of efficiency and effectiveness. There are a number of measures that are typically used, such as student satisfaction and employment and course completion. This paper is concerned with the last of these.

There are two fundamental concepts that arise when looking at completion rates. The first, module-completion<sup>1</sup> rates, is straightforward and is routinely published in the Australian National Report on the VET sector. It is simply the proportion of modules undertaken that are successfully completed, based on hours of training. The second, the rate at which qualifications are completed, is more problematic. The difficulties arise in two ways. First, technically it is far from straightforward because the VET student system is not based on a unique student identifier and the concept of commencement of a qualification is not well defined in the student system. For this reason, while key performance measures for the sector have included a course-completion rate (see, for example, KPM 1.5 of *Shaping our future* [ANTA 2004]), to date it has not been possible to calculate and publish such a measure. The second issue arises with the interpretation of the course-completion rate. The issue is that many individuals undertake particular VET modules with a view to obtaining particular skills rather than obtaining a complete qualification. Thus a completion rate has a difficult interpretation when we cannot distinguish between those people for whom a completed qualification is beneficial and those for whom particular modules are sufficient. One could argue that this argument is self-serving and it is not useful to conclude that course-completion rates are optimal because those who fail to complete have got what they needed from the VET system. A more useful approach is to treat both course-completion and module-completion rates as useful indicators, and to try to identify the groups for which the course-completion rate has particular salience. However, in any case, this argument is academic unless we have a measure of course-completion rates, and this is the topic of this paper. It uses mathematical techniques to obtain completion rates because direct observation is not possible.

The National Centre for Vocational Education Research (NCVER) collects enrolment and award details of VET students and the courses they enrol in. While the database is essentially cross-sectional by each year, there is enough information to match data over a number of years for individual VET students and the courses they undertake. Obtaining such a longitudinal dataset allows the use of mathematical techniques that rely on conditional probabilities to then calculate completion rates.

One such mathematical technique is a Markov chain. Markov chains have the property that the transition probabilities from one year to the next are not dependent on past transitions. While very simple, Markov chain models allow us to predict long-term completion rates.

We apply such tools to estimate the proportion of commencing VET course enrolments that will eventually be completed. From our VET collection, we create longitudinal data of the pathways that course enrolments take from 2004 to 2006, use that data to observe the one-step transitional probabilities of the state of course enrolments from 2005 to 2006 (with the states at the second period being: continuing study, completed the course enrolled in, and dropped-out), and then use

---

<sup>1</sup> Load pass rate in the terminology of the VET sector.

those probabilities and apply the techniques of absorbing Markov chain theory to calculate the completion rate of commencing course enrolments in 2005.

The next section describes the methodology. In the section that follows we present calculations of course-completion rates. We acknowledge that course-completion rates are not necessarily an appropriate measure of success for some groups for whom individual modules are more important. Thus we also present module-completion rates as a complementary measure. Tables are presented by state/territory, field of education and Australian Qualifications Framework (AQF) qualification level. Furthermore, a cut is made for young students studying full-time with no previous VET qualification, as well as for all VET students. One would expect completion rates to be relatively high for the former group, as they are typically new entrants to the labour market who are acquiring an initial post-school qualification. We find considerable variation by field of education for course-completion rates, while there is little variation for the module-completion rates. The paper ends with some concluding comments.



# Methodology

---

We create a longitudinal dataset of VET course enrolments, undertaken by matching students and the courses they enrol in over a number of years (from 2004 to 2006).

Each client record in NCVER's VET database has a student's date of birth, sex, and a variable called the 'encrypted\_id', which is a set of characters derived from the student's name via an encryption algorithm.<sup>2</sup> For the purposes of this project, a unique combination of these three variables defines an individual VET student, and thus enables us to track this student over the yearly datasets.

After identifying the student, we match each unique course they enrol in through the years via the variable 'course\_record\_id'. For each VET course enrolment in the longitudinal dataset, variables are set up to indicate the years (from 2004 to 2006) in which they are enrolled in the course, which year the course is commenced, and which year the course is completed (if it does get completed). This information is available from NCVER's VET derived course enrolment and award databases. These indicator variables will be used in definitions to determine what state of the Markov chain a course enrolment is in each year.

Demographic information (such as disability status, Indigenous status, state or territory of training) on the student and the course enrolment is also provided, as well as the AQF qualification or level of the course enrolled. This allows analysis on selected sub-groups.

Furthermore, we restrict our attention to VET course enrolments in AQF courses—Certificate I level and above. We thus exclude enrolments in non-AQF courses (secondary school level courses, non-award courses and courses not identifiable by level) for which the concept of completion is problematic.

## The state space of the Markov chain

In any one-year period, a VET course enrolment is classified as being in one of four states:

- ✧ commencing course year
- ✧ continuing course year
- ✧ dropped out of the course
- ✧ completed the course.

The assumption will be that a VET course enrolment can be in only one state for each year. So a course commenced in 2005, say, is counted as a commencing course in 2005 even if the course is completed that year. The course will then be classified as completed in the following year in 2006. This is acceptable because this will capture the completion of the VET student in the smallest time step (of one year).

---

<sup>2</sup> It may be possible that a change in a student's name will create a different encrypted id, which would prevent us from successfully tracking that certain individual; this is the best that we have to work with.

We model a VET course enrolment's path as a discrete-time stochastic process that can take the value of one of the four states in each yearly time step. We assume that the one-step transition from one state to another depends only on the present state that the course is in, and is independent of the past. This assumption is called a Markovian property, and hence this process is called a Markov chain.

Furthermore, the last two states are called absorbing states, since a course that has been completed or dropped out of forever remains a completed course or dropped out course. Hence, the model is an absorbing Markov chain. The remaining two states are transient states, in that a student in such a state will eventually leave it for another state (and finally be in one of the two absorbing states).

An alternative model to that presented here would be to consider the likelihood of a newly commencing VET student completing a course. The course that a student completes need not necessarily be the same course he or she commences; what counts is that a student completes a course.

In this case, the state space is changed to focus on the individual VET student rather than the individual VET course enrolment, that is:

- ✧ commencing student
- ✧ continuing student
- ✧ student dropped out of the VET course system
- ✧ course completing student.

Modelling a VET student does have its set of modelling difficulties, even though the interpretation is initially straightforward. The main difficulty is the problem of students who complete a course and enter the absorbing state of 'course completer', but then 'reappear' in the model again if they commence a new course. So the strict absorbing nature of the Markov model is compromised. At the same time, completion of a course by a recommencing student is also of interest. There are also difficulties in segmenting the data.

This methodology and the difficulties of this approach are discussed in more detail in appendix A, and we do not pursue this model further for this paper. The proper approach is to look at completion rates of VET course enrolments.

## Definitions and assumptions

We need to define the four states of a VET course enrolment listed above in terms of the enrolment and award data in our longitudinal dataset.

The definitions of the states will be for a year  $n$ . The states are defined by yearly enrolment and award information for each student over a two-year period, which is the year of consideration (year  $n$ ) and the year before (year  $n - 1$ ). The definitions may be visualised in the flow diagram in figure B in appendix B.

As a technical addition due to the nature of the dataset, a state will have to be made called 'Not in VET system', which is to exclude observations in the dataset that will not play a role in the year  $n$  of consideration for analysis, for example, a course in the dataset that is not enrolled or completed in years  $n - 1$  or  $n$  is not of interest in year  $n$ , and hence we shall assign this course enrolment to this dummy state.

### Commencing course in year $n$

A VET course in the longitudinal dataset is said to have commenced in year  $n$  by being enrolled in year  $n$  and the commencing flag variable states that this is a commencing year.

### Continuing course in year $n$

A VET course in the longitudinal dataset is said to be a continuing course enrolment in year  $n$  by being enrolled in both year  $n - 1$  and year  $n$ , and is not indicated as being completed in year  $n - 1$ . It is also a continuing course if it is not enrolled or indicated as completed in year  $n - 1$ , enrolled in year  $n$ , and not indicated as a commencing year in year  $n$ .

### Dropped out of course in year $n$

A course is considered to be dropped out of by a VET student in year  $n$  if it is enrolled in the previous year  $n - 1$ , but not enrolled in this year  $n$ , and is never indicated as being completed during this two-year window.

### Course completed in year $n$

The first rule will be that any course indicated to be completed in year  $n - 1$  will be regarded as a completed course in year  $n$ . This reflects the absorbing nature of this state. This rule will also capture the completion of any course that was both enrolled and completed in the previous year.

Secondly, any course not enrolled in year  $n$ , and awarded in year  $n$ , will also be a completed course in year  $n$ . This will cover courses that are enrolled in year  $n - 1$  but are not given their award until the following year  $n$ .

### Not in VET system

When classifying states of courses in year  $n$ , we are not interested in those in our dataset that have not been enrolled or given an award in either of year  $n - 1$  or year  $n$ . To exclude such course enrolments from our analysis in year  $n$ , we shall assign them this dummy state of Not in VET system for year  $n$ .

Using the definitions created, each course enrolment record in the longitudinal dataset is assigned a state in both year  $n = 2005$  and year  $n = 2006$ . Once the states are assigned, we look at the transitional probabilities of courses moving from one state in 2005 to another in 2006. Of interest to us are the transitional probabilities of commencing and continuing course enrolments in 2005 to the other states in 2006. Using these transitional probabilities, a completion rate for 2005 commencing course enrolments can be calculated using the formulae derived in appendix C.

Furthermore, the analysis can be refined by looking at certain subsets of the dataset based upon demographic data of students undertaking the courses. Examples include state and territory, occupation groups, or fields of education.

# Findings

Looking at the whole dataset, the national estimated completion rate of VET course enrolments commencing in 2005 is calculated at 27.1%. However, when restricting the analysis to only full-time VET students 25 years old and under in 2005 who also have no prior VET qualification, the estimated completion rate becomes 34.7%.

The calculated national completion rate of all course enrolments of 27.1% could be pulled down by students who enrol in a course to get certain skills from it, but never had the intention of completing the course the moment they have achieved this. So it should not be surprising that the completion rates after filtering the dataset to contain only full-time students 25 years old and under with no prior post-secondary school qualification increase. Such students are most likely to be completing their first after-school qualification to get a job and set up their careers (cf. university students). Also, this group is unlikely to contain VET students who have already gained qualifications and who are enrolling in courses merely to top up skills.

The whole population analysis is based upon following 1 310 143 VET course enrolments in 2005, of which 938 007 are commencing course enrolments and 372 136 are continuing course enrolments. The sample sizes by state and territory are presented in appendix D.

## State and territory course-completion rates

**Table 1** Calculated completion rates (and load pass rates) for 2005 commencing cohort of VET course enrolments, by state and territory, for the whole population and for full-time VET students aged 25 years and under with no prior VET qualification.

State	Completion rate of commencing course enrolments (%)		Load pass rates (%)	
	Whole population	Full-time, aged 25 years and under, with no prior VET qualification	Whole population	Full-time, aged 25 years and under, with no prior VET qualification
NSW	40.7	49.0	79.3	78.0
Vic.	22.4	30.6	77.5	76.4
Qld	19.6	27.7	81.4	77.3
SA	27.7	39.2	87.4	84.5
WA	24.4	36.8	74.7	77.7
Tas.	19.5	33.1	79.9	80.3
NT	11.0	17.9	72.8	77.6
ACT	41.0	58.9	80.4	79.4
<b>Australia</b>	<b>27.1</b>	<b>34.7</b>	<b>79.1</b>	<b>77.5</b>

The sample sizes to calculate the completion rates can be viewed in tables D1 and D2 of appendix D.

## Completion rates by broad fields of education

We now look at the estimated completion rates dissected by broad fields of education. This is the field of education at the 2-digit level, which is part of the Australian Standard Classification of Education. These are:

- FOE01 - Natural & physical sciences
- FOE02 - Information technology
- FOE03 - Engineering & related technologies
- FOE04 - Architecture & building
- FOE05 - Agriculture, environmental & related studies
- FOE06 - Health
- FOE07 - Education
- FOE08 - Management & commerce
- FOE09 - Society & culture
- FOE10 - Creative arts
- FOE11 - Food, hospitality & personal services
- FOE12 - Mixed field programmes

The estimated completion rates for a course commenced in 2005 are presented in table 2, along with the module-completer rates presented as load pass rates.

**Table 2** Calculated completion rates (and load pass rates) for 2005 commencing cohort of VET course enrolments, by broad fields of education (FOE), for the whole population and for full-time VET students aged 25 years and under with no prior VET qualification

FOE	Completion rate of commencing course enrolments (%)		Load pass rates (%)	
	Whole population	Full-time, aged 25 years and under, with no prior VET qualification	Whole population	Full-time, aged 25 years and under, with no prior VET qualification
Natural & physical sciences	29.5	28.5	74.6	74.0
Information technology	20.3	25.3	70.2	69.4
Engineering & related technologies	24.9	34.8	85.1	78.2
Architecture & building	27.0	42.1	84.8	82.4
Agriculture, environmental & related studies	18.1	35.8	83.2	81.2
Health	44.9	51.7	83.1	82.6
Education	48.3	24.3	84.0	73.6
Management & commerce	29.5	33.2	79.6	78.1
Society & culture	35.9	41.6	80.1	82.2
Creative arts	26.9	39.6	77.5	81.2
Food, hospitality & personal services	26.5	37.4	82.2	81.1
Mixed field programmes	13.3	28.5	56.0	61.2
<b>Total</b>	<b>27.1</b>	<b>34.7</b>	<b>79.1</b>	<b>77.5</b>

The sample sizes to calculate the completion rates can be viewed in tables D3 and D4 of appendix D.

## Completion rates by AQF levels of qualification

We now look at estimated completion rates of a cohort of VET students commencing a course in 2005 broken up into AQF qualification categories.

**Table 3** Calculated completion rates (and load pass rates) for 2005 commencing cohort of VET course enrolments, by AQF levels of qualification, for the whole population and for full-time VET students aged 25 years and under with no prior VET qualification.

AQF qualification level	Completion rate of commencing course enrolments (%)		Load pass rates (%)	
	Whole population	Full-time, aged 25 years and under, with no prior VET qualification	Whole population	Full-time, aged 25 years and under, with no prior VET qualification
Diploma and above	28.6	36.3	79.5	79.0
Certificate IV	31.7	34.5	77.0	76.2
Certificate III	33.5	40.3	83.6	79.5
Certificate I/II	19.9	28.8	72.7	70.8
<b>Total</b>	<b>27.1</b>	<b>34.7</b>	<b>79.1</b>	<b>77.5</b>

The sample sizes to calculate the completion rates can be viewed in tables D5 and D6 of appendix D.

## Final remarks

---

This report has outlined a methodology which uses an absorbing Markov chain model for estimating completion rates of commencing VET course enrolments. This methodology has then been applied to the Australian VET data collection from 2004 to 2006 to estimate the completion rates in 2005, and the report publishes some of its findings.

The main reason for adopting this methodology is that it minimises the number of years for which data need to be matched. This is a great advantage from a practical point of view (matching large datasets is never straightforward) but it is also an advantage in being able to calculate up-to-date completion rates. So, once 2009 data are available in mid-2010 we will be able to estimate completion rates for 2008. If we were to track students to completion we would need to match five or six years of data, and even then there may be continuing students who had not yet completed.

# References

---

- ANTA (Australian National Training Authority) 2004, *Shaping our future – Australia’s national strategy for vocational education and training 2004–2010: Measuring the future, key performance measures for vocational education and training 2004 to 2010, an implementation plan*, ANTA, Brisbane.
- Bhat, UN 1972, *Elements of applied stochastic processes*, John Wiley & Sons, New York.
- Isaacson, DL & Madsen RW 1976, *Markov chains: Theory and applications*, John Wiley & Sons, New York.
- Kemeny, JG & Snell, JL 1976, *Finite Markov chains*, Springer-Verlag, New York.



# Appendix A:

## Completion rate of new VET students

---

An alternative to tracking every individual course enrolment is to track a student's course enrolments over time. That is, each year we observe whether a student is enrolled in a course (or courses) or not each year, and whether a course is completed that year or not. The actual course or courses undertaken by the student is not of consideration. Thus, the longitudinal dataset will be rows of students, as opposed to course enrolments. In doing so, we are modelling the likelihood of a student commencing a VET course completing a VET course (be it the one that was commenced or another one).

Tracking a student this way has an advantage over tracking individual course ids because it is possible for course ids (or codes) for the same course to change from year to year. Changes in course id over the years make courses difficult to track when matching by course id. One example of a course changing its id is when the version number (typically the year of endorsement) of a training package changes. A student may enrol in a course with a training package endorsed in 1997, say, and continue in it the following year but in a training package endorsed in 2004. While the course is essentially the same, the change in version number will change the course id, and hence matching by course id will fail (thus creating a false dropout).

Another reason for a course's id to change is when the id has the qualification level of the course embedded in the code. Consider a student who enrolls in a course and then completes it but at a different qualification level. There is a change in qualification level, so there will not be a match in course ids and hence the enrolment will not be detected as being a completion. Conversely, tracking a student by enrolment, regardless of the course, would match the enrolment with the completion.

The methodology for calculating completion rates for a commencing VET student is similar to calculating completion rates for course enrolments. The only difference is in defining a commencing student; we will not use the commencing course flag variable as that is a variable specific for individual courses. Rather, a student is said to commence in year  $n$  if a student is not enrolled in it in the previous year  $n - 1$  but is enrolled in it this year  $n$ , and not given an award in either year  $n - 1$  or year  $n$ . See figure B of appendix B for a flow diagram that gives a visualisation of the definitions of the state space:

- ✧ commencing student
- ✧ continuing student
- ✧ student dropped out of the VET course system
- ✧ course completing student.

A drawback to analysing VET student enrolments is that the absorbing Markov model does not allow recommencing students to be included in calculations. That is because a student absorbed to a state such as course completer means that student is forever a course completer, and therefore new commencements should not take part in the calculations. Because of the absorbing nature of the model, the analysis is filtered to include only those students with no prior VET qualification in the year of consideration. Such students are assumed to be new VET students.

As with the method to be applied to course enrolments, completion rates can be calculated for various characteristics of the student. Calculations for demographic characteristics of the student such as age (at commencement) and sex are very straightforward.

We can attempt to apply the methodology to course characteristics by considering only students who have enrolled in a particular type of course (as a new student or as a continuing student). However, such an approach is a little unsatisfactory because a continuing student (in year 1 of our Markov chain) may have enrolled in a course different from that enrolled in as a commencing student. That is, trying to segment the data by course characteristics poses a difficulty for this student model.

As another example of segmenting by course characteristics being unsatisfactory, suppose a student commences a course, but changes and completes a course in a different field of education. Segmenting the commencing cohort by field of education creates a problem in interpretation when assigning the field of education for the completing student. Assigning the completer to the field of education of the course that was completed does not make sense as we seek completions by a commencing cohort.

# Appendix B: Flow diagram

The following flow diagram is a visualisation of the definitions of the states in the Markov chain model. Start from the left and work across to the right. The numbers labelled for each state in the diagram are arbitrarily chosen for programming and calculation purposes as listed in appendix C.

**Figure B1** Defining the state of a VET course enrolment in year  $n$

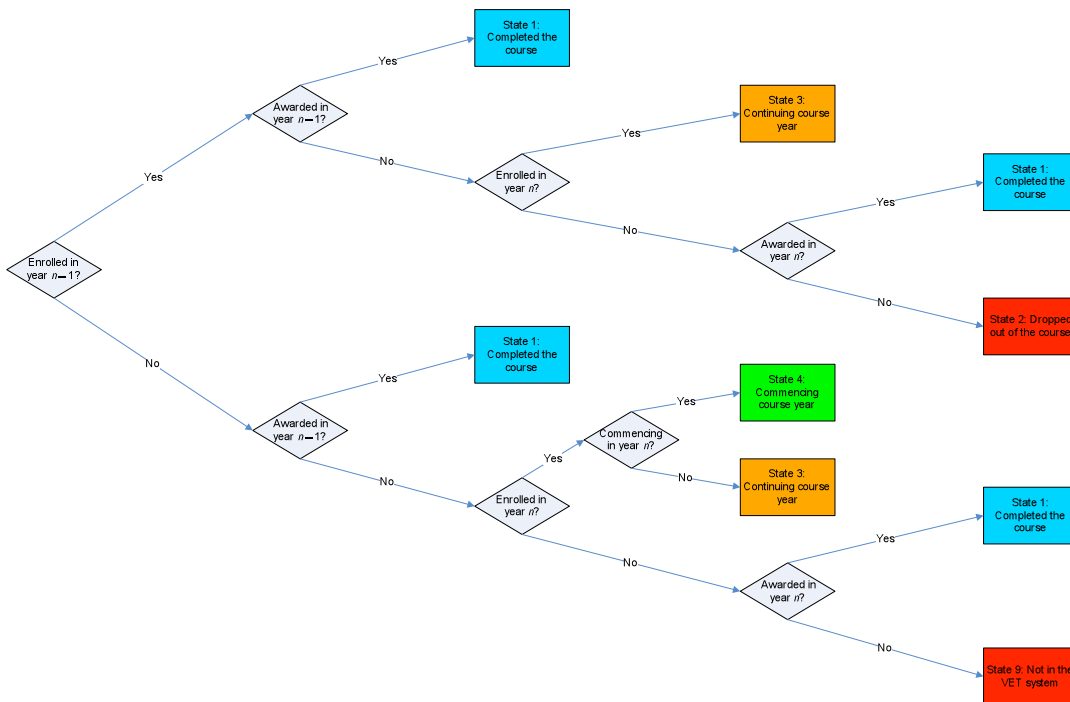
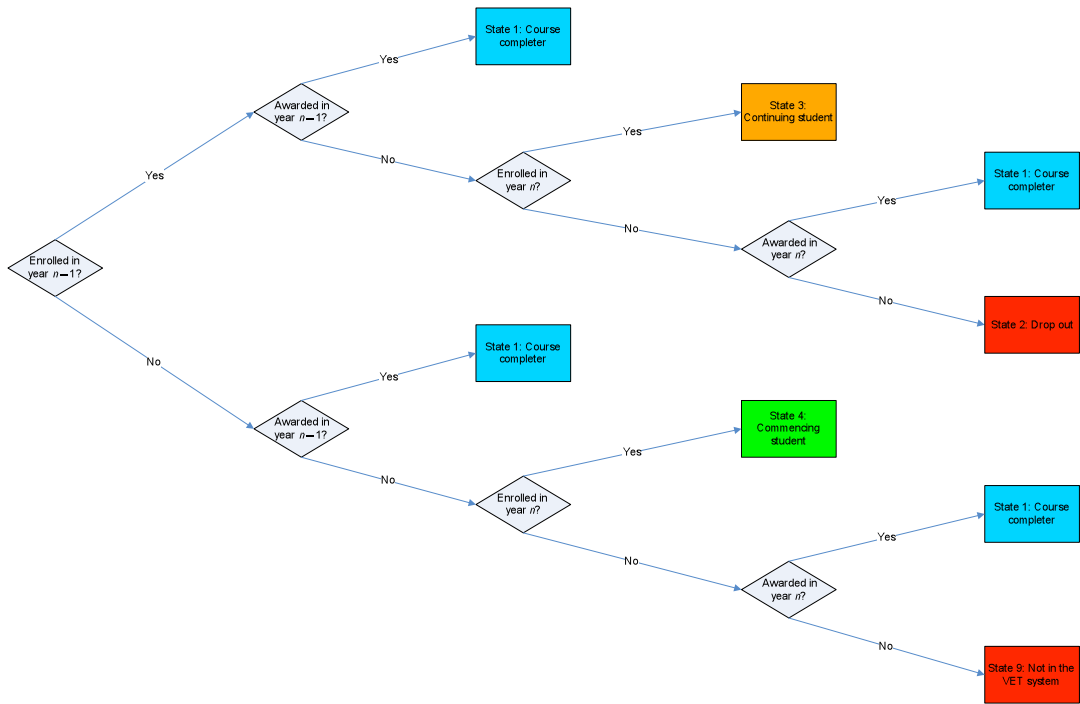


Figure B2 Defining the state of a new VET student in year  $n$



# Appendix C: Completions formula

---

The theory of absorbing Markov chains is reviewed in texts such as Bhat (1972), Isaacson and Madsen (1976), or Kemeny and Snell (1976). We do not repeat the theory here, but make use of their results to derive our formula. The model here is based on course enrolments, and can be repeated for new VET students (see appendix A).

## The transition probability matrix

Let  $p_{ij}$  be the conditional probability of a course enrolment moving from state  $i$  to state  $j$  in a one-step transition, also known as a transitional probability. Since there are four states in our model,  $i$  and  $j$  take values from 1 to 4 corresponding to the states as follows:

- 1 completed the course
- 2 dropped out of the course
- 3 continuing course year
- 4 commencing course year.

These conditional probabilities can be arranged in a four-by-four square matrix  $P = [p_{ij}]$ , called the transition probability matrix. The rows will naturally sum to 1.

As states 1 and 2 are absorbing states, we immediately have  $p_{11} = p_{22} = 1$  and the rest of the first two rows will be zeroes. Also, since a transition to a commencing course year from any of the four states is not possible, we have  $p_{34} = p_{44} = 0$ . With this information, our transition probability matrix will look like

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ p_{31} & p_{32} & p_{33} & 0 \\ p_{41} & p_{42} & p_{43} & 0 \end{bmatrix}.$$

So all that we need to find from our longitudinal dataset are the first three entries of the bottom two rows.

## Calculating the completion rate of a commencing course enrolment

Using the theory of absorbing Markov chains, the conditional probability of a commencing VET course enrolment eventually being completed is derived, to be

$$\Pr(\text{eventually completing course}) = p_{41} + p_{43} \frac{p_{31}}{p_{31} + p_{32}}.$$

This probability will be taken as the estimated completion rate of commencing course enrolments. The formula makes heuristic sense as well, as what it says is that the probability is equal to the conditional probability of a commencing course enrolment being completed in one time step, plus the conditional probability of a commencing course enrolment becoming a continuing course enrolment the following year times the net probability of a continuing course enrolment eventually being completed.

The probability of a commencing course enrolment not being completed is then simply 1 minus the probability above, or alternatively

$$\text{Pr(eventually dropping out)} = p_{42} + p_{43} \frac{p_{32}}{p_{31} + p_{32}}.$$

This will be taken as the estimated dropout rate of commencing course enrolments.

# Appendix D: Sample sizes

**Table D1** Number of VET course enrolments commencing or continuing in 2005, by state and territory

State	Commencing courses	Continuing courses	Total
NSW	253 730	82 745	336 475
Vic	262 135	123 517	385 652
Qld	190 822	89 043	279 865
SA	57 985	20 916	78 901
WA	111 064	30 351	141 415
Tas	31 709	12 452	44 161
NT	17 687	6 073	23 760
ACT	12 875	7 039	19 914
<b>Australia</b>	<b>938 007</b>	<b>372 136</b>	<b>1 310 143</b>

**Table D2** Number of VET course enrolments by a full-time student aged 25 years and below with no post-school qualification commencing or continuing in 2005, by state and territory

State	Commencing courses	Continuing courses	Total
NSW	10 959	2 709	13 668
Vic	23 526	9 187	32 713
Qld	10 577	4 990	15 567
SA	2 502	517	3 019
WA	7 057	1 305	8 362
Tas	122	100	222
NT	577	91	668
ACT	151	67	218
<b>Australia</b>	<b>55 471</b>	<b>18 966</b>	<b>74 437</b>

**Table D3 Number of VET course enrolments commencing or continuing in 2005, by broad fields of education**

FOE	Commencing courses	Continuing courses	Total
Natural & physical sciences	4 171	1 857	6 028
Information technology	47 203	14 452	61 655
Engineering & related technologies	155 061	86 808	241 869
Architecture & building	40 805	35 410	76 215
Agriculture, environmental & related studies	59 902	15 659	75 561
Health	22 709	10 669	33 378
Education	31 180	8 172	39 352
Management & commerce	241 086	83 666	324 752
Society & culture	121 010	43 924	164 934
Creative arts	32 408	12 095	44 503
Food, hospitality & personal services	89 032	31 611	120 643
Mixed field programmes	93 440	27 813	121 253
<b>Total</b>	<b>938 007</b>	<b>372 136</b>	<b>1 310 143</b>

**Table D4 Number of VET course enrolments by a full-time student aged 25 years and below with no post-school qualification commencing or continuing in 2005, by broad fields of education**

FOE	Commencing courses	Continuing courses	Total
Natural & physical sciences	716	268	984
Information technology	4 354	1 359	5 713
Engineering & related technologies	7 319	4 134	11 453
Architecture & building	2 932	1 523	4 455
Agriculture, environmental & related studies	2 260	550	2 810
Health	1 076	467	1 543
Education	353	28	381
Management & commerce	15 348	4 032	19 380
Society & culture	6 741	2 190	8 931
Creative arts	4 213	1 665	5 878
Food, hospitality & personal services	3 359	1 183	4 542
Mixed field programmes	6 800	1 567	8 367
<b>Total</b>	<b>55 471</b>	<b>18 966</b>	<b>74 437</b>

**Table D5 Number of VET course enrolments commencing or continuing in 2005, by levels of qualification**

AQF qualification level	Commencing courses	Continuing courses	Total
Diploma and above	110 841	72 685	183 526
Certificate IV	149 626	52 199	201 825
Certificate III	310 761	164 881	475 642
Certificate I/II	366 779	82 371	449 150
<b>Total</b>	<b>938 007</b>	<b>372 136</b>	<b>1 310 143</b>



**Table D6 Number of VET course enrolments by a full-time student aged 25 years and below with no post-school qualification commencing or continuing in 2005, by levels of qualification**

<b>AQF qualification level</b>	<b>Commencing courses</b>	<b>Continuing courses</b>	<b>Total</b>
Diploma and above	16 358	9 244	25 602
Certificate IV	11 012	2 093	13 105
Certificate III	13 615	4 865	18 480
Certificate I/II	14 486	2 764	17 250
<b>Total</b>	<b>55 471</b>	<b>18 966</b>	<b>74 437</b>