Unpacking VET Completion

How do we arrive at VET completion rates, with a special focus on what is driving the likelihood of VET completion

National Centre for Vocational Education Research
16 August 2018
Today’s Presenters

Brad McDonald
Data Analyst
National Collections

Adrian Ong
Research Officer
National Surveys
What's available out there?
Outlines of today’s presentation

1. How is program completion rate estimated?

2. What are the factors influencing the likelihood of completing a government-funded VET course?
Program completion rate for all Australian VET programs, Certificate I and above, that commenced in 2016

47%
The Completion Rate Methodology
The transition states

- Commencing program year
- Continuing program year
- Dropped out of the program
- Completed the program
Elements in the completion rate methodology

- Can the transition states be classified using information collected by NCVER?
- Is this information able to match programs and completions across years?
- Are there three years of data available?
<table>
<thead>
<tr>
<th>Program enrolments in 2016</th>
<th>Completed (1)</th>
<th>Dropped out (2)</th>
<th>Continuing (3)</th>
<th>Commencing (4)</th>
<th>Not in the system</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed (1)</td>
<td>6 779</td>
<td>10 793</td>
<td>4 229</td>
<td>0</td>
<td>552 451</td>
<td>574 252</td>
</tr>
<tr>
<td>Dropped out (2)</td>
<td>750</td>
<td>0</td>
<td>15 101</td>
<td>0</td>
<td>983 961</td>
<td>999 812</td>
</tr>
<tr>
<td>Continuing (3)</td>
<td>243 250</td>
<td>242 549</td>
<td>141 946</td>
<td>0</td>
<td>0</td>
<td>627 745</td>
</tr>
<tr>
<td>Commencing (4)</td>
<td>707 945</td>
<td>827 379</td>
<td>367 608</td>
<td>0</td>
<td>0</td>
<td>1 902 932</td>
</tr>
<tr>
<td>Not in the system</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 746 321</td>
<td>0</td>
<td>1 746 321</td>
</tr>
<tr>
<td>Program enrolments in 2016</td>
<td>Completed (1)</td>
<td>Dropped out (2)</td>
<td>Continuing (3)</td>
<td>Commencing (4)</td>
<td>Not in the system</td>
<td>Total</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Completed (1)</td>
<td>1.2%</td>
<td>1.9%</td>
<td>0.7%</td>
<td>0.0%</td>
<td>96.2%</td>
<td>100%</td>
</tr>
<tr>
<td>Dropped out (2)</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1.5%</td>
<td>0.0%</td>
<td>98.4%</td>
<td>100%</td>
</tr>
<tr>
<td>Continuing (3)</td>
<td>38.7%</td>
<td>38.6%</td>
<td>22.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100%</td>
</tr>
<tr>
<td>Commencing (4)</td>
<td>37.2%</td>
<td>43.5%</td>
<td>19.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100%</td>
</tr>
<tr>
<td>Not in the system</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
<td>0.0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Pr(eventually completing a course) = \( P_{41} + P_{43} \frac{P_{31}}{P_{31} + P_{32}} \)

\[ = 37.2 + 19.3 \frac{38.7}{38.7 + 38.6} = 46.9\% \]
The current matching process

- Student identifier, program identifier and training organisation/RTO are used in the matching process.

Student identifier = encrypted id + sex + date of birth

96ff9a470ff0bd3f592e20_F_01012000

Student enrolment in a single program (the matching key) =

Student identifier + program id + training organisation id

96ff9a470ff0bd3f592e20_F_01012000_ACM10110_123456
Can we use USI?

• The unique student identifier (USI) as the student identifier

  Student identifier = USI
  
  XDT60UYS5R

Individual enrolment in a single program (the matching key)=

  Student identifier + program id + training organisation id

  XDT60UYS5R_ACM10110_123456
A transition period to USI

- To ensure matching between enrolments and completions before and after 2015, when USI was implemented

- USI will only be used if:
  - The USI is valid
  - The program commenced 2015 or later
  - USI appears as both enrolment and completion

- The transition period will continue until 2020
Registered training organisation

- Individual RTO completion rates can be calculated with the current methodology

- Small program enrolments can distort projected completion rates

- NCVER has developed a set of rules on when completion rates can be calculated for an RTO
Rules for calculating RTO completion rates

• RTO must maintain a minimum of 50 commencing programs in the completion rate year and the following year to calculate an overall RTO rate

• For large RTOs, greater than 1000 enrolments, multiple demographic or training attributes may be calculated

• It is possible to group multiple RTOs to produce a combined completion rate
Publication

- Total VET completion rates to be published for first time

- A combined publication of government-funded and TVA released 13 August

- A total of 11 tables in the publication and a data visualisation product
VET program completion rates 2016: Data Slicer
What factors explain the likelihood of completing a VET qualification?
What factors explain the likelihood of completing a VET qualification?

Adrian Ong and Michelle Consoli
National Centre for Vocational Education Research
The Context

• A data analysis of some 2.4 million government-funded course enrolment records from 2011 and 2012

• The undertaking of this research is to explore the feasibility of using advanced data analytics while examining the factors that influence the likelihood of completing a VET qualification
Data Limitations

• The VET Provider Collection does not collect information about course duration

• Each state and territory is unique and there are certain artefacts which may not be captured in our data analysis
Our Student Cohort

Government-Funded Commencing Cohort of 2011 / 2012

Completers
• Students who were subsequently awarded a VET qualification between 2011 and 2016

Not Yet Completers
• Students whom we do not have any information about their completion status between 2011 and 2016
What we have looked at ....

- Age at commencement
- Client Apprenticeship Flag
- Disability Status
- Gender
- Highest Prior Education Level
- Indigenous Status
- Labour Force Status
- Mode of Attendance
- Course Qualification Level
- Student's At School Flag Status
- Student's Remoteness Status
- Whether the course was commenced full-time
- Student's self-assessment of their level of ability to speak English
- Whether the program was part of a training package
- Training Provider Type
- Course Field of Education
- Student's Socioeconomic Status
- Training Provider Course Enrolment Size
- State/territory that administered the funding of the training activity
Re-grouping the attributes

- Age at commencement
- Client apprenticeship flag
- Disability status
- Gender
- Highest prior education level
- Indigenous status
- Labour force status
- Last known mode of attendance
- Course field of education
- Course qualification level
- State/territory that administered the funding of the training activity
- Student’s at school flag status
- Student’s remoteness status
- Student’s self-assessment of their level of ability to speak English
- Student’s socioeconomic status
- Reason for undertaking training
- Training package flag
- Training provider course enrolment size (based on percentile rank)
- Training provider type
- Whether the course was commenced full-time
The Analytical Methods

Method 1: Generalised Logistic Mixed Regression

"Build" the Model

Challenges in handling skewed data

Limit how the information can describe completions

Method 2: Decision Tree Technique

Attempt to “Grow” the Model
(continuously re-evaluates other predictive variables while building the tree)

Inherently distribution-free
(Able to handle highly skewed data)

An easy way to learn and explore the interaction effects between student characteristics, provider characteristics and course characteristics
Method 1:
Generalised Logistic Mixed Regression

Ascertaining the likelihood of Completion Outcomes
In summary, who’re least likely to complete

**Model-based Mean Difference**

- Mixed field programs;
- Non full-time study;
- Last known mode of Attendance (Electronic only);
- Not part of an apprenticeship and/or traineeship scheme;
- Last known mode of Attendance (Correspondence only);
- Enrolled in a Certificate I program level of education

Baseline: 40%
Students more likely to complete their courses are:

- Community Education Providers (43%) versus Universities (32%).
- TAFEs (42%) versus Universities (32%).
- Other Registered RTOs (44%) as opposed to TAFEs (42%).
Students more likely to complete their courses are:

- Being part of an apprenticeship and/or traineeship scheme (51%) versus not being part of the scheme (30%).
- Diploma & above (44%) versus those enrolled in Certificate I (31%), and Certificate II (39%), and Certificate IV (42%).
- Certificate III (45%) versus Certificate IV (42%).
Likelihood of Completion Outcomes

Students more likely to complete their courses are:

- Enrolled in a program field of education other than in a Mixed Field program

Model-based Completion Rate by Course Field of Education:

- Mixed Field: 22%
- Society & Culture: 51%
- Health: 50%
- Education: 46%
- Management & Technology: 46%
- Natural & Physical Science: 42%
- Architecture & Building: 41%
- Engineering & Technology: 41%
- Food, Hospitality: 39%
- Agriculture: 38%
- Creative Arts: 35%
- Unknown: 34%
Students more likely to complete their courses are:

- Non-Indigenous (45%) as opposed to Indigenous (33%)
- Females (46%) versus Males (41%)
- Students with prior education (42%) versus Student without any prior education (38%).
- Employed full-time (44%) versus Not employed (not seeking employment) (39%).
Likelihood of Completion Outcomes

Students more likely to complete their courses are:

- Full-time study (56%) versus Non full-time study (26%)
- Mode of attendance (last known): Multimodal mode of learning (56%) as opposed to those who were enrolled purely in classroom (34%), electronic-based learning (30%) or employment-based learning (36%) or Correspondence learning (31%)
Method 2: Decision Tree Technique
(Classification & Regression Tree)
Decision Tree Technique

1. Decision tree technique results in different “pathways”

2. This technique gives us the ability to compute the overall relative contributing importance score for each attribute
Our Decision Tree (2011 cohort)

Overall classification accuracy: 68.6%

Likelihood of completion

Mean = 0.395 (N=1,148,672)

Course qualification level

Certificate III; Certificate IV; Diploma and above
Mean = 0.463 (66.0%)

Certificate I; Certificate II
Mean = 0.263 (34.0%)

Whether the course was commenced fulltime

No
Mean = 0.404 (46.9%)

Yes
Mean = 0.605 (19.1%)

Course field of education

Society and culture; Engineering and related technologies; Management and commerce; Health; Information technology; Creative arts
Mean = 0.354 (16.5%)

Architecture and building; Education; Mixed field programmes; Agriculture, environmental and related studies; Food, hospitality and personal services; Natural and physical sciences
Mean = 0.176 (17.4%)

Note: Mean refers the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Source: NCVER 2016 National VET Provider Collection.
The Top 10 “Influencers” (for 2011 cohort)

Most Important Attributes

1. Course field of Education
2. Labour force status
3. Course qualification level
4. Client apprenticeship flag (whether the course is part of an apprenticeship and/or traineeship)
5. Training Provider Type
6. Whether the course was commenced full-time
7. Last known mode of attendance
8. Training package flag (whether the course is part of a training package)
9. State/territory that administered the funding of the training activity
10. Reason for undertaking training

Less Important
Re-grouping the attributes

- Administration of Funding
- Provider Attributes
  - Likelihood of Completion
  - Provider Attributes
  - Course Attributes
  - Student Attributes
- Student Choice

Attributes:
- Age at commencement
- Client apprenticeship flag
- Disability status
- Gender
- Highest prior education level
- Indigenous status
- Labour force status
- Last known mode of attendance
- Course field of education
- Course qualification level
- State/territory that administered the funding of the training activity
- Student’s at school flag status
- Student’s remoteness status
- Student’s self-assessment of their level of ability to speak English
- Student’s socioeconomic status
- Reason for undertaking training
- Training package flag
- Training provider course enrolment size (based on percentile rank)
- Training provider type
- Whether the course was commenced full-time
Contributing Factors to the likelihood of completion

2011 cohort

- Course Attributes: 36%
- Provider Attributes: 10%
- Administration of Funding: 5%
- Student Choice: 14%
- Student Attributes: 35%
Attributes of Importance (2011 cohort)
Attributes of Importance (2011 cohort)

<table>
<thead>
<tr>
<th>State/territory</th>
<th>Course field of education</th>
<th>Course qualification level</th>
<th>Client apprenticeship flag</th>
<th>Training package flag</th>
<th>Training provider type</th>
<th>Training provider course enrolment size</th>
<th>Labour force status</th>
<th>Reason for undertaking training</th>
<th>Student at school flag status</th>
<th>Disability status</th>
<th>Age at commencement</th>
<th>Highest prior education level</th>
<th>Student's remoteness status</th>
<th>Indigenous status</th>
<th>Student's self-assessment on their ability to speak English</th>
<th>Student's socioeconomic status</th>
<th>Gender</th>
<th>Whether the course was commenced full-time</th>
<th>Student choice</th>
</tr>
</thead>
</table>
Attributes of Importance (2011 cohort)

- State/territory that administered the funding of the training activity
- Course field of education
- Course qualification level
- Client apprenticeship flag
- Training package flag
- Training provider type
- Training provider course enrolment size
- Labour force status
- Reason for undertaking training
- Student's at school flag status
- Disability status
- Age at commencement
- Highest prior education level
- Student's remoteness status
- Indigenous status
- Student's self-assessment on their ability to speak English
- Student's socioeconomic status
- Gender
- Last known mode of attendance
- Whether the course was commenced full-time
- Student choice
In Summary….

Field of education, labour force status and course qualification level are the three main factors as to whether a vocational education and training (VET) student is likely to complete a qualification.

Students enrolled in a certificate I or II qualification are among those less likely to complete.

Students in apprenticeships or traineeships and those students who enroll full-time are more likely to complete their VET courses.
https://www.ncver.edu.au/data/data/infographics/
what-factors-explain-the-likelihood-of-completing-a-vet-qualification
Let’s stay connected

Subscribe
at www.ncver.edu.au

Follow

twitter.com/ncver or @ncver

linkedin.com/company/ncver