2023 VET Provider / VIS Collection Assurance Program

NCVER

Introduction

NCVER introduced a number of data quality checks and reports from 2016 into the AVETMISS Validation Software (AVS) to enable State training authorities (STAs) and Boards of Studies (BoSs) to self-assess the quality of their data prior to submission to the annual National VET Provider and VET in Schools Collections. From 2019, these data quality checks and reports are applied to quarterly submissions to the National VET Provider Collection for STAs.

Data quality exemptions

Where data quality checks generate a data quality error that prevents the data being submitted, the data submitter may apply to NCVER for a data quality exemption through the Collection Details page within AVS. For STA quarterly submissions (Q1-Q3), the data quality exemptions are automatically approved unlike in the annual collections where they are approved by NCVER's Collection coordinator. Once the exemption has been granted, the data submitter will need to re-validate the data in AVS. If the data is resubmitted, the exemption will need to be requested and applied again.

The following tables provide a list of these data quality checks and reports, outlining when they are applied[1].

Training Organisation (NAT00010A) supplementary file - data quality checks

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
10001	Training Org Id not on TGA	Details the *Training Organisation identifier(s)* which are not listed as a Registered Training organisation (RTO) on <training.gov.au> (TGA). Where the reported name matches a name recorded on TGA, the TGA details are provided. *Training organisation identifier(s)* should be listed on TGA.	STA/BoS	STA only

2023 VET Provider / VIS Collection Quality Assurance Program

Subject (NAT00060) - data quality checks

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
60002	Subject with greater than 400 Nominal Hours	Details the *Subject identifier(s)* reported in the current data collection where the *Nominal hours* are greater than 400 hours. This analysis excludes activity where *Outcome identifier - national* is '60 - credit transfer/national recognition', '61 - Superseded subject', '85 - Not yet started' and if the *Subject identifier* is reported on the agreed hours table with a *Nominal hours* value greater than 400. It is unlikely that subjects would be greater than 400 *Nominal hours*.	STA/BoS	STA only

NCVER

Client (NAT00080) - data quality checks

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
80001	Changing Client IDs across collections	Details the *Client identifier(s)* in the current data collection that may have changed since the previous annual data collection when the fields are matched by *Submitter reference*, *Training organisation identifier*, *Name for encryption*, *Gender* and *Date of birth*. It is unlikely that *Client identifier*(s) would change between collections.	STA/BoS	STA only
80002	Multiple Client IDs for matched clients	Details the *Client identifier*(s) in the current data collection where the same Submitter reference, *Training organisation identifier*, *Date of birth*, *Gender*, *Postcode* and *Name for encryption* are repeated for different *Client identifiers*. It is unlikely that the same student studying at the same Training organisation would have multiple *Client identifiers* in the current collection.	STA/BoS	STA only
80004	Common Date of birth	Details the *Date of birth*(s) reported for clients in the current data collection that represents more than 10% of clients per *Training organisation identifier*. This analysis excludes activity where the *Date of birth* is completely filled with '@ - not specified'. It is unlikely that more than 10% of students studying at the same Training organisation would have the same *Date of birth*.	STA/BoS	STA only
80008	Indigenous ID check across collections	Details *Client identifier*(s) that appear in the previous annual data collection with *Indigenous status identifier* is '1, 2, 3 - Aboriginal and/or Torres Strait Islander' but in the current VET data collection appear with *Indigenous status identifier* is '4 - neither Aboriginal nor Torres Strait Islander'. It is unlikely that the *Indigenous status identifier* would change between collections.	STA/BoS	STA only
80011 (VIS only)	Unexpected At school value	Details the *Client identifier*(s) in the current VET in Schools collection where the *At school flag* does not equal 'Y - the client is still attending secondary school'. It is unlikely that the clients in the VET in Schools collection would not be attending secondary school.	BoS only	No
80012	Count of compromised USI within a RTO	Details the *Unique student identifier*(s) [USI] that have been used multiple times for different clients when the fields are matched by *Training organisation identifier*, *Name for encryption, Gender*, and *Date of birth*. This analysis excludes exempt USIs. It is unlikely that within a training organisation the same USI, which is not an exemption code, would be reported against more than one student (Client identifier).	STA/BoS	STA only
80013	Blank USI by Training Organisation	Details the RTOs in the current collection where more than USI Validation allowed threshold (%) of their students are eligible for a USI have been reported with blank USIs. A USI is expected if the training is reported by an RTO listed on TGA and the *Client Identifier* is reported with a subject listed on TGA in the Training activity (NAT00120) file or reported with a program listed on training.gov.au in the Program completed (NAT00130) file. This analysis excludes RTOs with less than 100 students.	STA/BoS	STA only

Training activity (NAT00120) - data quality checks

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
12011	Duplicate subject activity	Details duplicate records in the current data collection when the fields are matched by *Training organisation identifier*, *Training organisation delivery location identifier*, *Client identifier*, *Program identifier, Subject identifier*, and *Outcome identifier - national* for records where the *Outcome identifier - national* is '20 - Competency achieved / pass', '51 - Recognition of prior learning (RPL) granted' or '61 - Superseded subject'. It is unlikely that a client would complete more than one identical subject with the same successful outcome for the same program in the same collection period.	STA/BoS	STA only
12015	Replicate subject activity	Details replicate records in the current data collection when the fields are matched by *Training organisation identifier, Client identifier*, *Program identifier*, *Subject identifier*, and where the *Outcome identifier - national* is '70 - Continuing activity' and where there are additional enrolment(s) where the *Outcome Identifier - National* is '20 - Competency achieved/ pass', '51 - Recognition of Prior Learning (RPL) granted', '60 - Credit transfer/national recognition', or '61 - Superseded subject'. It is unlikely that a client would achieve a successful outcome and be continuing in an identical subject in the same year at the same training organisation.	STA/BoS	STA only
12016	Continuing studies	Details training activity in the current data collection (when the fields are matched by *Client identifier*, *Subject identifier*, *Program identifier*) that were previously reported as expected to end during this collection period and have subsequently been reported with a new expected *Activity end date* in a future collection. The *Activity end date* should reflect the expected end point for training and assessment; it's not expected that this should be adjusted too often.	STA/BoS	No
12017	Discontinued Study	Details training activity reported in the previous annual data collection when the *Activity end date* (year) was the current collection year and the *Outcome identifier - national* was '70 - Continuing enrolment', but the training activity is not reported in the current data collection. Activity is matched on *Submitter reference*, *Training Organisation identifier*, *Client identifier*, *Program identifier* and *Subject identifier*. It is expected that training activity that was previously reported as continuing with an end date in the current collection year would be expected to be reported with a final outcome in the current collection.	STA/BoS	STA only
12019 (VIS only)	School programs	Details the training activity in the current VET in Schools Collection where the *Program level of education identifier* is lower than Certificate I (912-999). These programs are out of scope of the VET in Schools Collection.	BoS	No

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
12020	Client profiles	Details the *Client identifier*(s) in current data collection where the total of a client's nominal hours exceeds 1,500 hours. This analysis excludes activity where the *Outcome identifier - national* is '51 - Recognition of prior learning granted', '52 - Recognition of prior learning not granted', '60 - Credit transfer/national recognition', '61 - Superseded subject' and '85 - Not yet started'. It is unlikely that a client would participate in more than 1,500 hours of training in a single collection period.	STA/BoS	STA only
12022	Non-Assessable activity	Details subjects where all students who satisfactorily completed the subjects without assessment ('81 – Non-assessable – satisfactorily completed') but any unsuccessful subjects were reported with assessable outcomes ('40 – Withdrawn/discontinued' or '30 – Competency not achieved/fail'). Non-assessable outcomes are not included in NCVER's load pass rate calculation, a measure of an organisation's success in terms of the subjects that were successfully completed. A mixture of successful non-assessment and unsuccessful assessment outcomes will negatively impact load pass rate calculations.	STA/BoS	STA only
12023	Matching apprenticeship client and contract IDs to apprentice and trainee collection	Details the *Client identifier – apprenticeships*, *Training contract identifier* and *Program Identifier* / superseded *Program Identifier* combinations reported in the current data collection that do not exist in the Training contract transaction (APP00150) file as part of the National Apprenticeship Collection. This check references your state's National Apprenticeship Collection data that was submitted for the previous quarter via AVS. This check references the Accredited courses/Qualifications superseded tables from Training.gov.au and is updated at the beginning of each collection quarter.	STA/BoS	STA only
12026	Excessive activity	Details the *Client identifier*(s) in the current data collection with more than 40 subjects. This analysis excludes training activity where the *Outcome identifier - national* is '60 - Credit transfer/national recognition', '61 - Superseded subject' and '85 - Not yet started'. This analysis is cross tabulated by *Outcome identifier - national*. It is unlikely that a client would participate in more than 40 enrolments in a single collection period.	STA/BoS	STA only
12078	Matching Client identifier – apprenticeships to apprentice and trainee collection	Details *Client identifier - apprenticeships* reported in current data collection but do not exist in the Training Contract Transaction (APP00150) file as part of the National Apprenticeship Collection. This check references the previous quarter of the National Apprenticeship Collection submitted to AVS. All clients with a *Client identifier - apprenticeship* reported must have a matching identifier in the National Apprenticeship Collection.	STA/BoS	STA only
12086	Re-reported training activity from previous collection.	Details training activity reported in the current data collection that were also reported in the previous annual data collection with an *Outcome identifier - national* of '20 - Competency achieved / pass', '81 - Non assessed enrolment - satisfactorily completed', or '51 - RPL-granted'. Activity is matched on *Training organisation identifier*, *Client identifier*, *Program identifier*, *Subject identifier* and *Activity start date*. Subjects reported to a previous collection as being satisfactorily completed are not expected to be reported again in the current collection.	STA/BoS	STA only

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
12087	Previously reported successful subject activity	Details training activity reported in the current data collection where the client was reported in the previous annual data collection as completing the subject with an *Outcome identifier – national* of '20 - Competency achieved / pass', '81 - Non assessed enrolment - satisfactorily completed', or '51 - RPL-granted'. Activity is matched on *Training organisation identifier*, *Client identifier* and *Subject identifier*. This analysis excludes training activity reported in the current data collection where the *Outcome identifier - national* is '60 - Credit transfer/national recognition'. Clients are not expected to be reported in the current collect as undertaking subjects that were reported as being satisfactorily completed in the previous collection.	STA/BoS	STA only
12088	Suspect Client identifier – apprenticeships	Shows the frequencies of potentially inappropriate *Client identifier - apprenticeships* in the Training activity (NAT00120) file in the current collection (such as 'NA', 'TBA', 'X'). A potentially inappropriate *Client identifier - apprenticeships* is defined as not including any numeric digits (excludes where the field is '@' filled or blank).	STA/BoS	STA only
12089	Suspect Training Contract identifier (apprenticeship)	Shows the frequencies of potentially inappropriate *Training contract identifier*(s) in the Training activity (NAT00120) file in the current collection (such as 'NA', 'TBA', 'X'). A potentially inappropriate *Training contract identifier* is defined as not including any numeric digits (excludes where the field is '@' filled or blank).	STA/BoS	STA only
12111	International funding source	Details where *Funding source - national* is '31 - International client onshore - other revenue' or '32 - International client offshore - other revenue' for at least one subject but not all subjects, for that client, when matched on *Training organisation delivery location identifier*, *Client identifier*, and *Program identifier*. When a client is reported with *Funding source - national* of '31 - International client onshore - other revenue' or '32 - International client offshore - other revenue', it is expected that all their enrolments would be reported with the same funding source - national.	STA/BoS	STA only
12112	High number of Client Identifier – apprenticeship within an RTO reported as unknown	Details the number of enrolments for each training organisation which are more than the allowed threshold (2% STA, 20% BoS) of their apprentice or trainee enrolments reported with '@@@@@@@@@ - not specified'. This is based on the *Client Identifier – apprenticeship* in the Training activity (NAT00120) file being reported with '@@@@@@@@@ - not specified'. It is unlikely that within a training organisation the use of '@@@@@@@@@ - not specified' to denote an apprentice or trainee enrolment would exceed the allowed threshold of all apprentice or trainee enrolments in a single collection period.	STA/BoS	STA only
12113	High number of Training contract identifiers within an RTO reported as unknown	Details the number of enrolments for each training organisation where more than the allowed threshold (2% STA, 20% BoS) of their apprentice or trainee enrolments reported with '@@@@@@@@@ - not specified'. This is based on the *Training contract identifier* in the Training activity (NAT00120) file being reported with '@@@@@@@@@. - not specified'. It is unlikely that within a training organisation the use of '@@@@@@@@@ - not specified' to denote an apprentice or trainee enrolment would exceed the allowed threshold of all apprentice or trainee enrolments in a single collection period.	STA/BoS	STA only

Program completed (NAT00130) - data quality checks

Check No.	Check Name	Data Quality Check – Description	Active for annual	Active for quarterly
13001	Program completed again	Details the *Client identifier*(s) in the current data collection where the same *Training Organisation identifier*, *Client identifier*, *Program identifier* and the *Date Program Completed* (in combination) were reported in the previous annual data collection. Clients are not expected to complete the same program more than once.	STA/BoS	STA only
13005 (VIS only)	Program level	Details the program(s) completed in the current VET in Schools collection but the *Program level of education* is lower than Certificate I (912-999). These programs are not in scope for the VET in Schools collection.	BoS	No
13007	Year of Completion	Details the program(s) reported in the current data collection as completing in a previous annual data collection but where there is training activity for that *Client identifier*, *Training organisation identifier*, *Program identifier* (in combination) in the Training activity (NAT00120) file with an *Activity start date* on or after 1 January of the current collection year. Clients are not expected to have training activity in the current collection year when they are reported as completing the same *Program identifier* in a previous year.	STA/BoS	STA only
13008	Completions without subject activity missing enrolments	Details the program(s) completed in the current year where the client has no subject activity for that year. This analysis excludes records where a client is identified in the previous annual data collection's Training activity (NAT00120) file as an Apprentice/Trainee. Clients are expected to have training activity in the current collection year if they were reported as having completed a program in the same time period.	STA/BoS	STA only

Collection summary reports

Report Name	Data Quality Check – Description	Active for annual	Active for quarterly
Summary	Summary report detailing for each quality check whether any records were identified as meeting the check conditions (identified records), the total records submitted (base records), the recorded threshold (identified records divided by base records) and the allowed threshold. Where multiple training organisation reporting types exist within a submission, checks will be displayed by each reporting type.	STA/BoS	STA only
VET Scope	A summary of reported training activity, students and hours excluded and included in VET Scope; includes both the Publication and Government-funded scopes.	STA only	STA only
VET in Schools Scope	A summary of reported training activity, students and hours excluded and included in VET in Schools Scope.	BoS only	No
Program Completed Scope	A summary of reported program completions, identifying excluded activity from the likely published numbers. Scope exclusions are cumulative; program completions are kept unless exclusions applied to all reported activity for that program completion.	STA/BoS	STA only
Activity rates by submission	Details the average number of subjects and *Nominal hours* (based on reported Subject [NAT00060] file) per student, and the difference between the current quarter and previous corresponding quarter. Also details the average number of *Nominal hours* per subject and the difference between the current and previous collection years. Excludes subject activity with an *Outcome identifier – national* of '60 – Credit transfer/national recognition', '61 – Superseded subject', and '85 – Not yet started', and any students only associated with a completed program(s).	STA/BoS	STA only
Activity rates by Training organisation	Replicates the Activity rates by submission report for each reported training organisation; detailing the average number of subjects and *Nominal hours* (based on reported Subject [NAT00060] file) per student, and average *Nominal hours* per subject and the difference between current quarter and the corresponding previous quarter. Excludes subject activity with an *Outcome identifier – national* of '60 – Credit transfer/national recognition', '61 – Superseded subject', and '85 – Not yet started', and any students only associated with a completed program(s).	STA/BoS	STA only
Training organisation (COMP FREQ)	Comparative frequency report of the Training organisation (NAT00010 or 10A) files that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only

NCVER

Report Name	Data Quality Check – Description	Active for annual	Active for quarterly
Training organisation delivery location (COMP FREQ)	Comparative frequency report of the Training organisation delivery (NAT00020) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only
Program (COMP FREQ)	Comparative frequency report of the Program (NAT00030 or 30A) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only
Subject (COMP FREQ)	Comparative frequency report of the Subject (NAT00060) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only
Unknown Data Freq	Comparative analysis on the reporting of '@ - not specified' in the Client (NAT00080) file between the current collection quarter and the corresponding previous quarter. Analyses *Highest school level completed*, *Gender*, *Date of birth*, *Postcode*, *Indigenous status identifier*, *Language identifier*, *Labour force status identifier*, *Country identifier*, *Disability flag*, *Prior educational achievement flag*, and *At school flag*.	STA/BoS	STA only
Client (COMP FREQ)	Comparative frequency report of the Client (NAT00080) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only
USI compliance by Training Organisation	Provides an analysis for each Training organisation of their USI compliance for the current data collection. Noting the blank USI counts may include records not required to have a USI e.g. non-nationally recognised training).	STA/BoS	STA only
Disability (COMP FREQ)	Comparative frequency report of the Disability (NAT00090) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only
Prior educational achievement (COMP FREQ)	Comparative frequency report of the Prior educational achievement (NAT00100) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only

2023 VET Provider / VIS Collection Quality Assurance Program

Report Name	Data Quality Check – Description	Active for annual	Active for quarterly
Outcome brief	A cross-tab analysis of the current quarter collection's subject activity and the previous corresponding quarter, where the *Outcome identifier - national* is cross-tabbed by *Subject flag* and *VET flag* and is measured in terms of enrolments and nominal hours	STA/BoS	STA only
Funding Source brief	A cross-tab analysis of the current quarter collection's subject activity and previous corresponding quarter, where each *Funding source - national* is measured in terms of enrolments and *Nominal hours*.	STA/BoS	STA only
Training activity (COMP FREQ)	Comparative frequency report of the Training activity (NAT00120) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only
Program completed (COMP FREQ)	Comparative frequency report of the Program completed (NAT00130) file that compares the current quarter with the previous corresponding quarter.	STA/BoS	STA only

1. Please note:

   - the term 'data collection' used throughout the data quality checks refers to either the National VET Provider or VET in Schools (VIS) collections.

   - where the data quality check or report refers to the:

     - previous annual data collection means the previous year's January – December [Q4] data collection (e.g. current: 2019 Jan-Mar [Q1] versus previous: 2018 Jan-Dec [Q4]).

     - previous corresponding quarter means the previous year's corresponding quarter (e.g. current: 2019 Jan-Mar [Q1] versus previous: 2018 Jan-Mar [Q1])