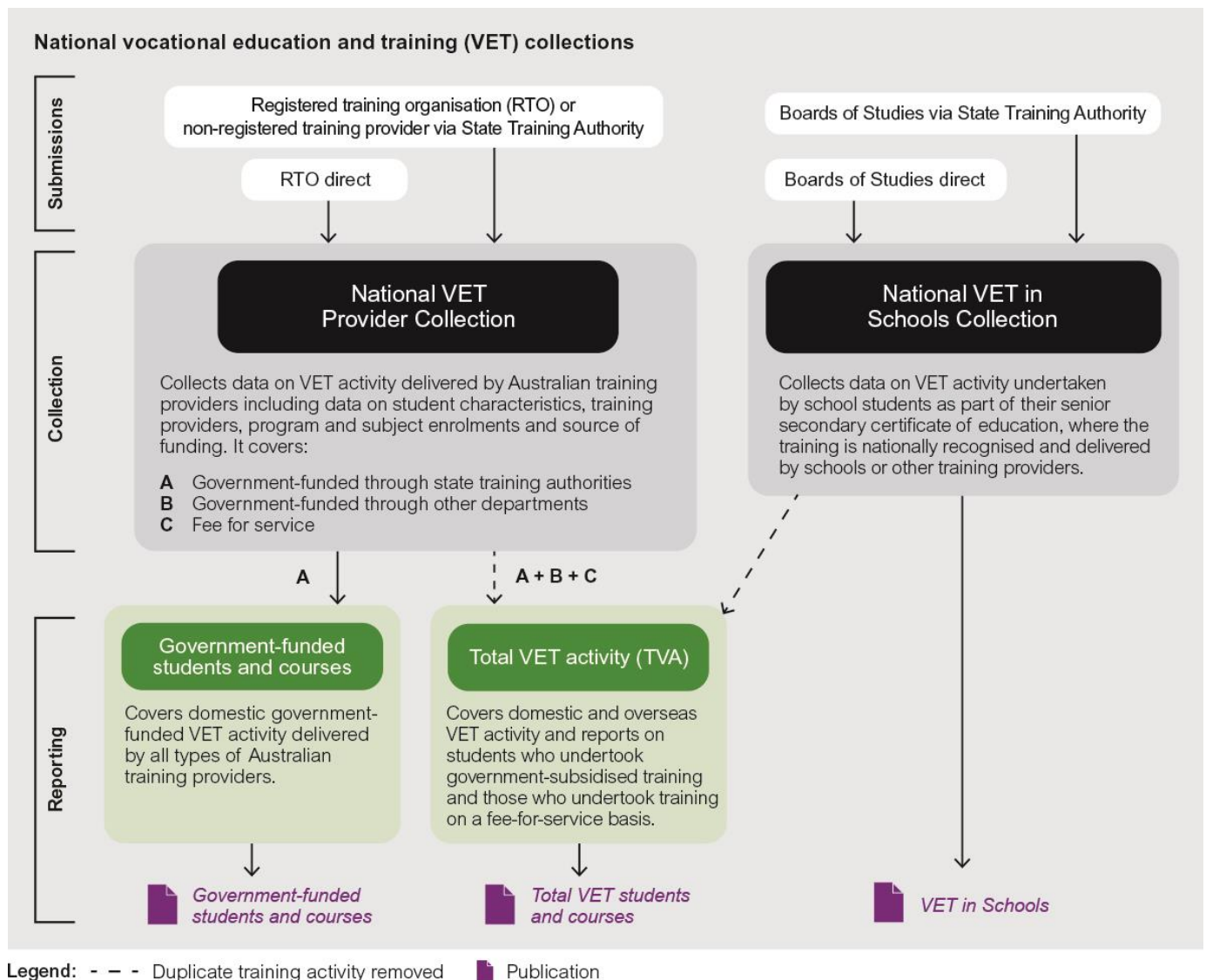


# TVA FACT SHEET

## De-duplication of training activity and student counts in 'total VET activity'

Total VET activity (TVA) data are sourced from both the National VET Provider Collection and the National VET in Schools Collection. These collections are compiled under the Australian Vocational Education and Training Management Information Statistical Standard (AVETMISS).

The below diagram illustrates how TVA data are sourced and where duplicate training activity is removed.



# FACT SHEET

## De-duplication of training activity

Since the introduction of total VET activity (TVA), the National Centre for Vocational Education Research (NCVER) has applied a process to identify and remove duplicate training activity where the same activity is reported for the same training provider in the same collection period via different data submitters.

When duplicate training activity is identified, the following hierarchy is applied such that only the training activity at the highest hierarchy level is reported:

- data submitted by a state training authority
- data submitted by a Board of Studies (or state training authority on behalf of a Board of Studies)
- data submitted directly to NCVER by a registered training organisation.

The results of the de-duplication process for subject enrolments are shown in table 1. The table shows the number of subject enrolments reported on data submission and the number after de-duplication of training activity. The last column shows the level of duplication. The table also shows the results of the de-duplication prior to total VET scope being applied (submitted data), and after total VET scope has been applied (published data). For further information on TVA scope, refer to the fact sheet *What is Total VET activity?* at <[https://www.ncver.edu.au/\\_\\_data/assets/pdf\\_file/0023/56381/What-is-total-VET-activity.pdf](https://www.ncver.edu.au/__data/assets/pdf_file/0023/56381/What-is-total-VET-activity.pdf)>.

Table 1 De-duplication of subject enrolments, 2017

	Subject enrolments		
	Before de-duplication	After de-duplication	Duplication
Submitted data	32 573 400	31 389 200	1 184 200
Published data	30 154 400	29 038 100	1 116 200

Source: National VET Provider Collection, 2017, National VET in Schools Collection, 2017

## De-duplication of student counts

Vocational education and training (VET) by nature is diverse, spanning a wide range of learning engagements from full-time programs across multiple years, short 'skill set' programs, to single subject enrolments. Student participation can also be wide ranging, with some students receiving training from multiple training organisations within the same year.

Prior to 2015, if a student was enrolled with more than one training provider during a collection period, it was possible for them to be counted more than once. From 1 January 2015, all new and continuing students undertaking nationally recognised VET in Australia are required to have a unique student identifier (USI) to receive a qualification or statement of attainment. The USI is a randomly generated alpha-numeric code recorded against any nationally recognised training undertaken and remains with an individual for life.

The implementation of the USI provides a mechanism with which to better estimate student counts by identifying and potentially removing duplicate student records. The NCVER has used a two-step process to de-duplicate student counts in *Total VET students and courses 2017*, which uses the USI where available and a count of distinct client identifiers within each submission for the residual data.

Table 2 shows the number of students with subject enrolments or program completions that were reported with a USI by submission pathway prior to total VET publication scope being applied.

# FACT SHEET

**Table 2 The number of students with subject enrolments or program completions that were reported with a USI by submission pathway**

Submission pathway	2015					2016					2017				
	USI reported		USI missing		Total students	USI reported		USI missing		Total students	USI reported		USI missing		Total students
	('000)	%	('000)	%	('000)	('000)	%	('000)	%	('000)	('000)	%	('000)	%	('000)
State training authority	1594.1	71	654.0	29	2248.1	1871.8	85	327.9	15	2199.7	1914.5	92	168.7	8	2083.3
Direct submitter	1712.2	74	600.1	26	2312.3	2302.4	83	471.2	17	2773.6	2833.5	95	156.2	5	2989.7
Boards of studies	103.7	49	108.7	51	212.3	129.2	62	80.6	38	209.7	125.4	63	75.2	37	200.7
<b>Total</b>	<b>3409.9</b>	<b>71</b>	<b>1362.8</b>	<b>29</b>	<b>4772.7</b>	<b>4303.4</b>	<b>83</b>	<b>879.6</b>	<b>17</b>	<b>5183.0</b>	<b>4873.4</b>	<b>92</b>	<b>400.1</b>	<b>8</b>	<b>5273.6</b>

Note: Includes all reported training activity. Does not take in to account training activity that is exempt from reporting a USI. No reporting scope has been applied to the figures in the table.

Sources: National VET Provider Collection, 2015-17, National VET in Schools Collection, 2015-17.

Table 3 provides estimated student counts before and after the application of the two-step de-duplication process, and prior to the total VET publication scope being applied.

**Table 3 Estimated student counts before and after application of the two-step de-duplication process**

	Estimated students		
	Before de-duplication	After de-duplication	Level of duplication
2015	4 772 700	4 181 200	12.4
2016	5 183 000	4 338 000	16.3
2017	5 273 600	4 357 500	17.4

Note: Includes all reported training activity. Does not take in to account training activity that is exempt from reporting a USI. No reporting scope has been applied to the figures in the table.

Sources: National VET Provider Collection, 2015-17, National VET in Schools Collection, 2015-17.

Caution must be used when comparing year-on-year de-duplicated student counts and participation rates based on these estimated student counts, due to differing rates of duplication across reported variables (for example, age, state or territory of student residence), which in part could be due to varying rates of USI provision. The table below shows the different rates of duplication for selected student characteristics from 2015 to 2017.

**Table 4 Rates of duplication in estimated student counts by selected student characteristics, 2015-17**

	Duplication rate (%)		
	2015	2016	2017
<b>Age</b>			
14 years and under	3.9	5.6	3.7
15 to 19 years	16.4	22.9	23.4
20 to 24 years	11.3	15.8	17.0
25 to 44 years	12.4	16.2	17.5
45 to 64 years	10.9	13.7	14.6
65 years and over	4.0	5.0	5.8
Not known	5.3	2.6	2.2
<b>Disability (including impairment or long-term condition)</b>			
With a disability	6.8	10.4	11.5
Without a disability	13.3	17.6	18.6
Not known	9.2	11.0	12.1

# FACT SHEET

**Table 4 cont. Rates of duplication in estimated student counts by selected student characteristics, 2015-17**

	Duplication rate (%)		
	2015	2016	2017
<b>Indigenous status</b>			
Indigenous	12.1	15.9	18.8
Non-Indigenous	13.8	17.7	18.8
Not known	4.1	6.6	5.5
<b>Sex</b>			
Males	14.5	18.6	20.0
Females	9.8	13.6	14.3
Not known	12.5	17.7	18.2
<b>School status</b>			
At school	17.9	23.2	24.1
Not at school	12.8	16.7	17.7
Not known	6.2	9.2	10.3
<b>Study mode</b>			
Full-time	24.1	30.4	31.9
Part-time	10.6	14.1	15.3
<b>Apprentice/trainee status</b>			
Apprentice or trainee undertaking off-the-job training	18.5	24.4	25.5
Not an apprentice or trainee	11.8	15.5	16.6
<b>Labour force status</b>			
Full-time employee	14.8	17.1	18.6
Part-time employee	11.3	14.5	15.4
Employed – other	7.5	9.8	10.2
Unemployed	9.8	14.0	15.4
Not employed – not seeking employment	7.9	11.3	12.1
Not known	13.2	19.8	20.3
<b>Provider type</b>			
TAFE	2.3	3.9	3.5
University	0.7	0.5	0.6
School	15.0	18.5	21.5
Community education provider	2.1	1.8	1.9
Enterprise provider	1.1	5.0	5.8
Private training provider	9.1	10.8	12.0
Attending more than one provider type	*	*	*
<b>SEIFA (IRSD)</b>			
Quintile 1 - most disadvantaged	11.4	15.9	17.0
Quintile 2	11.9	16.0	16.9
Quintile 3	12.0	15.6	16.3
Quintile 4	10.9	14.4	15.3
Quintile 5 - least disadvantaged	11.2	14.5	15.6
Not known	*	*	*

# FACT SHEET

**Table 4 cont. Rates of duplication in estimated student counts by selected student characteristics, 2015-17**

	Duplication rate (%)		
	2015	2016	2017
<b>Student remoteness region (ARIA+)</b>			
Major cities	11.4	15.5	16.5
Inner regional	12.8	17.0	17.9
Outer regional	14.4	18.5	19.5
Remote	15.3	19.6	20.7
Very remote	16.4	18.2	19.8
Overseas	2.2	3.5	5.3
Not known	24.4	23.9	27.0
<b>Student status</b>			
International students	8.4	12.0	15.1
Domestic students	12.7	16.6	17.6
<b>State or territory of student residence</b>			
New South Wales	12.7	17.9	18.9
Victoria	9.6	12.7	13.6
Queensland	12.4	15.9	17.0
South Australia	9.2	14.5	13.8
Western Australia	20.1	25.0	26.9
Tasmania	12.8	16.3	17.4
Northern Territory	14.4	20.3	22.1
Australian Capital Territory	16.0	15.5	16.2
Overseas	2.2	3.5	5.3
Other	18.1	17.6	20.4
<b>State or territory where the training was delivered</b>			
New South Wales	9.9	14.5	15.8
Victoria	8.1	10.6	11.2
Queensland	10.0	13.0	14.0
South Australia	7.2	11.9	10.8
Western Australia	18.9	21.7	22.4
Tasmania	11.4	14.1	14.9
Northern Territory	11.2	16.0	17.4
Australian Capital Territory	10.7	9.1	9.8
Overseas	0.0	0.0	0.0
Other	*	*	*
<b>TOTAL</b>	<b>12.4</b>	<b>16.3</b>	<b>17.4</b>

\*Student numbers in this category increased after de-duplication.

Sources: National VET Provider Collection, 2015-17, National VET in Schools Collection, 2015-17.

## Summary of USI data quality

Students reported with unknown or not stated demographic information are also most likely to be reported without a USI (table 5). Despite a general improvement in USI reporting over time, the percentage of students with missing USIs who also have missing demographics for age has increased year-on-year since 2015. As these students are reported without both demographic information and USIs, any existing or future supplementary de-duplication processes that rely on this variable will be ineffectual if data quality does not improve.

# FACT SHEET

**Table 5 Students with missing USIs by demographic characteristics**

Characteristic	Demographic	Total students and percentage with missing USIs					
		2015		2016		2017	
		Total students	% USI missing	Total students	% USI missing	Total students	% USI missing
Age	14 years and under	24 700	44	26 500	35	21 400	37
	15 to 19 years	896 100	31	932 400	18	937 000	16
	20 to 24 years	746 500	30	776 600	16	796 300	12
	25 to 44 years	1 962 800	26	2 143 800	14	2 206 300	11
	45 to 64 years	991 100	25	1 118 900	14	1 151 700	11
	65 years and over	78 400	45	84 400	32	85 900	28
	Not known	73 200	76	100 300	83	75 000	91
Sex	Males	2 521 400	25	2 644 500	14	2 748 100	11
	Females	2 067 700	32	2 346 700	19	2 381 300	16
	Not known	183 700	35	191 800	37	144 200	31
Indigenous status	Indigenous	163 900	32	204 400	13	179 000	11
	Non-Indigenous	3 945 100	27	4 344 900	15	4 538 200	12
	Not known	663 700	40	633 700	35	556 300	28
Student remoteness (ARIA)+region	Major cities	2 795 900	29	3 017 600	16	3 103 500	14
	Inner regional	944 000	26	1 009 200	15	1 009 900	11
	Outer regional	490 800	24	509 900	13	501 200	9
	Remote	90 400	22	92 100	12	89 000	8
	Very remote	54 500	24	53 300	14	51 700	8
	Overseas	173 900	52	179 300	42	193 600	29
	Not known	223 300	27	321 500	28	324 700	20
Disability	With a disability	197 700	27	207 600	14	208 500	10
	Without a disability	3 832 900	28	4 201 500	15	4 317 800	12
	Not known	742 200	34	773 900	29	747 300	23
<b>Total</b>		<b>4 772 700</b>	<b>29</b>	<b>5 183 000</b>	<b>17</b>	<b>5 273 600</b>	<b>14</b>

Notes: Percentages are of the row totals.

International students undertaking their entire VET course outside of Australia (offshore) are not required to have a USI.

Sources: National VET Provider Collection 2015-17; National VET in Schools Collection, 2015-17

It is worth noting, that although instances of missing USIs are decreasing over time, the number is unlikely to reach zero while valid exemptions exist. In addition, it is possible for students to have multiple USIs and although a mechanism exists to account for this in the USI Transcript Service, it can result in small numbers of duplication within the data.

The de-duplication process has been applied to estimated student counts in all 2017 TVA data products and tools. It has, however, only been applied to a limited number of demographic variables and for this reason some data products and tools do not contain estimated student counts (for example, estimated student counts are not available in VOCSTATS).

## FACT SHEET

It is also important to note that the information presented in the 'Atlas of total VET' data visualisation tool pertains to the student's usual residence, which is based on the statistical area level 2 (SA2). Using the SA2 of the student's usual residence as the basis of analysis means that there will be differences in training activity counts at the state or territory level from those published in *Total VET students and courses 2017*, where information is primarily reported on the state or territory of where the training was delivered.

For further information on the de-duplication of student counts in *Total VET students and courses 2017*, refer to the technical paper *Counting students in total VET students and courses*, available at [https://www.ncver.edu.au/\\_\\_data/assets/pdf\\_file/0028/796213/Technical-paper-Counting-students-in-Total-VET-students-and-courses.pdf](https://www.ncver.edu.au/__data/assets/pdf_file/0028/796213/Technical-paper-Counting-students-in-Total-VET-students-and-courses.pdf).